

# ANALISIS *CLUSTER* PENDIDIKAN FORMAL PENDUDUK INDONESIA

Nandang Rusmana, MT

Jurusan Teknik Manufaktur, Politeknik Manufaktur Negeri Bandung

Jln. Ir. H. Juanda Komplek Kanayakan Dago, Tromol Pos 851, Bandung 40008 Indonesia

Phone: (022)2500241, Fax : (022)2502649, E-mail : alamat\_email@polman-bandung.ac.id

## Abstrak

*Kemajuan suatu bangsa dapat diukur dari tingkat pendidikan yang dialami oleh masyarakatnya. Pendidikan yang diterima oleh masyarakat disediakan secara formal oleh pemerintah dan swasta. Pendidikan formal yang disediakan tersebut dimulai dari Sekolah Dasar(SD/MI), SMP/MTs, SMU/MA dan SMK, D1/D2, D3/Akademi, D4/S1, S2 dan S3. Tingkat pendidikan yang rendah dapat menyebabkan tertinggalnya pengetahuan suatu daerah dari daerah lainnya. Hal ini mengakibatkan harus dibedakan cara pendidikan yang dilakukan oleh lembaga pendidikan formal terhadap suatu daerah tertentu.*

*Pengolahan data statistik dilakukan menggunakan Analisis cluster terhadap pemerataan pendidikan yang didapatkan oleh masyarakat di seluruh provinsi yang ada sehingga dapat mengelompokkan tiap provinsi ke dalam suatu permasalahan yang sama. Dari hasil analisis, lembaga pendidikan formal dapat memanfaatkannya untuk mengelompokkan siswa dari daerah tertentu dengan pendidikan yang khusus.*

**Kata Kunci :** Pendidikan formal, Pemerataan Pendidikan, Analisis Cluster

## I. PENDAHULUAN

### Latar belakang

Jumlah sekolah dan perguruan tinggi negeri dan swasta tersebar di seluruh provinsi yang ada di Indonesia. Jumlah provinsi yang ada di Indonesia pada saat ini terdiri dari 33 propinsi. Hampir semua provinsi terdiri dari penduduk pedesaan dan perkotaan.

Berdasarkan hasil survei Biro Pusat Statistik (BPS) tahun 2007 yang lalu dihasilkan data yang menunjukkan jumlah penduduk yang telah mengalami pendidikan formal sampai pada tingkatan tertentu berdasarkan ijazah terakhir yang mereka miliki. Data yang dihasilkan berupa presentasi jumlah penduduk yang mengalami pendidikan formal pada tingkatan tertentu terhadap seluruh pendidikan formal yang ada. Tabel yang lengkap dapat dilihat pada tabel Welfare 2007[3] yang terdiri dari survei yang dilakukan di daerah pedesaan dan survei di perkotaan.

Tidak meratanya pendidikan formal yang diterima oleh sebagian penduduk dapat menyebabkan kesenjangan dalam ilmu pengetahuan. Siswa yang berasal dari daerah tertentu harus mendapatkan pendidikan yang lebih dibandingkan yang berasal dari daerah lain. Oleh karena itu dibutuhkan sekelompok siswa yang berasal dari daerah-daerah tertentu untuk memudahkan dalam penyampaian pendidikannya.

### Tujuan dan Manfaat Penelitian

Dari hasil data statistik yang dikumpulkan, dapat dibuat analisis terhadap pemerataan pendidikan yang didapatkan oleh masyarakat di seluruh provinsi yang ada. Tujuan dari analisis ini adalah:

1. Untuk mengetahui daerah mana saja yang masih mengalami ketertinggalan dalam masalah pendidikan formal.
2. Untuk mengetahui sejauh mana pemerataan pendidikan formal yang ada di seluruh provinsi Indonesia.
3. Untuk mengelompokkan suatu daerah pada permasalahan yang sama sehingga memudahkan dalam menyelesaikan persoalannya.
4. Untuk mengetahui karakteristik pendidikan formal penduduk di pedesaan dan perkotaan.

Dengan tujuan analisis yang ada, maka analisis yang digunakan adalah analisis *cluster*. Analisis *cluster* dapat mengelompokkan tiap provinsi ke dalam suatu permasalahan yang sama.

Manfaat yang didapatkan dari analisis ini adalah untuk memudahkan dalam memahami daerah mana saja yang harus mendapatkan perhatian pada tingkatan pendidikan yang sesuai, sehingga penyelesaian masalah yang ada bisa lebih sederhana. Lebih jauh lagi dapat dianalisis variabel-variabel yang

08 MAR 2012

mempengaruhi karakteristik pendidikan formal penduduk di suatu provinsi.

## II. SUMBER DATA

Sumber data yang digunakan adalah data yang berasal dari Biro Pusat Statistik (BPS) dengan judul Statistik Kesejahteraan Rakyat 2007 (*Welfare Statistics 2007*). Tabel yang digunakan dalam bentuk persentase penduduk umur sepuluh tahun ke atas menurut propinsi dan ijazah/STTB tertinggi yang dimiliki tahun 2007.

Variabel yang digunakan untuk analisis *cluster* ini terdiri dari 9 variabel yaitu:

- X1= jumlah penduduk tidak berijazah
- X2= jumlah penduduk berijazah SD/MI
- X3= jumlah penduduk berijazah SMP/MTs
- X4= jumlah penduduk berijazah SMU/MA
- X5= jumlah penduduk berijazah SMK
- X6= jumlah penduduk berijazah Diploma I/II
- X7= jumlah penduduk berijazah Diploma III/Akademi
- X8= jumlah penduduk berijazah Diploma IV/Universitas
- X9= jumlah penduduk berijazah S2/S3

Obyek penelitian yang dilakukan terhadap penduduk di seluruh provinsi di Indonesia terdiri dari 33 propinsi seperti terlihat pada Tabel 1.

Tabel 1. Propinsi di Indonesia

No.	Propinsi	No.	Propinsi
1	Nanggroe Aceh D.	18	Nusa Tenggara Barat
2	Sumatera Utara	19	Nusa Tenggara Timur
3	Sumatera Barat	20	Kalimantan Barat
4	R i a u	21	Kalimantan Tengah
5	J a m b i	22	Kalimantan Selatan
6	Sumatera Selatan	23	Kalimantan Timur
7	Bengkulu	24	Sulawesi Utara
8	Lampung	25	Sulawesi Tengah
9	Kep. Bangka Belitung	26	Sulawesi Selatan
10	Kep. Riau	27	Sulawesi Tenggara
11	DKI Jakarta	28	Gorontalo
12	Jawa Barat	29	Sulawesi Barat
13	Jawa Tengah	30	Maluku
14	DI Yogyakarta	31	Maluku Utara
15	Jawa Timur	32	Papua Barat
16	Banten	33	P a p u a
17	B a l i		

## III. PENGUJIAN ASUMSI

Data yang akan digunakan harus diuji dulu validitasnya. Oleh karena itu terdapat dua asumsi yang harus diujikan terhadap data yang ada. Adapun kedua asumsi tersebut adalah:

1. Asumsi data yang *representatif*  
Data yang digunakan merupakan data yang valid karena data diperoleh dari Biro Pusat Statistik (BPS) yang berdasarkan data penyusunan tabulasi susenas 225,18 juta jiwa dengan jumlah sample 285.904 rumah tangga.
2. Asumsi data yang *multikolinearitas*  
Untuk menguji data yang *multikolinearitas* yaitu data yang bukan merupakan kombinasi linear antara satu variabel dengan variabel yang lainnya menggunakan pengujian korelasi Pearson karena datanya berupa data metrik.

Hasil yang didapatkan berupa matrik korelasi antara variabel jumlah penduduk di perkotaan berupa Index Korelasi Pearson.

Dari analisis bivariate yang dihasilkan, X3(SLTP/MTs) dan X7(D3/Akademi) menghasilkan signifikansi (0,993) yang terbesar diantara variabel yang lainnya. Hal ini menyatakan bahwa kedua variabel tersebut cenderung saling independent, artinya korelasi antara keduanya sangat kecil ( $=0,002$ ).

Dengan demikian variabel yang akan digunakan pada analisis *cluster* adalah kedua variabel tersebut, sehingga asumsi *multikolinearitas* dapat terpenuhi.

## IV. PENENTUAN JUMLAH CLUSTER

Penentuan jumlah *cluster* dilakukan dengan metoda *Hierarchical*. Hal ini dilakukan karena belum diketahui jumlah *cluster* yang akan buat. Variabel yang dijadikan dasar pengelompokkan adalah X3 dan X7.

Beberapa hal lain yang harus diperhatikan adalah:

- Untuk menentukan jumlah *cluster* dengan penjelasan terbesar digunakan *Agglomeration Schedule* sehingga penentuan jumlah *cluster* akan dilihat melalui tabel *aglomerasi* yang dihasilkan.
- Akan lebih mudah melihat data yang dikelompokkan melalui *dendogram*.
- Metode yang digunakan dalam pembentukan kelompok adalah metode Ward yang didasarkan pada hilangnya informasi akibat penggabungan obyek menjadi kelompok.
- Pengukuran jarak yang digunakan adalah jarak kuadrat *euclidean* sehingga diperoleh jarak yang tidak dipengaruhi oleh perbedaan tanda +/- antar obyek data.

- Standarisasi data menggunakan *Z scores* karena variabel yang dianalisis memiliki interval nilai yang berbeda.

Hasil yang didapatkan dari tabel *Agglomeration Schedule* terdiri dari 32 stage. Pada stage terakhir tersebut menyatakan jumlah *cluster* hanya satu buah. Semakin mengecil stage yang ada maka jumlah *cluster* akan semakin membesar. Pada stage pertama terlihat obyek 5 bergabung dengan obyek 7 karena mempunyai koefisien jarak yang paling rendah. Obyek-obyek lain pada saat itu belum mempunyai pasangan sehingga jumlah *cluster* pada saat itu ada 32 *cluster*.

Pendeteksian *Outlier* dapat dilihat pada tabel *dendogram*. *Outlier* terlihat ketika ada obyek yang baru bergabung pada stage akhir. Dari *dendogram* yang dihasilkan tidak terdapat *outlier*.

Jumlah *Cluster* yang dipilih berdasarkan pada tabel *agglomeration*. Terdapat beberapa kemungkinan jumlah *cluster* yang akan dipilih. Hal ini dapat dilihat pada Tabel 2.

Tabel 2. Selisih *Agglomeration Coefisient*

Jumlah Cluster	Selisih <i>Agglomeratio Coefisient</i>
1	0.00
2	20.12
3	16.95
4	10.08
5	3.46
6	3.29
7	2.57
8	1.26
9	1.11
10	0.82

Pada stage terakhir dari tabel 3 selisih *agglomeration coefisient* adalah nol dengan jumlah *cluster* hanya satu. Apabila jumlah *cluster* dua maka selisih *agglomeration coefisient* dengan stage sebelumnya adalah 20,12. Begitu pula selisih stage 31 dan 30 adalah 16,95. Setelah itu selisihnya terus mengecil sehingga jumlah *cluster* yang dapat digunakan adalah 2 atau 3 *cluster* saja sesuai dengan perubahan *coefisient* yang cukup besar.

Analisis dengan menggunakan *hierarchical cluster* dilakukan untuk jumlah *cluster* sebanyak 2 dan 3 *cluster* saja. Hasil yang didapat dibandingkan dengan metoda *K-Means* untuk jumlah *cluster* sebanyak 2 dan 3 *cluster*.

Dari hasil perbandingan dua metoda *hierarchical* dan *K-Means* untuk 2 kelompok dan 3

kelompok, didapatkan pengelompokkan dengan 2 kelompok menghasilkan lebih banyak obyek yang sama, sehingga dipilih dua kelompok yang dianggap valid.

Menggunakan *Hierarchical cluster* untuk dua kelompok, maka didapatkan dua kelompok yaitu *Cluster 1* penduduk perkotaan dan *Cluster 2* penduduk perkotaan.

Dari kedua *cluster* yang terbentuk, sebenarnya tidak begitu terlihat perbedaannya. hal ini dikarenakan data yang ada cenderung hampir sama. Jumlah penduduk perkotaan di tiap propinsi yang paling banyak memiliki ijazah tertinggi SMA/MA (X4) sebesar 22.95% berada di Claster 1. Untuk melihat lebih jauh perbedaan tiap *cluster*, maka pembagian jumlah *cluster* diperbanyak dengan melihat *dendogram* yang ada. Dari *dendogram* penduduk perkotaan, maka *cluster* dapat pula dikelompokkan menjadi 6 kelompok. Berikut hasil pengelompokkan *cluster* berdasarkan *dendogram* yang ada.

Dengan memperbanyak jumlah kelompok pada *cluster*, maka terlihat pada *cluster 4* dan *cluster 5* penduduk yang lulus SD/MI lebih banyak dibandingkan dengan *cluster* yang lainnya. Selain *cluster 4* dan 5 lulusan SMU/MA merupakan jumlah penduduk terbanyak di perkotaan. Perhatian harus lebih ditujukan lagi pada *cluster 4* dimana jumlah lulusan SD/MI cukup besar (25,42%) dan jumlah penduduk yang tidak sekolah sebesar 24,95%. Adapun yang termasuk ke dalam *cluster* tersebut adalah provinsi; Gorontalo, Sulawesi Barat, Nusa Tenggara Barat dan Bali.

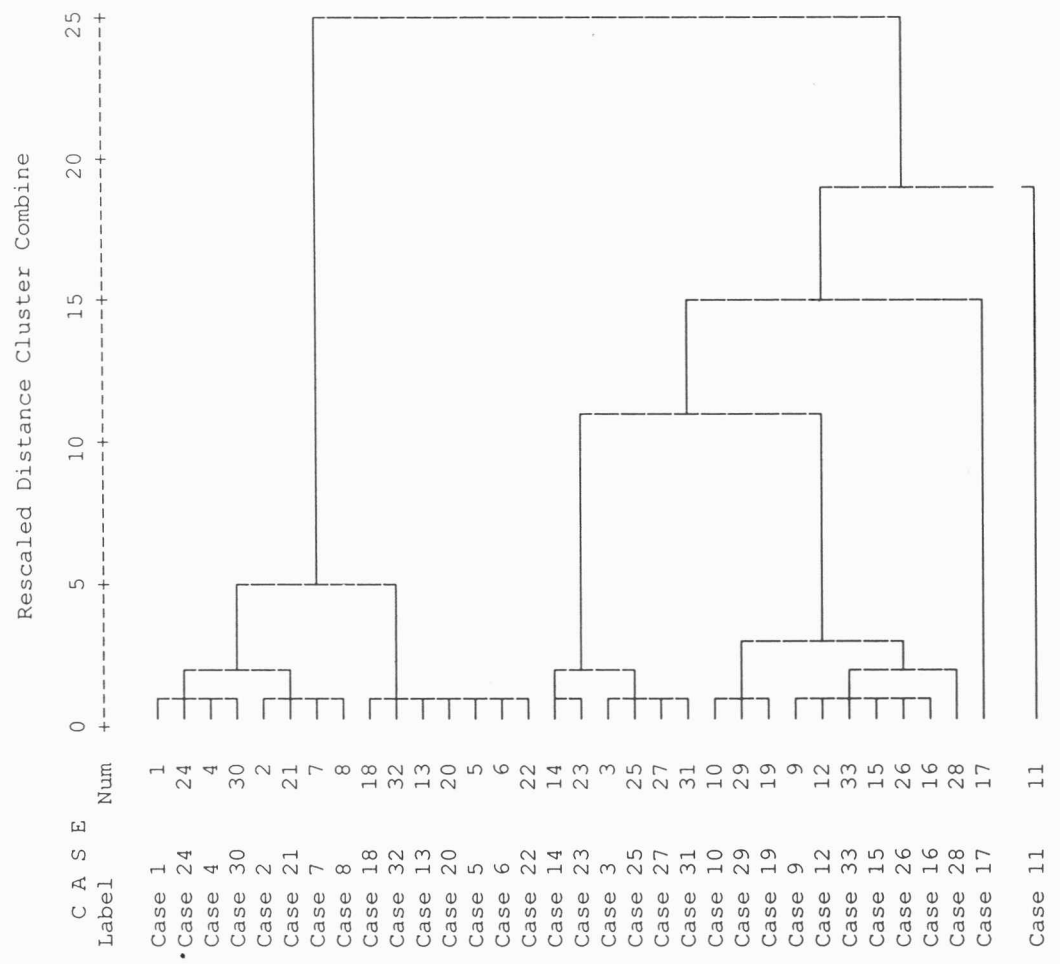
Pengolahan data yang sama dilakukan terhadap data pendidikan formal pada penduduk di daerah pedesaan. Hal ini dilakukan untuk melihat karakteristik yang terjadi di perkotaan dibandingkan dengan pedesaan.

Dari tabel analisis bivariate yang dihasilkan, X3(SLTP/MTs) dan X9(S2/S3) menghasilkan signifikansi (0,991) yang terbesar diantara variabel yang lainnya. Hal ini menyatakan bahwa kedua variabel tersebut cenderung saling independent, artinya korelasi antara keduanya sangat kecil ( $=0,002$ ). Dengan demikian variabel yang akan digunakan pada analisis *cluster* adalah kedua variabel tersebut, sehingga asumsi *multikolinearitas* dapat terpenuhi.

08 MAR 2012

HIERARCHICAL CLUSTER ANALYSIS

Dendrogram using Ward Method



Tabel 3 Jumlah *cluster* penduduk pedesaan

Jumlah <i>Cluster</i>	Selisih Agglomeratio Coefisient
1	0.00
2	19.09
3	13.91
4	10.90
5	7.91
6	3.62
7	1.74
8	1.52
9	1.50
10	0.90

Perbandingan banyaknya-kelompok pada penduduk pedesaan menggunakan dua metoda *hierarchical* dan *K-Means* untuk 2 kelompok dan 3 kelompok, didapatkan pengelompokkan dengan 2 kelompok menghasilkan lebih banyak obyek yang sama, sehingga dipilih dua kelompok yang dianggap valid.

Dari pengelompokkan penduduk pedesaan dengan menggunakan analisis *cluster*, maka terlihat dua kelompok yang berbeda. Pada *cluster* pertama presentase terbesar didominasi oleh jumlah penduduk yang lulus SD/MI. Hal ini bisa terlihat pada presentasi rata-rata  $X_2$  sebesar 33,82%. Provinsi yang termasuk ke dalam *cluster* ini ada 15 provinsi, yaitu;

Tabel 4. *Cluster* 1

No.	Propinsi	No.	Propinsi
1	Nanggroe Aceh D.	18	Nusa Tenggara Barat
2	Sumatera Utara	20	Kalimantan Barat
4	R i a u	21	Kalimantan Tengah
5	J a m b i	22	Kalimantan Selatan
6	Sumatera Selatan	24	Sulawesi Utara
7	Bengkulu	30	Maluku
8	Lampung	32	Papua Barat
13	Jawa Tengah		

Sedangkan pada *cluster* 2 terdiri dari 18 propinsi dimana presentase terbesar berada pada jumlah penduduk yang tidak sekolah. Provinsi yang termasuk ke dalam *cluster* ini ditunjukkan pada Tabel 5.

Walaupun demikian, perbedaan antara *cluster* 1 dan *cluster* 2 tidak terlalu jauh. Tabel di atas lebih mencerminkan perbedaan secara menyeluruh jumlah penduduk yang mendapatkan pendidikan. Untuk melihat perbedaan yang lebih besar lagi maka jumlah *cluster* dapat diperbanyak.

Tabel 5 *Cluster* 2

No.	Propinsi	No.	Propinsi
3	Sumatera Barat	19	Nusa Tenggara Timur
9	Kep. Bangka Belitung	23	Kalimantan Timur
10	Kep. Riau	25	Sulawesi Tengah
11	DKI Jakarta	26	Sulawesi Selatan
12	Jawa Barat	27	Sulawesi Tenggara
14	DI Yogyakarta	28	Gorontalo
15	Jawa Timur	29	Sulawesi Barat
16	Banten	31	Maluku Utara
17	B a l i	33	P a p u a

## V. KESIMPULAN

Dari analisis *cluster* yang dilakukan, maka dapat disimpulkan beberapa hal :

1. Terdapat perbedaan karakteristik pendidikan formal pada penduduk di pedesaan dan perkotaan. Jumlah penduduk di perkotaan paling banyak memiliki ijazah SMA/MA, sedangkan di pedesaan banyak penduduk tidak bersekolah dan hanya memiliki ijazah SD/MI saja.
2. Perbedaan *cluster* pada penduduk perkotaan tidak begitu terlihat karena datanya menyebar hampir sama. Sehingga dapat disimpulkan bahwa pendidikan di perkotaan pada masing-masing provinsi menyebar merata. Untuk melihat perbedaan yang lebih jelas, maka pengelompokkan *cluster* tidak cukup hanya dengan dua kelompok saja, tetapi dikelompokkan menjadi lebih banyak lagi.
3. Perbedaan *cluster* pada penduduk pedesaan terlihat pada penduduk yang tidak bersekolah dan penduduk yang hanya lulus SD. Perhatian pemerintah sebaiknya ditujukan pada pendidikan SD pada propinsi yang berada di *cluster* 2 karena masih banyak penduduknya yang tidak bersekolah (35,46%), sedangkan pada *cluster* 1 perhatian pemerintah lebih ditujukan pada pendidikan setelah SD yaitu SMP.
4. Dari analisis yang didapatkan, survei antara penduduk pedesaan dan perkotaan harus dibedakan, karena keduanya mempunyai pola karakteristik pendidikan yang berbeda.

## DAFTAR PUSTAKA

1. Laboratorium Perencanaan dan Optimasi Sistem Industri (2008), Multivariate Analysis Training, ITB

2. Hair J. F. et al. (1998), Multivariate Data Analysis, Prentice Hall
3. Biro Pusat Statistik (2007), Statistik Kesejahteraan Rakyat 2007 (Welfare Statistics 2007)