

**PEMILIHAN *CLUSTER* OPTIMUM PADA *FUZZY C-MEANS*
(Studi kasus: Pengelompokan Kabupaten/Kota di Jawa Tengah
berdasarkan Indikator Indeks Pembangunan Manusia)**



SKRIPSI

Disusun Oleh:

SARITA BUDIYANI PURNAMASARI

NIM. 24010210130073

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

2014

PEMILIHAN *CLUSTER* OPTIMUM PADA *FUZZY C-MEANS*
(Studi kasus: Pengelompokan Kabupaten/Kota di Jawa Tengah
berdasarkan Indikator Indeks Pembangunan Manusia)

Oleh:

SARITA BUDIYANI PURNAMASARI

NIM. 24010210130073

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Sains pada Jurusan Statistika

JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG

2014

HALAMAN PENGESAHAN I

Judul Skripsi : Pemilihan *Cluster* Optimum pada *Fuzzy C-Means*
(Studi kasus: Pengelompokan Kabupaten/Kota di Jawa Tengah berdasarkan Indikator Indeks Pembangunan Manusia)

Nama Mahasiswa : Sarita Budiyan Purnamasari

NIM : 24010210130073

Jurusan : Statistika

telah diujikan pada sidang Tugas Akhir tanggal 9 Juni 2014 dan dinyatakan lulus pada tanggal 18 Juni 2014.

Semarang, 19 Juni 2014

Mengetahui,

a.n. Ketua Jurusan Statistika

Sekretaris Jurusan Statistika

FSM Universitas Diponegoro



Drs. Agus Rusgiyono, M.Si

NIP. 196408131990011001

Panitia Penguji Ujian Tugas Akhir

Ketua



Dra. Hj. Dwi Ispriyanti, M.Si

NIP. 195709141986032001

HALAMAN PENGESAHAN II

Judul Skripsi : Pemilihan *Cluster* Optimum pada *Fuzzy C-Means*
(Studi kasus: Pengelompokan Kabupaten/Kota di Jawa
Tengah berdasarkan Indikator Indeks Pembangunan
Manusia)

Nama Mahasiswa : Sarita Budiyan Purnamasari

NIM : 24010210130073

Jurusan : Statistika

telah diujikan pada sidang Tugas Akhir tanggal 9 Juni 2014.

Semarang, 19 Juni 2014

Pembimbing I



Hasbi Yasin, M. Si.
NIP. 198212172006041003

Pembimbing II



Triastuti Wuryandari, S.Si, M.Si
NIP. 197109061998032001

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul **“Pemilihan Cluster Optimum pada Fuzzy C-Means (Studi kasus: Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah berdasarkan Indikator Indeks Pembangunan Manusia)”**.

Penulis menyadari bahwa penyusunan Tugas Akhir ini tidak akan berjalan dengan baik tanpa bantuan berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih kepada :

1. Ibu Dra. Dwi Ispriyanti, M.Si sebagai Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Bapak Hasbi Yasin, M.Si dan Ibu Triastuti Wuryandari, S.Si, M.Si selaku dosen pembimbing I dan dosen pembimbing II yang telah memberikan bimbingan dan pengarahan hingga terselesaikannya Tugas Akhir ini.
3. Bapak dan Ibu dosen Jurusan Statistika Universitas Diponegoro yang telah memberikan ilmu yang sangat bermanfaat.
4. Pihak-pihak yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa penulisan Tugas Akhir ini masih jauh dari sempurna. Sehingga penulis mengharapkan kritik dan saran demi kesempurnaan penulisan selanjutnya.

Semarang, Juni 2014

Penulis

ABSTRAK

Analisis *cluster* adalah proses pemisahan objek ke dalam kelompok, sehingga objek-objek yang tergabung dalam sebuah kelompok merupakan objek yang mirip satu sama lain dan berbeda dengan objek dalam kelompok lainnya. Salah satu metode dalam *clustering* adalah *Fuzzy C-Means Cluster* (FCM). FCM digunakan karena keberadaan tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan yang bernilai antara 0 sampai dengan 1. Penelitian ini menggunakan dua jenis jarak yaitu Manhattan dan Euclidean. Untuk menentukan jarak yang tepat pada kasus pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator Indeks Pembangunan Manusia (IPM) digunakan rasio simpangan baku, dimana nilai yang lebih kecil menunjukkan pengelompokan yang lebih baik. Sedangkan jumlah *cluster* optimum diperoleh dari Nilai Xie Beni yang minimum. Variabel-variabel yang digunakan merupakan indikator IPM tahun 2012 yang terdiri dari: Angka Harapan Hidup-AHH (tahun); Angka Melek Huruf-AMH (persen); Rata-rata Lamanya Sekolah-RLS (tahun) dan Paritas Daya Beli-PPP (ribuan rupiah). Hasil yang didapatkan dari penelitian adalah jarak Euclidean memberikan kualitas ketepatan pengelompokan lebih baik dan jumlah *cluster* optimum diberikan saat jumlah *cluster* adalah lima dengan nilai Xie Beni terkecil dibanding *cluster* lainnya, yakni 0,50778.

Kata kunci: analisis *cluster*, *Fuzzy C-Means* (FCM), IPM, *cluster* optimum.

ABSTRACT

Cluster analysis is a process of separating the objects into groups, so that the objects that belong to the same group are similar to each other and different from the other objects in another group. One method of clustering is Fuzzy C-Means (FCM). FCM is used because each data in a cluster determined by a degree of membership that have value between 0 and 1. This research use two kinds of distance, Manhattan and Euclidean. To determine the proper distance in clustering district / city in Central Java based on indicators of Human Development Index (HDI), we have to calculate the ratio of the standard deviation, where the smaller value indicates a better clustering. While the optimum number of groups obtained from the minimum value of Xie Beni. Variables that used in this research are the indicators of HDI in 2012 for district / city in Central Java, consists of: Life Expectancy Value (years), Literacy Rate (percent), Average Length of School (years), and Purchasing Power Parity (thousands rupiah). The results from this research are the distance that gives a better quality is Euclidean and the optimum cluster given when the number of cluster is five with the smallest value of Xie Beni is 0,50778.

Keywords: cluster analysis, Fuzzy C-Means (FCM), HDI, optimum cluster.

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 TABEL	ix
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Permasalahan	4
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
BAB II TINJAUAN PUSTAKA	
2.1 Konsep Pembangunan Manusia	6
2.2 Indeks Pembangunan Manusia	7
2.3 Indikator Indeks Pembangunan Manusia	8
2.4 Penyusunan IPM	9
2.5 Analisis <i>Cluster</i>	10
2.6 Himpunan <i>Crisp</i> dan Himpunan <i>Fuzzy</i>	13
2.7 Fungsi Keanggotaan	14
2.8 Konsep Jarak	15
2.9 <i>Fuzzy C-Means</i> (FCM)	17
2.10 Algoritma FCM	19
2.11 Formula Jarak Terbaik dan <i>Cluster Optimum</i>	20
BAB III METODOLOGI PENELITIAN	
3.1 Sumber Data dan Variabel Penelitian	23

3.2 Langkah-langkah Analisis Data	23
3.3 Diagram Alir Penelitian.....	25
BAB IV HASIL DAN PEMBAHASAN	
4.1 Analisis Deskriptif Variabel Penelitian	26
4.2 Analisis Hasil Pengelompokan dengan FCM.....	27
4.2.1 Pengelompokan FCM dengan Formula Jarak Manhattan	28
4.2.2 Pengelompokan FCM dengan Formula Jarak Euclidean	37
4.3 Pemilihan Formula Jarak Terbaik	45
4.4 Penentuan Jumlah <i>Cluster</i> Optimum.....	46
4.5 Interpretasi Hasil Pengelompokan dengan FCM.....	47
4.5.1 Hasil Pengelompokan.....	47
4.5.2 Karakteristik <i>Cluster</i>	48
BAB V PENUTUP	
5.1 Kesimpulan.....	52
5.2 Saran	53
DAFTAR PUSTAKA	54
LAMPIRAN.....	55

DAFTAR TABEL

	Halaman
Tabel 2.1 Nilai Maksimum dan Minimum Indikator IPM	9
Tabel 4.1 Statistik Deskriptif Indikator IPM Provinsi Jateng Tahun 2012...	26
Tabel 4.2 Derajat Keanggotaan Awal FCM Jarak Manhattan 2 <i>Cluster</i>	29
Tabel 4.3 Hasil Perhitungan Pusat <i>Cluster</i> 1 pada Iterasi Pertama FCM Jarak Manhattan	30
Tabel 4.4 Hasil Perhitungan Pusat <i>Cluster</i> 2 pada Iterasi Pertama FCM Jarak Manhattan	31
Tabel 4.5 Pusat <i>Cluster</i> FCM dengan Jarak Manhattan untuk 2 <i>Cluster</i> pada Iterasi Pertama	32
Tabel 4.6 Perhitungan Fungsi Objektif Iterasi Pertama FCM Jarak Manhattan	33
Tabel 4.7 Nilai Derajat Keanggotaan Baru Iterasi Pertama FCM Jarak Manhattan	34
Tabel 4.8 Pusat <i>Cluster</i> FCM dengan Jarak Manhattan untuk 2 <i>Cluster</i> pada Iterasi Terakhir	35
Tabel 4.9 Pusat <i>Cluster</i> FCM dengan Jarak Manhattan untuk 3 <i>Cluster</i> pada Iterasi Terakhir	36
Tabel 4.10 Pusat <i>Cluster</i> FCM dengan Jarak Manhattan untuk 4 <i>Cluster</i> pada Iterasi Terakhir	36
Tabel 4.11 Pusat <i>Cluster</i> FCM dengan Jarak Manhattan untuk 5 <i>Cluster</i> pada Iterasi Terakhir	36
Tabel 4.12 Derajat Keanggotaan Awal FCM Jarak Euclidean 2 <i>Cluster</i>	38

Tabel 4.13 Hasil Perhitungan Pusat <i>Cluster</i> 1 pada Iterasi Pertama FCM	
Jarak Euclidean	39
Tabel 4.14 Hasil Perhitungan Pusat <i>Cluster</i> 2 pada Iterasi Pertama FCM	
Jarak Euclidean	40
Tabel 4.15 Pusat <i>Cluster</i> FCM dengan Jarak Euclidean untuk 2 <i>Cluster</i>	
pada Iterasi Pertama	41
Tabel 4.16 Perhitungan Fungsi Objektif Iterasi Pertama FCM Jarak Euclidean	41
Tabel 4.17 Nilai Derajat Keanggotaan Baru Iterasi Pertama FCM	
Jarak Euclidean	42
Tabel 4.18 Pusat <i>Cluster</i> FCM dengan Jarak Euclidean untuk 2 <i>Cluster</i>	
pada Iterasi Terakhir	44
Tabel 4.19 Pusat <i>Cluster</i> FCM dengan Jarak Euclidean untuk 3 <i>Cluster</i>	
pada Iterasi Terakhir	44
Tabel 4.20 Pusat <i>Cluster</i> FCM dengan Jarak Euclidean untuk 4 <i>Cluster</i>	
pada Iterasi Terakhir	44
Tabel 4.21 Pusat <i>Cluster</i> FCM dengan Jarak Euclidean untuk 5 <i>Cluster</i>	
pada Iterasi Terakhir	45
Tabel 4.22 Hasil Perhitungan Rasio $S_w S_B$	45
Tabel 4.23 Nilai Xie Beni	46
Tabel 4.24 Karakteristik Hasil pengelompokkan Kabupaten/Kota di Jateng	
menggunakan FCM Jarak Euclidean.....	48

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Gambaran Umum Indek Pembangunan Manusia	7
Gambar 2.2 Himpunan <i>Fuzzy</i> Kelompok Umur	14
Gambar 2.3 Jarak Dua Data dalam Dua Dimensi.....	16
Gambar 3.1 Diagram Alir Penelitian.....	25

DAFTAR LAMPIRAN

	Halaman
Lampiran1. Nilai Indikator IPM Provinsi Jawa Tengah Tahun 2012.....	55
Lampiran2. Standarisasi Nilai Indikator IPM Provinsi Jawa Tengah Tahun 2012.....	56
Lampiran3. Syntax Program MATLAB	57
Lampiran4. Derajat Keanggotaan Awal dan Akhir FCM dengan Formula Jarak Manhattan.....	60
Lampiran5. Derajat Keanggotaan Awal dan Akhir FCM dengan Formula Jarak Euclidean	68
Lampiran6. Rata-rata Nilai Indikator IPM.....	76

BAB I

PENDAHULUAN

1.1 Latar Belakang

Gaung pemerintah dalam upaya mewujudkan cita-cita pembangunan nasional kian gencar dilakukan. Pembangunan nasional Indonesia menempatkan manusia sebagai titik sentral yang bercirikan dari rakyat, oleh rakyat, dan untuk rakyat. Dengan kata lain, rakyat dilibatkan dalam seluruh proses pembangunan, bukan hanya sebagai alat untuk mencapai hasil akhir pembangunan, tetapi sebagai tujuan akhir dari pembangunan itu sendiri. Untuk dapat ikut berpartisipasi dalam proses pembangunan, tentunya dibutuhkan masyarakat Indonesia yang tidak hanya unggul dari segi kuantitas, tetapi juga unggul pada segi kualitas (BPS, 2012)

Pengalaman selama krisis menunjukkan bahwa negara-negara yang mempunyai kualitas sumber daya manusia yang lebih baik, lebih cepat bangkit dari krisis yang melandanya, hal ini dapat dilihat pada negara-negara seperti Korea, Jepang dan Thailand (Yunitasari, 2007). Oleh karena itu dirumuskan suatu konsep baru dalam mengukur pembangunan suatu negara yang berorientasi pada manusia.

Tolok ukur keberhasilan pembangunan manusia telah dikembangkan oleh *United Nation Development Programe* (UNDP) yang dikenal dengan istilah *Human Development Index* (HDI) atau dikenal dengan Indeks Pembangunan Manusia (IPM) (BPS, 2012).

Upaya mewujudkan pembangunan manusia yang produktif diperlukan monitoring yang berkelanjutan. Pada tahun 2011, nilai IPM Indonesia berada pada peringkat 124 di dunia, dan pada tahun 2012 Indonesia naik tiga posisi menjadi peringkat 121. Hal ini didukung dengan peningkatan indikator pembentuknya, terutama pada angka harapan hidup dan rata-rata lama sekolah yang pada tahun 2012 mencapai 69,8 tahun dan 12,9 tahun (Hartono, 2013).

IPM mencakup tiga dimensi yang dianggap mendasar bagi manusia dan secara operasional mudah dihitung untuk menghasilkan suatu ukuran yang merefleksikan upaya pembangunan manusia. Adapun dimensi tersebut adalah peluang hidup (*longevity*), pengetahuan (*knowledge*) dan hidup layak (*living standards*). Peluang hidup dihitung berdasarkan angka harapan hidup ketika lahir (AHH), pengetahuan diukur berdasarkan rata-rata lama sekolah (RLS) dan angka melek huruf (AMH) penduduk berusia 15 tahun ke atas, dan hidup layak diukur dengan pengeluaran per-kapita yang didasarkan pada paritas daya beli (*purchasing power parity*) atau PPP (BPS, 2012).

Sejak 1 Januari 2001 pemerintah pusat memberikan kewenangan otonomi kepada pemerintah daerah. Masing-masing pemerintah daerah menyusun perencanaan pembangunan dan anggaran keuangan daerahnya, bukan hanya untuk meningkatkan pertumbuhan ekonomi, tetapi juga meningkatkan kesejahteraan masyarakat melalui pembangunan manusia yang meliputi sektor pendidikan, pelayanan kesehatan, dan kebijakan-kebijakan lain yang secara langsung memperbaiki kualitas hidup (Yunitasari, 2007). Dengan demikian daerah dapat meningkatkan kesejahteraan masyarakat melalui peningkatan

pembangunan manusia yang tercermin dari indikator Indeks Pembangunan Manusia (IPM).

IPM Provinsi Jawa Tengah pada tahun 2012 menempati peringkat 15 secara nasional dengan nilai 73,36. Angka ini naik 0,42 dari tahun sebelumnya yang berada di angka 72,94. Provinsi Jawa Tengah sendiri terdiri dari 29 kabupaten dan 6 kota, dengan IPM tertinggi di Kota Surakarta 78,60 dan IPM terendah di Kabupaten Brebes 69,37 (BPS, 2013).

Tinggi rendahnya IPM Kabupaten/Kota di Provinsi Jawa Tengah hanya ditunjukkan melalui indeks komposit, tetapi tidak ditunjukkan indikator mana yang dominan terhadap tinggi/rendahnya peringkat IPM. Padahal, nilai dari tiap-tiap indikator pembentuk IPM perlu dilihat agar dapat diketahui pencapaian dari tiap-tiap indikator. Pengelompokan wilayah kabupaten/kota di Jawa Tengah juga perlu dilakukan sebagai bahan perencanaan dan evaluasi sasaran program pemerintah untuk meningkatkan angka pembangunan manusia.

Pengelompokan wilayah bertujuan untuk membagi wilayah-wilayah ke dalam beberapa kelompok dengan karakteristik yang memiliki keserupaan tinggi di dalam setiap kelompok dan berbeda antar kelompok. Ada beberapa algoritma pengelompokan yang dapat digunakan, salah satu diantaranya adalah *Fuzzy C-Means*.

Fuzzy C-means Clustering (FCM) adalah suatu teknik pengelompokan data yang mana keberadaan tiap titik data dalam suatu kelompok ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Dunn pada tahun 1973. Konsep dasar FCM, pertama kali adalah menentukan pusat *cluster* yang akan menandai lokasi rata-rata untuk tiap-tiap *cluster*. Pada kondisi awal,

pusat *cluster* ini masih belum akurat. Tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat *cluster* akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimalisasi fungsi obyektif (Kusumadewi, 2002).

Dalam penelitian ini akan digunakan formula jarak Manhattan dan Euclidean sebagai formula jarak pada analisis FCM yang akan diterapkan untuk pengelompokan kabupaten/kota di Jawa Tengah berdasarkan indikator IPM tahun 2012. Kemudian akan dilakukan pemilihan jarak yang menghasilkan kualitas pengelompokan terbaik dengan memperhatikan nilai simpangan bakunya (Rasio S_w/S_B). Diharapkan nilai simpangan baku dalam kelompok akan bernilai minimum, sedang simpangan baku antar kelompok bernilai maksimum (Bunkers dan Miller, 1996). Penentuan jumlah *cluster* optimum pada jarak terpilih dilakukan dengan menggunakan perhitungan nilai validitas Xie Beni (Duo, et al, 2007).

1.2 Permasalahan

Berdasarkan latar belakang di atas, permasalahan yang akan dibahas dalam Tugas Akhir ini adalah sebagai berikut:

1. Bagaimana perbandingan hasil pengelompokan pada FCM dengan formula jarak Manhattan dan Euclidean untuk mengelompokkan kabupaten/kota di Jawa Tengah berdasarkan indikator IPM tahun 2012?
2. Berapakah jumlah *cluster* optimum berdasarkan perhitungan nilai validitas Xie Beni?

3. Bagaimana interpretasi hasil pengelompokan kabupaten/kota di Jawa Tengah berdasarkan indikator IPM?

1.3 Batasan Masalah

Penelitian ini dibatasi pada pengkajian metode *Fuzzy C-Means* dan penggunaannya untuk kasus pengelompokan kabupaten/kota di Jawa Tengah berdasarkan indikator IPM tahun 2012. Perbandingan hasil FCM dilakukan untuk formula jarak Manhattan dan Euclidean. Pengolahan data dilakukan menggunakan *software* MATLAB.

1.4 Tujuan Penelitian

Berdasarkan latar belakang yang telah dijabarkan sebelumnya, maka tujuan dari penulisan Tugas Akhir ini adalah sebagai berikut:

1. Membandingkan hasil pengelompokan pada FCM dengan formula jarak Manhattan dan Euclidean untuk mengelompokkan kabupaten/kota di Jawa Tengah berdasarkan indikator IPM tahun 2012
2. Mengetahui jumlah *cluster* optimum berdasarkan perhitungan nilai validitas Xie Beni
3. Melakukan interpretasi hasil pengelompokan kabupaten/kota di Jawa Tengah berdasarkan indikator IPM dimana hasil pengelompokan ini diharapkan mampu menjadi bahan pertimbangan bagi tiap-tiap daerah dalam pengambilan kebijakan yang berkenaan dengan peningkatan pembangunan manusia daerahnya.