

STKIP PGRI JOMBANG



Media Kampus

Jurnal Pendidikan



Volume 7, Nomor 1, Maret 2012
ISSN 0853-6937

Media Kampus

Volume 7 Nomor 1 Maret 2012

- I** **Daftar Isi**
- 1-8 Meningkatkan Keingintahuan Peserta Didik
(Hendra Darmawan)
- 9-15 Konsep dan Strategi Pembelajaran Literasi Media Melalui Pendekatan Bahasa Kritis
(Ahmad Sirulhaq)
- 16-23 Analisis Karakter Tokoh Utama Pada Novel Sosial: Sang Nabi Karya Abdillah Firmansyah Haem
(Syafuddin)
- 24-30 Analisis Pengelompokan (Cluster Analysis) dengan Metode K-Means Cluster
(Abd Rozak)
- 31-40 Islam dan pembelajaran yang menaruhbukkan Sikap dan mental kewirausahaan
(Munawaroh)
- 41-46 An Analysis On English Try Out Test
(Prilya Rochana Wati)
- 47-55 Black Feminist Spirit Against Racism And Sexism As Reflected In Sula
(Stella Rose Qur)
- 56-61 Femrocritique: Bahasa Anak Usia Dini
(Henry Sulistyowati)
- 62-65 Resonansi Dan Disonansi Pendidikan Nilai Bagi Peserta Didik
(Kusumo)
- 70-80 Meningkatkan Kemampuan Matematika Siswa Melalui Penerapan Pendekatan Kontekstual Dalam Model Pembelajaran Kooperatif Tipe Tai , Teknik Tsis Dan Pembelajaran Langsung
(Zulfan Ritonga)
- 81-85 Model Kurikulum Pendidikan Anti Korupsi, Kolusi Dan Nepotisme (Anti-KKN) Di Sekolah
(Nur Wahyuningstih)
- 86-92 Vietnamization In Richard Wright's Native Son
(Rif'ah Inayati)
- 93 **Petunjuk Bagi Penulis**

ANALISIS PENGELOMPOKAN (*CLUSTER ANALYSIS*) DENGAN METODE *K-MEANS CLUSTER*

Abd. Rozak, S.Pd., M.Si

ABSTRAK

Analisis pengelompokan merupakan salah satu analisis data eksploratori yang bertujuan untuk menentukan kelompok dari sekelompok data. Analisis *cluster* sendiri digunakan dalam berbagai aspek pengetahuan, diantaranya pengelompokan data mining, *market analysis* ataupun juga informasi perilaku atau karakteristik wilayah tertentu. Dalam kajian ini akan menganalisis pengelompokan Kabupaten di Jawa Timur dengan metode *K-means cluster* berdasarkan enam faktor yang mempengaruhi IPM, dengan analisis ini diharapkan dapat memberikan gambaran tentang kedekatan faktor yang mempengaruhi IPM sehingga dapat diambil kebijakan yang sesuai dengan permasalahan yang ada. Dari enam *cluster* yang ditentukan, untuk *cluster 3*, *cluster 4* dan *cluster 5* lebih banyak nilai variabel yang di bawah rata-rata, sedangkan untuk *cluster 1*, *cluster 2* dan *cluster 6* lebih banyak nilai variabel di atas rata-rata. Perlu diambil kebijakan pada *cluster* tertentu pada nilai variabel yang berada di bawah rata-rata sesuai dengan kebijakan yang strategis terhadap peningkatan IPM di Jawa Timur.

Kata kunci : *K-means cluster*, IPM Jawa Timur

1. Pendahuluan

Analisis pengelompokan (*cluster analysis*) merupakan salah satu analisis data eksploratori yang bertujuan untuk menentukan kelompok dari sekelompok data. Pertama kali metode ini dikembangkan dengan menemukan struktur pengelompokan di antara objek yang akan dikelompokkan. Paradigma pengelompokan mulai banyak diminati berbagai kalangan dan ditulis dalam berbagai paper dan jurnal. Analisis *cluster* sendiri digunakan dalam berbagai aspek pengetahuan, diantaranya adalah dalam formulasi hipotesis untuk pengelompokan data mining, *market analysis* ataupun juga informasi perilaku atau karakteristik wilayah tertentu.

Perkembangan analisis *cluster* dimulai dari metode *hierarchical* yang secara garis besar membentuk sebuah *tree diagram (treelike)* yang biasa disebut dengan dendogram yang mendeskripsikan pengelompokan berdasarkan jarak, metode *partitional* lebih dikenal dengan metode *non-hierarchy* termasuk didalamnya adalah metode *K-means cluster*.

Beberapa penelitian dengan menggunakan Analisis *Cluster* sudah dilakukan, diantaranya Pengelompokan kecamatan di kabupaten Tuban

berdasarkan 12 karakteristik pendidikan pernah dilakukan sebelumnya oleh Pravitasari (2006) dengan metode *Single Linkage*. Purwaningsih (2007) Analisis *Cluster* Terhadap Tingkat Pencemaran Udara Pada Sektor Industri di Jawa Tengah.

Pembangunan dapat diartikan sebagai usaha atau proses untuk melakukan perubahan ke arah yang lebih baik, dalam pelaksanaannya pembangunan bersifat multi dimensional dan memiliki berbagai kompleksitas masalah. Proses pembangunan terjadi di semua aspek kehidupan masyarakat, baik aspek ekonomi, politik, sosial, maupun budaya. Menurut *United Nations Development Programme / UNDP* (1990) pembangunan manusia adalah model pembangunan yang bertujuan untuk memperluas pilihan masyarakat melalui usaha-usaha untuk memberdayakan masyarakat. Sebagai indikator pembangunan manusia, UNDP telah mengembangkan *Human Development Index / Indeks Pembangunan Manusia (IPM)*. Penelitian yang sudah banyak dilakukan hanya pada menentukan faktor-faktor yang mempengaruhi IPM di wilayah tertentu (Nur, 2009). Kajian ini akan menganalisis pengelompokan Kabupaten di Jawa Timur berdasarkan enam faktor yang mempengaruhi IPM, dengan analisis ini diharapkan dapat memberikan gambaran tentang kedekatan faktor yang mempengaruhi IPM sehingga dapat diambil kebijakan yang sesuai dengan permasalahan yang ada.

2. Kajian Teori

A. Analisis Cluster

1. Definisi

Analisis multivariat merupakan analisis dimana masalah yang diteliti bersifat multidimensional dan menggunakan tiga atau lebih variabel, Analisis ini berhubungan dengan semua teknik statistik yang secara simultan menganalisis sejumlah pengukuran pada individu atau objek. (Santoso, 2010).

Analisis multivariat dibagi menjadi dua kategori utama yaitu:

1. *Dependence Method*.

Dependence Method atau analisis ketergantungan digunakan apabila tujuan dari analisis adalah untuk menjelaskan atau memprediksi variabel respon berdasarkan dua atau lebih variabel prediktor. Metode ini terdiri dari 4 macam yaitu *Multiple Regression Analysis*, Analisis Diskriminasi Berganda (*Multiple*

Discriminant Analysis), Analisis Multivariat Varians (*Multivariate Analysis of Variance*), dan Analisis Korelasi Kanonikal (*Canonical Correlation Analysis*).

2. *Interdependence Method*.

Metode yang digunakan untuk menjelaskan seperangkat variabel atau mengelompokkan berdasarkan variabel-variabel tertentu. Metode ini dikelompokkan menjadi tiga yaitu Analisis Faktor (*Factor Analysis*), Analisis *Cluster* (*Cluster Analysis*), dan Skala Multidimensional (*Multidimensional Scaling*).

Cluster dapat diartikan sebagai kelompok, pada dasarnya analisis *cluster* akan menghasilkan sejumlah kelompok. Analisis *Cluster* adalah teknik yang digunakan untuk mengidentifikasi objek atau individu yang mempunyai kemiripan dengan memperhatikan beberapa kriteria. Analisis *cluster* yaitu analisis untuk mengelompokkan objek yang mirip sebagai objek penelitian menjadi kelompok (*cluster*) yang berbeda dan *mutually exclusive*.

Analisis *cluster* termasuk dalam analisis statistik multivariat metode interdependen. Sebagai alat analisis interdependen maka tujuan analisis *cluster* tidak untuk menghubungkan ataupun membedakan dengan sampel / variabel lain. Analisis *cluster* merupakan salah satu alat analisis yang berguna sebagai peringkasan data. Dalam meringkas data ini dapat dilakukan dengan jalan mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu di antara objek-objek yang hendak diteliti.

Jika terdapat n obyek dan p - variat, maka observasi x_{ij} dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, p$, dapat digambarkan sebagai berikut:

	Variabel 1	Variabel 2	...	Variabel j	...	Variabel 5
Objek 1	X_{11}	X_{12}	...	X_{1j}	...	X_{1p}
Objek 2	X_{21}	X_{22}	...	X_{2j}	...	X_{2p}
.
.
.
Objek i	X_{i1}	X_{i2}		X_{ij}		X_{ip}
.
.
.
Objek n	X_{n1}	X_{n2}	...	X_{nj}	...	X_{np}

2. Konsep Dasar dalam Analisis Cluster

Analisis *cluster* merupakan suatu kelas teknik, dipergunakan untuk mengklasifikasi obyek atau kasus ke dalam kelompok yang relatif homogen atau mirip (*similarity*) yang disebut *cluster*. Pengelompokan dilakukan berdasarkan kemiripan (*similarity*) antar obyek. Kemiripan diperoleh dengan meminimalkan jarak antar obyek dalam kelompok (*within-cluster*) dan memaksimalkan jarak antar kelompok (*between-cluster*).

Proses pengolahan data sehingga data mentah dikelompokkan menjadi beberapa *cluster* adalah sebagai berikut:

1) Melakukan proses standarisasi data

proses ini dilakukan jika masing-masing data mempunyai perbedaan satuan yang cukup besar.

2) Menentukan ukuran jarak antar data/objek

Untuk mengetahui kemiripan antara objek dapat dilakukan dengan mengukur seberapa jauh kesamaan antara objek tersebut. Hal ini dapat dilakukan dengan:

a) mengukur korelasi antar objek

jika data tergabung dalam satu *cluster*, maka data tersebut akan mempunyai korelasi yang tinggi.

b) mengukur jarak antar objek

Jarak yang digunakan adalah jarak *euclidean*. Ukuran jarak atau ketidaksamaan antar obyek ke-i dengan obyek ke-h, disimbolkan dengan d_{ih} , diperoleh melalui rumus :

$$d_{ih} = \sum_{j=1}^p (x_{ij} - x_{hj})^2$$

dengan:

d_{ih} = jarak kuadrat Euclidean antar obyek ke-i dengan obyek ke-h.

p = jumlah variabel *cluster*.

x_{ij} = nilai atau data dari obyek ke-i pada variabel ke-j.

x_{hj} = nilai atau data dari obyek ke-h pada variabel ke-j

c) mengukur asosiasi antar objek

dilakukan jika data dalam bentuk non metrik.

3) Proses *Clustering*

Proses *cluster* atau pengelompokan data bisa dilakukan dengan dua metode:

a. Metode Hirarki (*Hierarchical Method*)

Metode ini memulai pengelompokan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian operasi diteruskan ke objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk semacam pohon (*treelike*) dimana ada hirarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai paling tidak mirip.

b. Metode Non-hirarki (*Hierarchical Method*)

Metode ini dimulai dengan proses penentuan jumlah *cluster* terlebih dahulu. Metode Non-hirarki yang digunakan adalah *k-means*. Metode *k-means* digunakan sebagai alternatif metode *cluster* untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan metode hirarki.

4) Melakukan validasi dan profiling *Cluster*

Dilakukan dengan tujuan untuk menjelaskan karakteristik setiap *cluster* berdasarkan profil tertentu dalam objek.

3. Asumsi dalam Analisis *Cluster*

Dalam Analisis *Cluster* diperlukan asumsi terhadap objek yang ada, asumsi tersebut adalah :

- 1) Sampel yang diambil mewakili dari populasi yang ada
- 2) Multikolinieritas, yaitu kemungkinan terjadi korelasi antar objek tidak ada, tetapi jika terdapat korelasi maka tidak terlalu tinggi.

Adapun ciri-ciri *cluster* yang baik adalah:

- 1) Homogenitas (kesamaan) yang tinggi antar anggota dalam satu *cluster* (*within-cluster*).
- 2) Heterogenitas (perbedaan) yang tinggi antar *cluster* yang satu dengan *cluster* yang lainnya (*between-cluster*).

4. K-Means Cluster

Penggunaan *k-means* untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam *cluster* tertentu berdasarkan rata-rata terdekat. Proses Pengelusteran dengan metode *k-means* adalah :

- 1) Menentukan besarnya *k*, yaitu banyaknya *cluster* dan menentukan *centroid* di tiap *cluster*.
- 2) Menghitung jarak tiap objek dengan setiap *centroid*.
- 3) Menghitung kembali rata-rata (*centroid*) untuk *cluster* yang baru terbentuk.
- 4) Mengulangi langkah 2 sampai tidak ada lagi pemindahan objek antar *cluster*

Setelah *cluster* terbentuk, baik dengan metode hirarki maupun non hirarki, langkah selanjutnya melakukan interpretasi terhadap *cluster* yang terbentuk, yang pada intinya memberi nama spesifik untuk menggambar isi *cluster* tersebut.

- 5) Melakukan validasi *cluster*.

Untuk menguji validasi *cluster* digunakan uji parsial F.

Hipotesis:

H_0 : variabel *i* bukan variabel pembeda dalam pengelusteran.

H_1 : variabel *i* merupakan variabel pembeda dalam pengelusteran

Dengan taraf signifikansi α dan Statistik uji

$$F = \frac{\text{Mean square kluster}}{\text{mean square error}}$$

Kriteria Uji :

Tolak H_0 jika $F > F_{\alpha, k-1, n-k}$ Atau $p\text{-value} < \alpha$

B. Indeks Pembangunan Manusia (IPM)

IPM merupakan suatu gambaran komprehensif mengenai tingkat pencapaian pembangunan manusia di suatu daerah dan sebagai dampak dari kegiatan pembangunan yang dilakukan di daerah tersebut (BPS, 2008). IPM dibangun melalui pendekatan tiga dimensi dasar, yang meliputi usia hidup (*longevity*), pengetahuan (*knowledge*), dan standar hidup layak (*decent living*).

1. Kategori IPM

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

2. Rumus Penghitungan IPM

Tahap pertama penghitungan IPM adalah menghitung indeks masing-masing komponen IPM dengan rumus (BPS, 2008)

$$Indeks(X_i) = \frac{(X_i - X_{\min})}{(X_{\max} - X_{\min})}$$

Dengan

X_i : Indikator komponen pembangunan manusia ke-i, $i= 1, 2, 3$

X_{\min} : Nilai minimum X_i

X_{\max} : Nilai Maksimum X_i

Tahapan kedua perhitungan IPM adalah menghitung rata-rata sederhana dari masing-masing-masing indeks X_i dengan rumus

$$IPM = \{X_1 + X_2 + X_3\} / 3$$

Dengan

X_1 : Indeks harapan hidup

X_2 : Indeks pendidikan = $2/3$ (indeks melek huruf) + $1/3$ (indeks rata-rata lama sekolah)

X_3 : Indeks standar hidup layak

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Data yang diambil adalah data IPM tahun 2007 Provinsi Jawa Timur. Adapun unit observasi penelitian ini adalah kabupaten/kota di kedua provinsi di mana pada tahun 2007, Provinsi Jawa Timur terdiri dari 38 kabupaten / kota.

3. Variabel Penelitian

Variabel- variabel yang digunakan adalah:

Tabel 1. Variabel Penelitian

No.	Kode	Nama Variabel	Tipe Variabel
1		Variabel Respon	
	Y	Indeks Pembangunan Manusia (IPM)	kontinu
2		Variabel Prediktor	
	X ₁	Persentase penduduk yang tinggal di daerah perkotaan	kontinu
	X ₂	Persentase penduduk yang berpendidikan di atas SLTP	kontinu
	X ₃	Rata-rata pendapatan perkapita	kontinu
	X ₄	Rasio ketergantungan penduduk	kontinu
	X ₅	Peranan sektor industri dalam PDRB	kontinu
	X ₆	Persentase penduduk miskin	kontinu

3. Metode Penelitian

Pada kajian ini metode yang digunakan adalah metode K-Mean *Cluster*, digunakan untuk melakukan pengelompokan kabupaten di Jawa Timur berdasarkan enam variabel-variabel yang mempengaruhi IPM (sesuai dengan penelitian sebelumnya, enam variabel tersebut signifikan), dengan langkah-langkah sebagai berikut:

- 1) Melakukan deskripsi data
- 2) Melakukan standarisasi data jika terdapat perbedaan satuan data yang cukup besar.
- 3) Menentukan besarnya k, yaitu banyaknya cluter dan menentukan *centroid* di tiap *cluster*.
- 4) Menghitung jarak tiap objek dengan setiap *centroid*.
- 5) Menghitung kembali rata-rata (*centroid*) untuk *cluster* yang baru terbentuk.
- 6) Mengulangi langkah 4 sampai tidak ada lagi pemindahan objek antar *cluster* (iterasi)
- 7) Melakukan validasi *cluster*.
- 8) Untuk menguji validasi *cluster* digunakan uji parsial F.
- 9) Melakukan analisis hasil.

4. Pembahasan

Dengan menggunakan perangkat lunak SPSS 13, dilakukan proses analisis.

- 1) Melakukan deskripsi data

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
X1	38	9.33	100.00	45.9350	30.94923
X2	38	12.64	66.50	39.7755	13.11670
X3	38	191.62	457.06	265.7589	67.50631
X4	38	.40	.69	.4929	.06116
X5	38	.96	72.64	15.9858	15.36608
X6	38	7.07	39.42	19.6426	7.57717
Valid N (listwise)	38				

- 2) Melakukan standarisasi data jika terdapat perbedaan satuan data yang cukup besar.

Pada tahapan ini dilakukan standarisasi data dengan tujuan agar selisih data tidak terlalu besar.

- 3) Menentukan besarnya k, yaitu banyaknya cluster dan menentukan *centroid* di tiap *cluster*.

Pada analisis ini diambil nilai *cluster* sebanyak enam kelompok, sebenarnya untuk penentuan banyaknya *cluster* sangat tergantung dengan kebijakan yang akan dijalankan. Diperoleh *cluster* awal :

Initial Cluster Centers

	Cluster					
	1	2	3	4	5	6
Zscore(X1)	1.74689	1.74689	-1.02313	-1.16045	.39629	1.68421
Zscore(X2)	1.45040	1.40999	-1.01287	-2.06878	.19399	1.56476
Zscore(X3)	1.12643	.61122	-.92182	-.90375	.11897	2.83382
Zscore(X4)	-.37437	-.37437	-1.51898	3.22300	.77025	-1.35547
Zscore(X5)	-.20733	3.68697	-.87894	-.97785	-.23987	1.29599
Zscore(X6)	-1.00600	-.78824	1.76020	2.61013	.20685	-1.64344

Pada tahap ini belum bias dilakukan analisis, dikarenakan masih ada proses iterasi sehingga didapatkan *cluster* akhir

Iteration History^a

Iteration	Change in Cluster Centers					
	1	2	3	4	5	6
1	.444	1.018	1.272	.438	1.100	.236
2	.227	.254	.462	.146	.175	.371
3	.038	.064	.038	.049	.012	.093
4	.006	.428	.003	.016	.143	.023
5	.001	.086	.000	.005	.010	.006
6	.000	.017	2.23E-005	.002	.001	.001
7	2.92E-005	.003	1.86E-006	.001	5.20E-005	.000
8	4.86E-006	.001	1.55E-007	.000	3.72E-006	9.07E-005
9	8.10E-007	.000	1.29E-008	6.68E-005	2.66E-007	2.27E-005
10	1.35E-007	2.74E-005	1.07E-009	2.23E-005	1.90E-008	5.67E-006

a. Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is 2.01E-005. The current iteration is 10. The minimum distance between initial centers is 2.561.

4) *Cluster Akhir*

Final Cluster Centers

	Cluster					
	1	2	3	4	5	6
Zscore(X1)	1.37525	.81714	-.74234	-.97014	-.40129	1.72600
Zscore(X2)	1.09635	.95142	-.74832	-1.84159	-.20574	1.76730
Zscore(X3)	1.07769	.68003	-.74686	-.96752	-.36941	2.28143
Zscore(X4)	-.63599	-.45613	-.38923	2.97772	.53126	-1.19195
Zscore(X5)	-.14199	2.25036	-.35233	-.87080	-.41828	.92113
Zscore(X6)	-.98251	-.42927	.72660	2.09146	-.05415	-1.61397

Tanda negatif menunjukkan berada di bawah rata-rata dan tanda positif menunjukkan nilai di atas rata-rata, hal ini karena data telah distandarkan.

5) *Jarak Tiap Cluster*

Distances between Final Cluster Centers

Cluster	1	2	3	4	5	6
1		2.560	3.774	6.431	3.041	1.965
2	2.560		3.933	6.450	3.487	2.787
3	3.774	3.933		3.842	1.419	5.416
4	6.431	6.450	3.842		3.762	8.073
5	3.041	3.487	1.419	3.762		4.758
6	1.965	2.787	5.416	8.073	4.758	

6) Untuk menguji validasi *cluster* digunakan uji parsial F.

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(X1)	6.220	5	.184	32	33.751	.000
Zscore(X2)	6.499	5	.141	32	46.149	.000
Zscore(X3)	6.611	5	.123	32	53.606	.000
Zscore(X4)	6.037	5	.213	32	28.353	.000
Zscore(X5)	5.612	5	.279	32	20.086	.000
Zscore(X6)	5.594	5	.282	32	19.830	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Analisis ini digunakan untuk mengetahui seberapa besar variabel tersebut berpengaruh dalam proses *cluster*, jika dilihat dari nilai F yang cukup besar, dan dari nilai sign < 0,05, (jika $\alpha = 5\%$) maka semua variabel signifikan dalam proses *cluster* atau merupakan variabel pembeda dalam proses *cluster*.

7) Jumlah kabupaten tiap *cluster*

Number of Cases in each Cluster

Cluster	1	5.000
	2	4.000
	3	11.000
	4	2.000
	5	13.000
	6	3.000
Valid		38.000
Missing		.000

8) Keanggotaan dan jarak *cluster*

Cluster Membership

Case Number	Kabupaten	Cluster	Distance
1	01. Pacitan	5	1.223
2	02. Ponorogo	5	.802
3	03. Trenggal	5	.694
4	04. Tulungag	5	.767
5	05. Blitar	5	.940
6	06. Kediri	5	.654
7	07. Malang	5	.903
8	08. Lumajang	3	.786
9	09. Jember	5	.698
10	10. Banyuwana	5	.631
11	11. Bondowos	3	.703
12	12. Situbond	3	1.524
13	13. Probolin	3	.901
14	14. Pasuruan	3	1.696
15	15. Sidoarjo	2	1.143
16	16. Mojokert	2	1.552
17	17. Jombang	5	1.092
18	18. Nganjuk	5	.737
19	19. Madiun	5	.543
20	20. Magetan	5	.613
21	21. Ngawi	3	.835
22	22. Bojonego	3	.651
23	23. Tuban	3	.873
24	24. Lamongan	3	.974
25	25. Gresik	2	1.319
26	26. Bangkala	4	.658
27	27. Sampang	4	.658
28	28. Pamekasa	3	1.787
29	29. Sumenep	3	1.674
30	30. Kota Ked	2	1.811
31	31. Kota Bli	1	.582
32	32. Kota Mal	6	.718
33	33. Kota Pro	1	.654
34	34. Kota Pas	1	.827
35	35. Kota Moj	1	1.074
36	36. Kota Mad	6	.779
37	37. Kota Sur	6	.195
38	38. Kota Bat	1	1.094

9) Melakukan analisis hasil.

Analisis

- a. *Cluster 1*, terdiri dari 5 kabupaten/kota dengan nilai X1, X2 dan X3 di atas rata-rata (jarak *cluster* bernilai positif) dan nilai X4, X5 dan X6 di bawah rata-rata (jarak *cluster* bernilai negatif). Kota tersebut adalah kota Blitar, kota Mojokerto, Kota Probolinggo, kota Pasuruan, dan kota Batu.
- b. *Cluster 2*, terdiri dari 5 kabupaten/kota dengan nilai X1, X2, X3 dan X5 di atas rata-rata dan nilai X4, dan X6 di bawah rata-rata. kabupaten/kota tersebut adalah kab. Mojokerto, Sidoarjo, Gresik dan kota Kediri.
- c. *Cluster 3*, terdiri dari 11 kabupaten/kota dengan nilai X6 di atas rata-rata dan nilai X1, X2, X3, X4 dan X5 di bawah rata-rata. kabupaten/kota tersebut adalah Kab. Lumajang, Bondowoso, Situbondo, Probolinggo, Pasuruan, Ngawi, Bojonegoro, Tuban, Lamongan, Pamekasan, dan Sumenep.
- d. *Cluster 4*, terdiri dari 2 kabupaten/kota dengan nilai X4 dan X6 di atas rata-rata dan nilai X1, X2, X3, dan X5 di bawah rata-rata. kabupaten/kota tersebut adalah Kab. Bangkalan dan Sampang
- e. *Cluster 5*, terdiri dari 13 kabupaten/kota dengan nilai X4 di atas rata-rata dan nilai X1, X2, X3, X5 dan X6 di bawah rata-rata. kabupaten/kota tersebut adalah Kab. Pacitan, Ponorogo, Trenggalek, Tuluagung, Blitar, Kediri, Malang, Jember, Banyuwangi, Jombang, Nganjuk, Madiun, dan Magetan.
- f. *Cluster 6*, terdiri dari 3 kabupaten/kota dengan nilai X1, X2, X3, dan X5 di atas rata-rata dan nilai X4 dan X6 di bawah rata-rata. kabupaten/kota tersebut adalah Kota Malang, Kota Madiun dan Kota Surabaya.

5. Kesimpulan

- 1) Penentuan banyaknya *cluster* tergantung dengan kebijakan yang akan dilakukan.
- 2) Untuk *cluster 3*, *cluster 4* dan *cluster 5* lebih banyak nilai variabel yang di bawah rata-rata. Sedangkan untuk *cluster 1*, *cluster 2* dan *cluster 6* lebih banyak nilai variabel di atas rata-rata.

- 3) Perlu diambil kebijakan pada *cluster* tertentu pada nilai variabel yang berada di bawah rata-rata sesuai dengan kebijakan yang strategis terhadap peningkatan IPM di Jawa Timur.

6. Referensi

- 1) BPS (2008), *Indeks Pembangunan Manusia 2006-2007*. BPS, Jakarta.
- 2) Nur, C. F. (2009), *Pemodelan IPM Provinsi Jawa Timur, Jawa Tengah, Jawa* Program Sarjana. Institut Teknologi Sepuluh Nopember, Surabaya.
- 3) Santoso, S., 2010, *Statistik Multivariat*, Graha Ilmu, Jogjakarta
- 4) Sharma, S., *Applied Multivariate Techniques*. John Willey and Son, Inc. New York.
- 5) Johnson, N. and D. Wichem, 1988, *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, NJ.
- 6) Pravitasari, A. A., (2006), *Sistem Pengambilan Keputusan (SPK) dengan Analisis Cluster untuk Membantu Pemerintah Daerah Kabupaten Tuban dalam Menentukan Kebijakan di bidang Pendidikan*. Tugas Akhir, Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember, Surabaya.

7) Lampiran

Data IPM Tiap Kabupaten di Jawa Timur dan Variabel – variabel yang Mempengaruhinya Tahun 2007

Kabupaten/Kota	IPM	X1	X2	X3	X4	X5	X6
01. Pacitan	70.48	12.94	35.17	195.16	0.55	4.18	23.31
02. Ponorogo	68.55	25.72	37.03	233.76	0.48	9.67	18.23
03. Trenggalek	71.68	24.22	32.84	224.93	0.53	8.47	22.79
04. Tulungagung	72	43.08	41.41	253.34	0.52	18.07	17.83
05. Blitar	72.28	25.4	34.91	276.45	0.55	2.45	16.47
06. Kediri	70.39	39.22	42.68	231.14	0.52	16.12	18.98
07. Malang	69.07	40.75	35.07	259.45	0.52	19.65	15.66
08. Lumajang	66.2	21.88	26.16	200.03	0.47	13.49	20.09
09. Jember	63.27	38.18	29.56	219.25	0.53	7.33	18.57
10. Banyuwangi	67.24	36.51	35.12	252.23	0.52	5.65	15.33
11. Bondowoso	60.76	23.68	26.9	208.7	0.43	9.4	24.23
12. Situbondo	62.64	38.69	31.61	247.52	0.44	9.39	15.6
13. Probolinggo	60.97	29.9	21.64	218.54	0.5	12.2	27.42
14. Pasuruan	65.52	40	29.16	225.27	0.49	32.01	19.88
15. Sidoarjo	74.87	87.95	58.56	345.59	0.43	49.13	13.05
16. Mojokerto	71.99	44.66	45.3	302.43	0.48	33.25	14.86
17. Jombang	71.44	58.2	42.32	273.79	0.54	12.3	21.21
18. Nganjuk	69.25	34.73	36.85	246.18	0.55	7.93	23.79
19. Madiun	68.24	24.6	38.14	231.75	0.52	4.1	20.98
20. Magetan	71.2	32.15	40.9	233.25	0.5	8.34	16.87
21. Ngawi	67.52	9.33	34.75	216.46	0.44	6.09	23.33
22. Bojonegoro	65.5	17.77	33.56	221.34	0.49	4.24	26.37
23. Tuban	66.61	19.77	31.23	193.13	0.46	20.48	28.51
24. Lamongan	67.88	17.5	39.37	242.61	0.49	5.36	25.79
25. Gresik	73	52.29	46.89	291.62	0.48	47.24	23.98
26. Bangkalan	62.97	21.8	18.6	196.14	0.66	4.25	31.56
27. Sampang	56.99	10.02	12.64	204.75	0.69	0.96	39.42
28. Pamekasan	62.49	19.77	28.69	191.62	0.55	1.15	32.43
29. Sumenep	63.71	14.27	26.49	203.53	0.4	2.48	32.98
30. Kota Kediri	74.45	100	58.27	307.02	0.47	72.64	13.67
31. Kota Blitar	75.88	100	58.8	341.8	0.47	12.8	12.02
32. Kota Malang	75.72	98.06	60.3	457.06	0.41	35.9	7.19
33. Kota Probolinggo	72.76	81.7	51.2	329.62	0.45	16.47	16.19
34. Kota Pasuruan	72.2	96.23	50.73	295.69	0.47	18.16	12.61
35. Kota Mojokerto	75.66	100	62.78	385.23	0.44	13.98	10.46
36. Kota Madiun	75.42	100	66.5	383.98	0.44	24.33	7.07
37. Kota Surabaya	75.87	100	62.07	418.27	0.41	30.19	7.98
38. Kota Batu	72.83	64.56	47.27	340.21	0.44	7.61	9.71