

**PENGELOMPOKAN PROVINSI DI INDONESIA  
BERDASARKAN KARAKTERISTIK KESEJAHTERAAN  
RAKYAT MENGGUNAKAN METODE *K-MEANS CLUSTER***



**SKRIPSI**

**Disusun oleh:**

**FITRA RAMDHANI**

**NIM 24010210141044**

**JURUSAN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG**

**2015**

**PENGELOMPOKAN PROVINSI DI INDONESIA  
BERDASARKAN KARAKTERISTIK KESEJAHTERAAN  
RAKYAT MENGGUNAKAN METODE *K-MEANS CLUSTER***

**Disusun oleh:  
FITRA RAMDHANI  
NIM 24010210141044**

Diajukan sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Statistika pada Jurusan Statistika

**JURUSAN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG**

**2015**

## HALAMAN PENGESAHAN I

Nama : Fitra Ramdhani  
NIM : 24010210141044  
Fakultas : Sains dan Matematika  
Jurusan : Statistika  
Judul Laporan : Pengelompokan Provinsi di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode *K-Means Cluster*

Telah diujikan pada sidang Tugas Akhir tanggal 20 Agustus 2015 dan dinyatakan lulus pada tanggal 4 September 2015.

Semarang, 7 September 2015

Mengetahui,  
Ketua Jurusan Statistika FSM  
Undip



Dra. Dwi Ispriyanti, M.Si  
NIP. 195709141986032001

Panitia Penguji Ujian Tugas  
Akhir  
Ketua,

A handwritten signature in black ink, likely belonging to the exam committee chair.

Drs. Sudarno, M.Si  
NIP. 196407091992011001

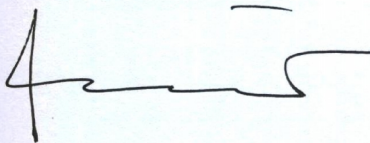
## HALAMAN PENGESAHAN II

Nama : Fitra Ramdhani  
NIM : 24010210141044  
Fakultas : Sains dan Matematika  
Jurusan : Statistika  
Judul Laporan : Pengelompokan Provinsi di Indonesia Berdasarkan  
Karakteristik Kesejahteraan Rakyat Menggunakan  
Metode *K-Means Cluster*

Telah disidangkan pada tanggal 20 Agustus 2015.

Semarang, 7 September 2015

Pembimbing I



Abdul Hoyyi, S.Si, M.Si  
NIP. 197202022008011018

Pembimbing II



Moch. Abdul Mukid, S.Si, M.Si  
NIP. 197808172005011001

## KATA PENGANTAR

Syukur Alhamdulillah, segala puja dan puji hanya bagi Allah SWT yang telah melimpahkan karunia-Nya, karena hanya dengan rahmat dan hidayah-Nya penulis dapat menyelesaikan penulisan Tugas Akhir ini dengan judul “Pengelompokan Provinsi di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode *K-Means Cluster*”.

Penulis menyadari bahwa di dalam penyusunan Tugas Akhir ini masih banyak terdapat kekurangan. Penulis sangat berharap setiap kekurangan yang ada dapat diperbaiki sehingga akan dapat memberikan manfaat bagi kita semua.

Tugas Akhir ini dapat diselesaikan karena bantuan beberapa pihak. Pada kesempatan ini penulis menyampaikan terima kasih kepada:

1. Ibu Dra. Dwi Ispriyanti, M.Si, selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Bapak Abdul Hoyyi, S.Si, M.Si selaku dosen pembimbing I dan Bapak Moch. Abdul Mukid, S.Si, M.Si selaku dosen pembimbing II.
3. Bapak dan Ibu dosen Jurusan Statistika FSM Undip.
4. Semua pihak yang tidak dapat penulis sebutkan satu persatu.

Penulis sadar bahwa penulisan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, saran dan kritik dari berbagai pihak sangat penulis harapkan. Besar harapan penulis semoga Tugas Akhir ini dapat bermanfaat bagi semua pihak.

Semarang, 7 September 2015

Penulis

## ABSTRAK

Kesejahteraan mengandung pengertian yang relatif, dinamis, dan kuantitatif. Rumusan kuantitatif dari kesejahteraan tidak pernah final karena akan terus berkembang seiring dengan perkembangan kebutuhan hidup manusia. Pada tahun 2011 Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) membuat bidang prioritas yang dapat dijadikan sebagai tolak ukur kesejahteraan suatu daerah. Dari bidang prioritas tersebut hendak dibuat klaster yang berisi 33 provinsi berdasarkan tingkat kesejahteraan rakyat di daerah tersebut menggunakan data tahun 2012 yang bersumber dari Badan Pusat Statistik (BPS). Metode yang digunakan untuk mengelompokkan ke-33 provinsi tersebut adalah metode *K-Means Cluster* dengan jumlah klaster sebanyak dua, tiga, empat, dan lima klaster. Metode *K-Means Cluster* merupakan salah satu metode analisis klaster yang dapat mempartisi data ke dalam bentuk satu atau lebih klaster, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu klaster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam klaster yang lain. Untuk mengetahui jumlah klaster yang optimal digunakan *Davies-Bouldin Index* (DBI). Diperoleh kesimpulan bahwa jumlah klaster yang optimal adalah tiga klaster dengan rincian provinsi pada klaster satu adalah provinsi yang memiliki keunggulan pada bidang APM SD atau MI, APM SMP atau MTs, AKB, dan akses listrik. Klaster dua merupakan kelompok provinsi yang memiliki keunggulan hanya pada satu bidang saja, yaitu TPT. Klaster tiga merupakan kelompok provinsi yang memiliki keunggulan di segala bidang.

**Kata Kunci:** Kesejahteraan, Bidang Prioritas TNP2K, Metode *K-Means Cluste*, *Davies-Bouldin Index* (DBI)

## ABSTRACT

Welfare have a relative explanation, dynamic, and quantitative. Quantitative formulation of welfare is never final because it will continue to evolve along with the development needs of human life. In 2011, the National Team for the Acceleration of Poverty Reduction (NTAPR) made priority sector that can serve as a benchmark the welfare in a region. From the priority sector will be made cluster or group which contains all 33 provinces based on the level of public welfare in the region uses data in 2012 were sourced from the Central Statistics Agency (CSA). The method that can be used to group the 33 provinces is K-Means Cluster method with number cluster as many as two, three, four, and five clusters. K-Means Cluster method is one of cluster analysis method who can partition the data into one or more clusters, so that the data with the same characteristics are grouped into the same cluster and data with different characteristics grouped into other clusters. To know the most optimal of the number of clusters we use Davies-Bouldin Index (DBI). We concluded that the optimal number of cluster is three with details the province in the first clusters have superiority in four sectors like net enrollment rate of primary school, net enrollment rate of junior high school, IMR (Infant Mortality Rate), and access to electricity. The province in the second clusters have superiority in one sector, that is open unemployment rate. The province in the third clusters have superiority in all sectors.

**Keywords:** Welfare, NTAPR Priority Sector, K-Means Cluster Method, Davies-Bouldin Index (DBI)

## DAFTAR ISI

	Halaman
HALAMAN JUDUL .....	i
HALAMAN PENGESAHAN I .....	ii
HALAMAN PENGESAHAN II .....	iii
KATA PENGANTAR .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
DAFTAR ISI .....	vii
DAFTAR GAMBAR .....	x
DAFTAR TABEL .....	xii
DAFTAR LAMPIRAN .....	xiii
 <b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan masalah .....	3
1.4 Tujuan .....	4
 <b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Hakikat dan Arti Kesejahteraan Rakyat .....	5
2.2 Indikator Kesejahteraan Rakyat .....	6
2.3 Pengertian Analisis Kluster .....	14
2.4 Tujuan Dasar .....	15

2.5	Asumsi pada Analisis Klaster .....	16
2.6	Metode Pengklasteran dalam Analisis Klaster .....	18
2.7	Metode <i>K-Means Cluster</i> .....	23
2.8	Memilih Ukuran Kemiripan .....	24
2.9	Jarak <i>Euclidean</i> .....	25
2.10	Standardisasi Data .....	26
2.11	Menentukan Jumlah Klaster .....	27
2.12	Menginterpretasi dan Memprofil Klaster .....	28
2.13	Kevalidan Hasil Klaster .....	28

### BAB III METODOLOGI PENELITIAN

3.1	Sumber Data .....	29
3.2	Variabel Penelitian .....	29
3.3	Metode Analisis Data .....	31

### BAB IV HASIL DAN PEMBAHASAN

4.1	Analisis Deskriptif .....	35
4.2	Pendeteksian Multikolinearitas .....	39
4.3	Proses Pengklasteran Metode <i>K-Means Cluster</i> untuk $k = 2$ .....	40
4.4	Proses Pengklasteran Metode <i>K-Means Cluster</i> untuk $k = 3$ .....	46
4.5	Proses Pengklasteran Metode <i>K-Means Cluster</i> untuk $k = 4$ dan $k = 5$ .....	52
4.6	Penentuan Jumlah Klaster .....	55
4.7	Interpretasi dan <i>Profiling</i> Hasil Klaster untuk $k = 3$ .....	60
4.8	Pemeriksaan Kevalidan Hasil Klaster .....	63

BAB V PENUTUP	
5.1 Kesimpulan .....	64
5.2 Saran .....	65
DAFTAR PUSTAKA .....	67
LAMPIRAN .....	69

## DAFTAR GAMBAR

	Halaman
Gambar 1 Angka Partisipasi Murni SD atau MI dan SMP atau MTs Tahun 2003-2012 di Indonesia .....	8
Gambar 2 Angka Kematian Bayi per 1.000 Kelahiran Hidup Tahun 2000-2012 di Indonesia .....	9
Gambar 3 Persentase Rumah Tangga Menurut Daerah dengan Sumber Air Bersih Tahun 2003-2012 di Indonesia .....	10
Gambar 4 Persentase Rumah Tangga Menurut Daerah yang Memiliki Akses Listrik Tahun 2009-2012 .....	11
Gambar 5 Persentase Tingkat Pengangguran Terbuka Tahun 2006-2012 di Indonesia .....	13
Gambar 6 Ilustrasi <i>Single Linkage</i> .....	19
Gambar 7 Ilustrasi <i>Complete Linkage</i> .....	20
Gambar 8 Ilustrasi <i>Average Linkage</i> .....	20
Gambar 9 Klasifikasi Prosedur Pengklasteran .....	22
Gambar 10 Ilustrasi Jarak <i>Euclidean</i> dari Teorema Pitagoras .....	25
Gambar 11 <i>Flowchart</i> Proses Algoritma <i>K-Means Cluster</i> .....	34
Gambar 12 Grafik APM SD atau MI dari 33 Provinsi di Indonesia Tahun 2012 .....	36
Gambar 13 Grafik APM SMP atau MTs dari 33 Provinsi di Indonesia Tahun 2012 .....	36
Gambar 14 Grafik AKB dari 33 Provinsi di Indonesia Tahun 2012 .....	37

Gambar 15 Grafik Persentase Rumah Tangga yang Memiliki Akses Air Bersih dari 33 Provinsi di Indonesia Tahun 2012 .....	37
Gambar 16 Grafik Persentase Rumah Tangga yang Memiliki Akses Listrik dari 33 Provinsi di Indonesia Tahun 2012 .....	38
Gambar 17 Grafik Persentase Tingkat Pengangguran Terbuka dari 33 Provinsi di Indonesia Tahun 2012 .....	39

## DAFTAR TABEL

	Halaman
Tabel 1 Susunan Observasi Analisis Klaster .....	15
Tabel 2 Statistik Deskriptif .....	35
Tabel 3 Nilai VIF dari Enam Variabel Penelitian .....	40
Tabel 4 Nilai <i>Centroid</i> untuk $k = 2$ .....	43
Tabel 5 Hasil Jarak <i>Euclidean</i> pada Setiap Iterasi untuk $k = 2$ .....	45
Tabel 6 Anggota Klaster dari Setiap Provinsi untuk $k = 2$ .....	46
Tabel 7 Nilai <i>Centroid</i> Awal untuk $k = 3$ .....	47
Tabel 8 Nilai <i>Centroid</i> Satu untuk $k = 3$ .....	48
Tabel 9 Hasil Jarak <i>Euclidean</i> untuk $k = 3$ pada Iterasi Satu dan Iterasi Dua .....	50
Tabel 10 Perubahan Posisi Klaster di Setiap Iterasi .....	51
Tabel 11 Anggota Klaster dari Setiap Provinsi untuk $k = 3$ .....	52
Tabel 12 Nilai <i>Centroid</i> Awal untuk $k = 4$ .....	53
Tabel 13 Anggota Klaster dari Setiap Provinsi untuk $k = 4$ .....	53
Tabel 14 Nilai <i>Centroid</i> Awal untuk $k = 5$ .....	54
Tabel 15 Anggota Klaster dari Setiap Provinsi untuk $k = 5$ .....	55
Tabel 16 Nilai DBI untuk $k = 2, 3, 4,$ dan $5$ .....	60
Tabel 17 Rata-rata Setiap Klaster Setelah Proses Standardisasi .....	61
Tabel 18 Rata-rata Setiap Klaster .....	61

## DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Data Enam Bidang Prioritas Kesejahteraan Rakyat ..	69
Lampiran 2 Data Enam Bidang Prioritas Kesejahteraan Rakyat yang Telah Melalui Proses Standardisasi ..	70
Lampiran 3 Tabel <i>Model Summary</i> untuk Melihat Nilai $R^2$ dari Enam Variabel Penelitian ..	71
Lampiran 4 <i>Output K-Means Cluster</i> dari Minitab 14 untuk $k = 2$ ..	74
Lampiran 5 Nilai setiap <i>centroid</i> dengan Menggunakan Jarak <i>Euclidean</i> untuk $k = 3$ ..	76
Lampiran 6 Hasil Jarak <i>Euclidean</i> pada Setiap Iterasi untuk $k = 3$ ..	77
Lampiran 7 <i>Output K-Means Cluster</i> dari Minitab 14 dengan $k = 3$ ..	80
Lampiran 8 <i>Output K-Means Cluster</i> dari Minitab 14 dengan $k = 4$ ..	82
Lampiran 9 <i>Output K-Means Cluster</i> dari Minitab 14 dengan $k = 5$ ..	84
Lampiran 10 Nilai setiap <i>centroid</i> dengan Menggunakan Jarak <i>Manhattan</i> untuk $k = 3$ ..	86
Lampiran 11 Hasil Jarak <i>Manhattan</i> pada Setiap Iterasi untuk $k = 3$ ..	87
Lampiran 12 Hasil Klaster dari Minitab 14, <i>Ms. Excel</i> dengan Jarak <i>Euclidean</i> , dan <i>Ms. Excel</i> dengan Jarak <i>Manhattan</i> ..	90

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pada tahun 2000, negara-negara yang tergabung di dalam Perserikatan Bangsa-Bangsa (PBB) mendeklarasikan kesepakatan Tujuan Pembangunan Milenium (*Millennium Development Goals/MDG's*) yang tertuang dalam delapan butir tahapan. Delapan tahapan tersebut adalah (1) menanggulangi kemiskinan dan kelaparan, (2) mencapai pendidikan dasar untuk semua, (3) mendorong kesetaraan gender dan pemberdayaan perempuan, (4) menurunkan angka kematian anak, (5) meningkatkan kesehatan ibu, (6) memerangi HIV/AIDS, malaria, dan penyakit menular lainnya, (7) memastikan kelestarian lingkungan hidup dan (8) mengembangkan kemitraan global untuk pembangunan. Delapan butir tersebut dibuat sebagai tahapan untuk mencapai kesejahteraan dan pembangunan masyarakat pada tahun 2015 (Bappeda Kabupaten Banjar, 2013).

Kesejahteraan rakyat sendiri sesungguhnya merupakan kondisi yang bentuknya dinamis. Rumusan kuantitatifnya tidak pernah final karena akan terus berkembang seiring dengan perkembangan kebutuhan hidup manusia. Oleh karena itu, seandainya kesejahteraan rakyat dalam pengertiannya yang hakiki belum dapat sepenuhnya diwujudkan, atau kesejahteraan rakyat belum bisa dicapai oleh kalangan masyarakat tertentu, pembangunan nasional harus tetap ditingkatkan untuk menciptakan kesejahteraan rakyat (Roestam, 1993).

Berbagai program telah dilaksanakan pemerintah guna mencapai kesejahteraan yang maksimal. Akan tetapi, kesejahteraan memiliki dimensi yang sangat luas dan kompleks sehingga membuat tarafnya hanya dapat dinilai

berdasarkan indikator-indikator yang terukur dari berbagai aspek pembangunan. Hal ini membuat tinggi rendahnya tingkat kesejahteraan di suatu daerah tidak hanya ditentukan oleh besaran pendapatan perkapita daerah tersebut, melainkan terdapat berbagai faktor lainnya (Bappeda Kabupaten Banjar, 2013).

Tidak dapat dipungkiri bahwa kemiskinan memang merupakan penyebab utama kesejahteraan (Haughton dan Khandker, 2012). Namun, terdapat beberapa bidang prioritas yang telah Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) buat sebagai tolak ukur kesejahteraan rakyat secara keseluruhan disamping bidang lainnya. Bidang-bidang tersebut terbagi menjadi empat indikator, yaitu pendidikan, kesehatan, perumahan, dan ketenagakerjaan.

Pada penulisan Tugas Akhir ini dilakukan suatu penelitian mengenai pengelompokan seluruh provinsi di Indonesia berdasarkan karakteristik kesejahteraan rakyat dari enam bidang prioritas yang telah dibuat oleh TNP2K. Enam bidang tersebut selanjutnya dijadikan sebagai variabel penelitian yang datanya merupakan data pada tahun 2012 yang berlaku di setiap provinsi. Data yang digunakan bersumber dari sensus Badan Pusat Statistik (BPS).

Sebelum proses pengklasteran dilakukan, variabel-variabel yang digunakan harus terbebas dari pengaruh multikolinearitas. Multikolinearitas adalah adanya hubungan yang linier di antara variabel penelitian. Apabila sudah dapat dipastikan bahwa variabel penelitian tidak memiliki masalah multikolinearitas, maka proses pengklasteran dapat dilakukan (Hair, *et al.*, 2006).

Proses pengklasteran dilakukan dengan menggunakan metode *K-Means Cluster* yang merupakan salah satu metode dalam Analisis Klaster dengan menggunakan jumlah klaster sebanyak dua, tiga, empat, dan lima. Jumlah klaster

yang optimal kemudian diperiksa menggunakan *Davies-Bouldin Index* (DBI). Jumlah kluster yang dipilih berdasarkan jumlah kluster yang memiliki nilai DBI terkecil (Permatadevi, *et al.*, 2013). Analisis kluster sendiri merupakan suatu metode pengelompokan yang didasarkan atas kesamaan atau kemiripan objek. Objek diklasifikasikan ke dalam satu atau lebih kluster sehingga objek-objek yang berada di dalam kluster akan mempunyai kemiripan karakter (Hair, *et al.*, 2006).

## 1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang di atas, maka dirumuskan pokok permasalahan yang akan menjadi kajian pada tugas akhir ini. Rumusan masalah tersebut adalah sebagai berikut:

1. Bagaimana menentukan jumlah kluster terbaik jika jumlah kluster yang digunakan adalah dua, tiga, empat, dan lima kluster?
2. Bagaimana menentukan anggota kluster dari 33 provinsi di Indonesia dengan menggunakan metode *K-Means Cluster* berdasarkan jumlah kluster terbaik?
3. Solusi apa yang dapat diberikan terhadap provinsi-provinsi yang ada pada setiap kluster terkait kondisi kesejahteraan rakyat?

## 1.3 Batasan Masalah

Permasalahan yang penulis angkat dibatasi pada pengelompokan provinsi berdasarkan data tahun 2012 dengan menggunakan metode *K-Means Cluster*. Pengelompokan didasarkan atas enam bidang prioritas kesejahteraan rakyat yang telah dibuat oleh TNP2K, yaitu: (1) angka partisipasi murni SD/MI, (2) angka partisipasi murni SMP/MTs, (3) angka kematian bayi per 1.000 kelahiran hidup,

(4) akses terhadap air bersih, (5) akses terhadap listrik, dan (6) tingkat pengangguran terbuka.

#### **1.4 Tujuan**

Berdasarkan uraian dari rumusan masalah di atas, maka dibuat tujuan yang hendak dicapai pada penelitian tugas akhir ini. Tujuan tersebut adalah sebagai berikut:

1. Menentukan jumlah klaster terbaik jika jumlah klaster yang digunakan adalah dua, tiga, empat, dan lima klaster.
2. Menentukan anggota klaster dari 33 provinsi di Indonesia dengan menggunakan metode *K-Means Cluster* berdasarkan jumlah klaster terbaik.
3. Menentukan solusi yang dapat diberikan terhadap provinsi-provinsi yang ada pada setiap klaster terkait kondisi kesejahteraan rakyat.