

PENGELOMPOKAN KABUPATEN/KOTA DI JAWA TENGAH MENGUNAKAN METODE K-MEANS DAN FUZZY C-MEANS

Rahman Hidayat¹⁾, Rochdi Wasono²⁾, Moh. Yamin Darsyah³⁾

^{1,2,3}Program Studi Statistika MIPA, Unimus

¹email: hidayatrahman010@gmail.com

Abstract

The Poverty is still a serious problem, especially in Indonesia. Central Java is the province with the highest percentage of poor people in Java Island 13, 19%, the figure is above the national poverty rate. In this research will be an analysis of the factors affecting poverty in Central Java. The statistical approach used in this case is cluster analysis. Cluster analysis is an analysis that aims to classify an object (region) based on similarity characteristics of data. The method used is the method of K-Means and Fuzzy C-Means. The object of research is grouped into 4 clusters. The result of grouping shows that K-Means method is the best method based on SW and SB ratio of 0.124.

Keywords: *Data Mining, K-means, Fuzzy C-Means*

1. PENDAHULUAN

Persoalan kemiskinan yang bersifat multidimensi masih membutuhkan perhatian serius diberbagai negara, khususnya bagi negara-negara berkembang seperti Indonesia. Berdasarkan data yang dirilis oleh BPS, pada tahun 2016 di semester 2 persentase kemiskinan tercatat 10,7 % dengan jumlah penduduk yang hidup di bawah garis kemiskinan sebanyak 27,8 juta jiwa. Penduduk yang hidup dibawah garis kemiskinan sebagian besarnya berada di Pulau Jawa dengan persentase 53,4% atau 14,8 juta jiwa dari total angka kemiskinan nasional. Dari keenam provinsi di Pulau Jawa Provinsi Jawa Tengah merupakan daerah dengan persentase tertinggi sebesar 13,19 dengan jumlah penduduk miskin terbanyak kedua setelah Jawa Timur sebanyak 4.493.750 jiwa, angka ini melebihi persentase kemiskinan nasional dengan selisih 2,49 %. Menurut data BPS dari 35 Kabupaten/Kota yang ada di Jawa Tengah tercatat 15 kabupaten yang persentase penduduk miskinnya di atas angka Provinsi dan di atas angka nasional ada 24 kabupaten. Tingginya angka kemiskinan di Jawa Tengah tersebut perlu dilakukan upaya-upaya pengentasan kemiskinan yang berbasis pada data dengan memahami karakteristik dari masing-masing wilayah, sehingga upaya pengentasan yang dilakukan bisa efektif dan tepat. Maka perlu dilakukan identifikasi karakteristik dari masing-masing kabupaten/kota di Jawa Tengah sebagai salah satu penunjang keberhasilan program pengentasan kemiskinan.

Menurut Soemartini (2017), dalam melaksanakan program pembangunan perlu adanya identifikasi berdasarkan karakteristik tingkat kesejahteraan rakyat tiap daerah agar dalam mengambil kebijakan dan strategi pembangunan bisa tepat sasaran dan tepat guna. Sebab salah satu prasyarat keberhasilan program- program pembangunan sangat tergantung pada ketepatan pengidentifikasian target group dan target area. Oleh karenanya, pengentasan kasus kemiskinan di Jawa Tengah diperlukan suatu pendekatan analisis untuk mengidentifikasi karakteristik kemiskinan daerah berdasarkan indikator-indikator kemiskinan dengan pengelompokkan daerah. Salah satu metode yang dapat digunakan untuk mengelompokkan variabel atau objek adalah analisis kluster (*cluster analysis*) (soemartini, 2017). Dalam analisis kluster objek-objek dikelompokkan berdasarkan kemiripan/kesamaan objek. Prinsip dasar dalam analisis kluster adalah mengelompokkan objek pada suatu kluster (kelompok) yang memiliki kemiripan sangat besar dengan objek lain dalam kluster yang sama (*similarity*), tetapi sangat tidak mirip dengan objek lain pada kluster yang berbeda (*dissimilarity*). Hal ini bahwa kluster yang baik akan mempunyai homogenitas yang tinggi

antar anggota dalam satu kelompok dan heterogenitas yang tinggi antar kluster yang satu dengan yang lainnya (Rachmatin, 2014). Secara umum metode dalam analisis kluster dibagi dua yakni *Hierarchical Clustering* (metode hirarki) dan *Non-Hierarchical Clustering* (metode tak hirarki) atau *partitioning*. Metode hirarki jumlah kelompok yang akan dibentuk belum ditentukan. Prosedur kelompok hirarki terdiri atas dua bagian, yaitu metode (penggabungan) *agglomerative* dan metode (pemisah) *divisive*. Metode tak hirarki merupakan metode pengelompokan dimana kluster yang ingin dibentuk ditentukan terlebih dahulu, sehingga objek-objek akan dikelompokkan kedalam k kelompok yang telah ditentukan. Metode yang sering digunakan adalah K-Means dan Fuzzy C-Means. Penelitian terdahulu yang menggunakan analisis kluster diantaranya: Karti dan Irhamah (2013) mengelompokkan kabupaten/kota di Jawa timur berdasarkan indikator pendidikan dengan metode *C-Means* dan *Fuzzy C-Means*. Machfudhoh dan wahyuningsih (2013) yang mengelompokkan kabupaten/kota di Jawa Timur berdasarkan pertumbuhan ekonomi. Ningrat et.al (2016) yang melakukan pengelompokan data obligasi korporasi menggunakan metode *K-Means* dan *Fuzzy C-means*.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk mengelompokkan kabupaten/ kota di Jawa Tengah berdasarkan faktor-faktor yang mempengaruhi kemiskinan tahun 2016 menggunakan metode *K-means* dan *Fuzzy C-Means* (FCM) serta membandingkan hasil pengelompokan kedua metode tersebut. Pengelompokan metode *K-Means* dan FCM merupakan bagian dari aplikasi metode *data mining*. *Data Mining* merupakan metode analisis data yang memadukan berbagai bidang keilmuan seperti statistik, teknik, mesin learning dan lainnya. Hand et al. (2001) mendefinisikan sebagai sebuah analisa dari observasi data dalam jumlah yang besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan metode baru untuk meringkas data agar mudah dipahami serta kegunaannya bagi pemilik data. Data mining secara umum dibagi dua yakni *Supervised Learning* dan *Unsupervised Learning*. *Supervised Learning* merupakan pengklasifikasian data berdasarkan suatu label tertentu, sementara *Unsupervised Learning* data dikelompokkan tanpa mengikuti label tertentu. *Supervised Learning* memiliki beberapa metode seperti *Artificial Neural Network* (ANN), *Naive Bayes*, *SVM* dan lainnya. *Unsupervised Learning* berupa analisis kluster seperti *Ward*, *Single Linkage*, *K-Means*, *Fuzzy C-Means* dan lainnya.

Tan (2006) dalam Prasetyo (2012) mendefinisikan analisis kluster sebagai pekerjaan mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya. Analisis kluster merupakan metode analisis untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok, sehingga akan diperoleh kelompok dimana objek-objek dalam satu kelompok memiliki banyak persamaan sedangkan dengan anggota kelompok lain memiliki banyak perbedaan (Johnson dan Winchern, 2007). Jadi, tujuan pengelompokan adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (tidak berhubungan) dengan objek dalam kelompok yang lain. Lebih besar kemiripannya (homogenitas) dalam kelompok dan lebih besar perbedaannya diantara kelompok yang lain (Prasetyo, 2012).

Salah satu hal yang paling penting dalam mengidentifikasi kelompok pengamatan yang mungkin ada dalam data adalah pengetahuan tentang bagaimana individu 'dekat' satu sama lain, atau seberapa jauh jaraknya (Everitt, 2011). Ukuran jarak yang sering digunakan adalah ukuran jarak Euclidean. Jarak Euclidean merupakan jarak antar objek, misalkan dua objek ke-*i* dan ke-*j* yang berada pada *p* dimensi berikut (Johnson dan Winchen, 2007). Persamaanya dapat dituliskan sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (1)$$

Dimana:

d_{ij} = jarak antara objek ke-*i* dan objek ke-*j*

p = Jumlah variabel kluster

X_{ik} = data dari subjek ke- i pada variabel ke- k

X_{jk} = data dari subjek ke- j pada variabel ke- k

Selain penentuan jarak, dalam analisis kluster membutuhkan beberapa asumsi. Setidaknya ada dua asumsi dalam analisis kluster yaitu:

1. Kecukupan sampel
2. Tidak terjadi multikolinearitas

Jika terjadi pelanggaran asumsi pada poin 1 maka akan dilakukan penambahan sampel, untuk poin 2 akan diatasi dengan cara menghilangkan variabel yang saling berkorelasi atau dengan menggunakan *Principal Component Analysis (PCA)*. *Principal Component Analysis (PCA)* atau Analisis Komponen Utama (AKU) merupakan salah satu solusi dalam analisis kluster jika terjadi multikolinearitas dalam data. *PCA* bertujuan untuk mereduksi variabel menjadi lebih sedikit dari jumlah variabel sebelumnya dimana K variabel baru $< K$ variabel lama. *Principal component (PC)* merupakan suatu kombinasi linear dari variabel-variabel asal. Pembentukan *PC* berdasarkan dua cara yaitu matriks kovarian atau matriks korelasi (Johnson dan Wichern, 2007). Tahapan menentukan *PC* berdasarkan matriks korelasi adalah sebagai berikut:

1. Membuat matriks Z yang berisi data dari variabel X yang telah distandarisasi.
2. Membuat matriks korelasi dari Z yaitu $Z'Z$. Pereduksian *PC* dimulai dengan cara mencari nilai eigen $\lambda_1, \lambda_2, \dots, \lambda_p$ yang diperoleh dari persamaan:

$$|Z'Z - \lambda I| = 0 \quad (2)$$

Jumlah *Principal Component (PC)* yang terpilih berdasarkan nilai eigen (λ) dimana nilai $\lambda > 1$ maka *PC* tersebut akan dipilih (Hardika et al, 2013). Setelah asumsi terpenuhi baru bisa dilakukan analisis kluster.

Dalam statistik dan mesin pembelajaran, pengelompokan *K-Means* merupakan metode analisis kelompok yang mengarah pada pemartisan N objek pengamatan ke dalam K kelompok (*cluster*) dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan *mean* (rata-rata) terdekat.

K-Means merupakan salah satu metode pengelompokan data non hirarki (sekatan) yang berusaha mempartisi data yang ada kedalam kelompok sehingga data berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok lain (Prasetyo, 2012). Menurut MacQueen (1967), metode ini dimulai dengan menentukan jumlah cluster terlebih dahulu yang diinginkan misalkan c *cluster*. Langkah berikutnya adalah *K-Means*.

Pengelompokan data dengan metode *K-Means* secara umum dilakukan dengan algoritma sebagai berikut (Prasetyo, 2012):

1. Menentukan k sebagai jumlah kluster yang ingin dibentuk.
2. Mengalokasikan data secara acak.
3. Menentukan pusat kluster dari data yang ada pada masing-masing kluster dengan persamaan:

$$C_{kj} = \frac{x_{1j} + x_{2j} + \dots + x_{nj}}{n} \quad (3)$$

dimana:

C_{kj} = pusat kluster ke- k pada variabel ke- j ($j=1,2,\dots,p$)

n = banyak data pada kluster ke- k

4. Menentukan jarak setiap objek dengan setiap centroid dengan perhitungan jarak setiap objek dengan setiap centroid menggunakan jarak euclidean
5. Menghitung fungsi objektif

$$J = \sum_{i=1}^n \sum_{j=1}^k a_{ij} d(x_i, C_{kj})^2 \quad (4)$$

6. Mengalokasikan masing-masing data ke centroid/rata-rata terdekat yang dirumuskan.

$$a_{ij} = \begin{cases} 1, & s = \min \{d(x_i, C_{kj})\} \\ 0, & \text{lainnya} \end{cases} \quad (5)$$

a_{ij} adalah nilai keanggotaan titik x_i ke pusat kluster C_{kj} , s adalah jarak terpendek dari data x_i ke pusat kluster. Mengulangi langkah 3 sampai 6 hingga tidak adalagi objek yang berpindah.

Untuk pengelompokan dengan metode Fuzzy C-Means (FCM) didasarkan pada teori logika fuzzy yang pertama kali diperkenalkan oleh Lotfi Zadeh pada tahun 1965 dengan nama himpunan fuzzy (*fuzzy set*). FCM merupakan suatu teknik pengklasteran data yang mana keberadaan tiap-tiap data dalam suatu kluster ditentukan oleh nilai keanggotaan. Teknik ini pertama kali diperkenalkan oleh Dunn tahun 1973 kemudian Jim Bezdek pada tahun 1981. Konsep dasar FCM, pertama kali adalah menentukan pusat kluster yang akan menandai lokasi rata-rata untuk tiap kluster. Pada kondisi awal, pusat kluster masih belum akurat. Tiap-tiap data yang memiliki derajat keanggotaan untuk tiap-tiap kluster. Dengan cara memperbaiki pusat kluster dan nilai keanggotaan tiap-tiap data secara berulang, maka akan dapat dilihat bahwa pusat kluster akan bergerak menuju lokasi yang tepat (Kusumadewi, 2006).

Adapaun Algoritma FCM adalah sebagai berikut (Kusumadewi, 2006):

1. Tentukan:
 - a. Matriks X berukuran $n \times m$, dengan n = jumlah data yang akan dikluster; dan m = jumlah variabel.
 - b. Jumlah kluster yang akan dibentuk = $C (\geq 2)$.
 - c. Pangkat (pembobot) = $W (>1)$
 - d. Maksimum iterasi
 - e. Kriteria penghentian = ξ (nilai positif yang sangat kecil)
 - f. Iterasi awal, $t=1$ dan $\Delta = 1$.

2. Bentuk matriks partisi awal U^0 sebagai berikut:

$$U = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \dots & \mu_{1n}(x_n) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & \dots & \mu_{2n}(x_n) \\ \vdots & \vdots & & \vdots \\ \mu_{c1}(x_1) & \mu_{c2}(x_2) & \dots & \mu_{cn}(x_n) \end{bmatrix} \quad (6)$$

(matriks partisi awal biasanya dipilih secara acak)

3. Hitung pusat kluster V untuk setiap kluster:

$$V_{ij} = \frac{\sum_{k=1}^n (\mu_{ik})^w x_{kj}}{\sum_{k=1}^n (\mu_{ik})} \quad (7)$$

4. Perbaiki derajat keanggotaan setiap data pada setiap kluster (perbaiki matriks partisi), sebagai berikut:

$$\mu_{ik} = \left[\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(w-1)} \right]^{-1} \quad (8)$$

Dengan :

$$d_{ik} = d(x_k - V_i) = \left[\sum_{j=1}^m (x_{kj} - V_{ij}) \right]^{1/2} \quad (9)$$

5. Tentukan kriteria berhenti yaitu perubahan matriks partisi pada iterasi sekarang dengan iterasi sebelumnya sebagai berikut :

$$\Lambda = \|U^t - U^{t-1}\| \quad (10)$$

Apabila $\Lambda \leq \xi$ maka iterasi dihentikan, namun apabila $\Lambda > \xi$ maka naikan iterasi ($t = t+1$) dan kembali ke langkah 3

Pemilihan metode terbaik dimaksudkan untuk melihat kinerja masing-masing pengelompokan khususnya dalam mengelompokan kabupaten/kota di Jawa Tengah. Untuk mengukur kinerja metode analisis yang digunakan, menurut Bunkers (dalam Komariah dan Akbar, 2011) dilakukan evaluasi kelompok dengan kriteria dua nilai simpangan baku, yaitu simpangan baku dalam kelompok (S_w) dan antar kelompok (S_B) sehingga diperoleh metode terbaik, berikut persamaan S_w :

$$S_w = K^{-1} \sum_{k=1}^K S_k$$

S_w = rata-rata simpangan baku dalam klaster

S_k = simpangan baku klaster ke- k

Dimana K adalah banyaknya kelompok yang terbentuk dan S_k merupakan simpangan baku kelompok ke- k . Nilai S_B dicari dengan persamaan berikut:

$$S_B = \left[(K-1)^{-1} \sum_{k=1}^K (\bar{x}_k - \bar{x})^2 \right]^{1/2}$$

S_B = simpangan baku antar kelompok

Dimana \bar{x}_k adalah rata-rata kelompok ke- k dan \bar{x} adalah rata-rata keseluruhan kelompok.

Semakin kecil nilai S_w dan semakin besar nilai S_B maka metode tersebut memiliki kinerja yang baik, artinya mempunyai homogenitas yang tinggi (Komariah dan Akbar, 2011).

2. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) pusat, Provinsi Jawa Tengah dan Pemerintah Daerah Jawa Tengah tahun 2016, semarangkab.go.id untuk data Upah minimum Kabupaten/Kota dan data Pendapatan Asli Daerah (PAD) dari situs pad-dppad.jatengprov.go.id. Data akan dianalisis dengan menggunakan bantuan aplikasi R. Dengan variabel-variabel penelitian sebagai berikut:

X_1 = Jumlah Penduduk, X_2 = Produk Domestik Regional Bruto (PDRB), X_3 = Pendapatan Asli Daerah (PAD), X_4 = Upah Minimum Kabupaten/Kota (UMK), X_5 = Gizi Buruk Balita
 X_6 = Pengangguran, X_7 = Indeks Pembangunan Manusia (IPM)

Untuk mencapai tujuan dari penelitian ini maka akan dilakukan langkah-langkah sebagai berikut:

1. Mengumpulkan data dari sumbernya serta mendefinisikan masing-masing variabel.
2. Jika data memiliki ukuran yang berbeda-beda maka akan dilakukan standarisasi data terlebih dahulu sebelum dilakukan analisis.
3. Pemeriksaan asumsi analisis klaster.
4. Jika terjadi multikolinearitas maka akan dilakukan *Principal Component Analysis* (PCA).
5. Pemilihan ukuran jarak, dalam penelitian ini akan digunakan jarak *euclidean*
6. Analisis data dengan metode *K-Means* dan *Fuzzy C-Means* dengan jumlah kelompok 4 berdasarkan UNDP dalam mengelompokan kemiskinan wilayah membagi 4 kelompok.
7. Membandingkan hasil pengelompokan dengan menggunakan nilai S_w dan S_B . Interpretasi kelompok hasil pengelompokan metode terbaik.

3. HASIL PENELITIAN

3.1. Analisis Deskriptif Objek Penelitian

Pada penelitian ini data yang dianalisis terdiri dari 7 variabel yang merupakan variabel yang mempengaruhi kemiskinan. Gambaran umum data penelitian disajikan pada Tabel 1 yang memuat nilai minimum, mean, nilai maksimum dan standar deviasi dari ketujuh variabel yang digunakan.

Tabel 1. Hasil Analisis Deskriptif Ketujuh Variabel

Variabel	Minimum	Mean	Maximum	Standar Deviasi
Jumlah Penduduk	121,1	972	1788,9	418,13
PDRB	7015380	31023254	145993676	27636120
PAD	4.98E+10	2.48E+11	3.46E+12	5.62E+11
UMK	1265000	1415553	1909000	141099,20
Pengangguran	3927	24680	62332	15362,57
Gizi Buruk Balita	0	28,06	92	23,10
IPM	63,98	70,61	81,19	4,54

N=35

Dari tabel di atas dapat diketahui nilai masing-masing variabel, untuk variabel jumlah penduduk memiliki nilai minimum sebesar 121,1 dimana nilai ini merupakan jumlah penduduk dari Kota Magelang, mean sebesar 972, nilai maksimum sebesar 1788,9 yang menggambarkan jumlah penduduk terbanyak di Provinsi Jawa Tengah berada di Kabupaten Brebes, dan standar deviasi sebesar 418,13. Variabel PDRB memiliki nilai minimum sebesar 7015380 yang berarti angka PDRB Kabupaten/Kota di Jawa Tengah yang paling rendah yakni Kota Magelang, nilai maksimum sebesar 145993676 menggambarkan angka PDRB tertinggi yang dimiliki oleh Kota Semarang, dengan mean 31023254 dan standar deviasi sebesar 27636120. Variabel PAD memiliki nilai minimum 49.799.359.310 yang merupakan PAD Kota Magelang sementara untuk nilai maksimum atau PAD tertinggi di Jawa Tengah sebesar 345.677.0704.460 yang merupakan PAD Kota Semarang. Variabel UMK dengan mean 1415553, standar deviasi 141099,20, nilai minimum sebesar 1265000 dimana angka ini merupakan UMK/UMR terendah di Provinsi Jawa Tengah. UMK tersebut merupakan UMK Kabupaten Banjarnegara sedangkan UMK tertinggi adalah di Kota Semarang sebesar Rp. 1.909.000. Variabel pengangguran menggambarkan jumlah pengangguran Kabupaten/Kota dengan rata-rata 24.680, standar deviasi 15362,57, nilai minimum sebesar 3927 dimana angka tersebut merupakan angka pengangguran terendah di Jawa Tengah yang berada di Kota Magelang, sementara angka pengangguran tertinggi berada di Kabupaten Cilacap sebesar 62332. Variabel gizi buruk balita merupakan gambaran kasus gizi buruk dengan rata-rata kejadian 28,06, standar deviasi 23,10, nilai minimum 0 artinya tidak ada kejadian gizi buruk dimana angka ini merupakan data Kota Surakarta dan nilai maksimum 92 yang berarti kejadian gizi buruk tertinggi 92 kejadian yang terjadi di Kabupaten Brebes. Selanjutnya variabel IPM dengan mean 70,61, standar deviasi 4,54, nilai minimum 63,98 yang merupakan IPM Kabupaten Brebes, nilai maksimum sebesar 81,19 yang merupakan IPM Kota Semarang.

3.2 Uji Asumsi Non-Multikolinearitas

Uji asumsi Non-Multikolinearitas dilakukan dengan tujuan untuk mengetahui apakah terjadi multikolinearitas atau tidak pada data yang dianalisis. Jika tidak terjadi maka data dapat langsung dianalisis, namun jika terjadi multikolinearitas maka harus dilakukan penanganan data terlebih dahulu sebelum data dianalisis menggunakan analisis kluster. Menurut Yamin dan Kurniawan (dalam Puspitasari dan Susanti, 2016) dikatakan terjadi multikolinearitas apabila nilai korelasi antar variabel lebih besar dari 0,70. Berdasarkan hasil analisis ada korelasi antar variabel yang lebih besar dari 0,70 yakni korelasi antara variabel X_1 dengan X_5 sebesar 0,872 dan variabel X_2 dengan X_3 sebesar 0,77. Terjadinya multikolinearitas ini perlu dilakukan upaya perbaikan dengan reduksi variabel melalui *Principal Component Analysis* (PCA) atau Analisis Komponen Utama (AKU) untuk memperoleh variabel baru yang akan digunakan dalam analisis kluster.

3.4. *Principal Component Analysis* (PCA)

Salah satu solusi untuk menangani data yang multikolinearitas adalah dengan *Principal Component Analysis* (PCA) atau disebut juga dengan Analisis Komponen Utama (AKU). PCA merupakan salah satu solusi multikolinearitas dimana akan mereduksi jumlah variabel dari K variabel menjadi variabel baru J dimana $J < K$.

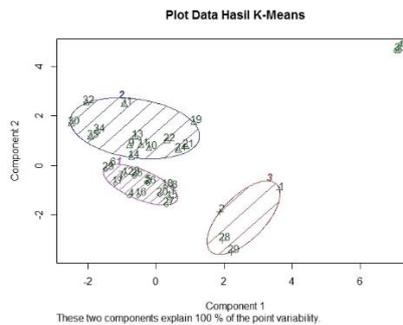
Tabel 2. Nilai Eigen, Proporsi Varian dan Proporsi Kumulatif

Komponen	Nilai Eigen	Proporsi Varian	Proporsi Kumulatif
1	3,168	0,453	0,453
2	2,533	0,362	0,815
3	0,411	0,059	0,874
4	0,358	0,051	0,925
5	0,256	0,037	0,962
6	0,161	0,023	0,985
7	0,111	0,016	1

Berdasarkan Tabel 2 di atas, dapat diketahui nilai eigen masing-masing komponen. Nilai eigen tersebut digunakan untuk menentukan jumlah komponen yang akan dipilih dengan ketentuan nilai eigen (λ) > 1. Tabel di atas memperlihatkan bahwa ada 2 komponen yang memiliki nilai eigen lebih besar dari 1 yakni komponen 1 dengan nilai eigen 3,168 dan komponen 2 sebesar 2,533. Berdasarkan ketentuan $\lambda > 1$, maka dipilih dua komponen utama dengan proporsi komponen masing-masing sebesar 0,45 untuk komponen 1 dan 0,362 untuk komponen 2. Adapun proporsi kumulatif dari kedua komponen tersebut sebesar 0,815, artinya sebesar 81,5% varians dari 7 variabel yang mempengaruhi kemiskinan dapat dijelaskan oleh 2 *Principal Component* (PC).

3.5. Pengelompokan Metode K-Means

Setelah dilakukan analisis melalui program R maka diperoleh hasil pengelompokan sebagai berikut:



Gambar 1. Plot hasil pengelompokan K-Means

Berdasarkan Gambar 1 di atas, dapat terlihat bahwa objek diplot ke dalam 4 kelompok. Untuk lebih jelasnya dapat dilihat pada Tabel 5.

Tabel 3. Hasil Pengelompokan Metode K-Means

Kelompok	Anggota Kelompok
1	Kab. Boyolali, Kab. Klaten, Kab. Sukoharjo, Kab. Karanganyar, Kab. Kudus, Kab. Demak, Kab. Semarang, Kab. Kendal, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Tegal, dan Kota Pekalongan
2	Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Purworejo, Kab. Wonosobo, Kab. Magelang, Kab. Wonogiri, Kab. Sragen, Kab. Grobogan, Kab. Blora, Kab. Rembang, Kab. Pati, Kab. Jepara, Kab. Temanggung, Kab. Batang, Kab. Pekalongan dan Kab. Pemalang
3	Kota Semarang
4	Kab. Cilacap, Kab. Banyumas, Kab. Tegal dan Kab. Brebes

Dari Tabel 3 di atas dapat diketahui masing-masing anggota kelompok dari keempat kelompok yang terbentuk. Kelompok 1 beranggotakan 13 kabupaten/kota, kelompok 2 beranggotakan 17 kabupaten, kelompok 3 beranggotakan 1 kota dan kelompok 4 beranggotakan 4 kabupaten.

3.6. Pengelompokan Metode *Fuzzy C-Means* (FCM)

Analisis kluster metode FCM dimulai dengan menentukan matriks $n \times m$ dimana data yang digunakan adalah data hasil standarisasi. Kemudian menentukan jumlah kluster $C=4$, pangkat pembobot $W=2$, kriteria penghentian $\xi = 10^{-5}$. Selanjutnya menentukan matriks partisi awal U^0 yang dipilih secara acak, menghitung pusat kluster dengan persamaan (7).

Tabel 4. Pusat Kluster Metode FCM

Kluster	X1	X2	X3	X4	X5	X6	X7
1	-0.0314	-0.3163	-0.1796	-0.5058	-0.1768	-0.0515	-0.5661
2	0.1210	0.4083	0.0812	1.2400	0.1241	-0.3235	0.1873
3	-1.1409	-0.3653	-0.1257	0.0090	-0.8523	-0.7523	1.1033
4	1.4741	0.4293	0.0725	-0.2840	1.7515	2.0118	-0.6698

Setelah menghitung pusat kluster, selanjutnya adalah menghitung derajat keanggotaan tiap kluster hingga diperoleh derajat keanggotaan akhir sebagai berikut :

Tabel 5. Derajat Keanggotaan Objek pada Tiap Kluster

Objek	Clust 1	Clust 2	Clust 3	Clust 4
1	0,17	0,25	0,11	0,47
2	0,05	0,05	0,02	0,88
3	0,89	0,05	0,04	0,02
4	⋮	⋮	⋮	⋮
33	0,22	0,31	0,23	0,25
34	0,13	0,12	0,72	0,03
35	0,11	0,07	0,8	0,02

Berdasarkan Tabel 5 di atas, maka dapat ditentukan setiap objek akan masuk ke kluster yang mana. Misal objek pertama yakni Kabupaten Cilacap memiliki nilai keanggotaan 0,17 untuk kluster 1, 0,25 kluster 2, 0,11 kluster 3 dan 0,47 untuk kluster 4. Dari nilai keanggotaan tersebut dapat ditentukan bahwa Kabupaten Cilacap masuk ke kluster 4 sebab derajat keanggotaan tertinggi berada pada kluster 4. Demikian juga untuk menentukan keanggotaan objek yang lain berdasarkan derajat keanggotaan tertinggi pada tiap klasternya, sehingga diperoleh hasil pengelompokan sebagai berikut:

Tabel 6. Hasil Pengelompokan Metode FCM

Kelompok	Anggota Kelompok
1	Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Purworejo, Kab. Wonosobo, Kab. Magelang, Kab. Klaten, Kab. Wonogiri, Kab. Sragen, Kab. Grobogan, Kab. Blora, Kab. Rembang, Kab. Pati, Kab. Jepara, Kab. Temanggung, Kab. Batang, Kab. Pekalongan dan Kab. Pemalang
2	Kota Semarang, Kab. Kudus, Kab. Demak, Kab. Kendal dan Kab. Semarang
3	Kab. Boyolali, Kab. Sukoharjo, Kab. Karanganyar, Kota Magelang, Kota Surakarta, Kota Salatiga dan Kota Tegal, dan Kota Pekalongan
4	Kab. Cilacap, Kab. Banyumas, Kab. Tegal dan Kab. Brebes

Dari Tabel 6 di atas dapat diketahui jumlah anggota setiap kluster, untuk kluster 1 beranggotakan 18 kabupaten, kluster 2 beranggotakan 4 kabupaten dan 1 kota, kluster 2

beranggotakan 1 kota dan 3 kabupaten, klaster 3 beranggotakan 3 kabupaten dan 5 kota, dan klaster 4 beranggotakan 4 kabupaten.

Tabel 7. Nilai S_w dan S_B Metode K-Means dan *Fuzzy C-Means*

Metode	S_w	S_B	Rasio
K-Means	0,413	3,335	0,124
Fuzzy C-Means	0,440	0,604	0,729

Dari Tabel 7 dengan melihat nilai simpangan baku masing-masing metode, dapat diketahui bahwa metode K-Means lebih baik daripada metode FCM pada kasus pengelompokan kabupaten/kota di Jawa Tengah berdasarkan faktor-faktor yang mempengaruhi kemiskinan.

Setelah dilakukan perbandingan metode pengklasteran berdasarkan rasio simpangan baku dalam klaster dan simpangan baku antar klaster dapat disimpulkan bahwa *Hierarchical Clustering* dengan metode K-Means adalah metode terbaik untuk mengelompokan kabupaten/kota di Jawa tengah berdasarkan faktor-faktor yang mempengaruhi kemiskinan tahun 2016. Hasil pengelompokan metode K-Means dapat dilihat pada Tabel 5 dimana kabupaten/kota dikelompokan 4 kelompok/klaster. Keempat klaster tersebut dapat diuraikan karakteristik masing masing berdasarkan variabel-variabel yang mempengaruhi kemiskinan. Tabel 8. Nilai Mean Variabel-variabel untuk Setiap Klaster

Variabel/Klaster	Klaster 1	Klaster 2	Klaster 3	Klaster 4
Jumlah Penduduk	705,45	973,35	1729,08	1643.073
PDRB	28963913,4	20943998.73	145993676,4	51810338,25
PAD	163786012319	129350046116	3456770704460	223322162117
UMK	1485942.54	1338711.76	1909000	1390000
Pengangguran	15922,31	22523,18	51229	55667.5
Gizi Buruk	12	30	14	75
IPM	74,43	67,86	81,19	67,22
Kemiskinan (%)	10,02	14,67	4,85	15.23

Berdasarkan Tabel 8 di atas dapat diuraikan karakteristik setiap klaster sebagai berikut:

1. Klaster 1 memiliki rata-rata jumlah penduduk paling kecil diantara 4 klaster, jumlah kejadian gizi buruk artinya wilayah-wilayah yang tergabung dalam klaster ini memiliki tingkat kesehatan khususnya pada kasus gizi balita lebih baik daripada wilayah yang lain, persentase kemiskinan terendah kedua setelah klaster 3 yakni Kota Semarang dengan rata-rata persentase kemiskinan dibawah angka provinsi bahkan angka nasional. Kemudian rata-rata nilai variabel lainnya menempati posisi tertinggi kedua setelah klaster 3.
2. Klaster 2 memiliki rata-rata UMK terendah jika dibandingkan klaster yang lainnya, artinya wilayah-wilayah yang tergabung dalam kelompok ini merupakan wilayah dengan rata-rata UMK yang rendah pada tahun 2016. Selain itu, rata-rata PDRB pada klaster ini juga paling rendah diantar klaster lainnya, artinya klaster ini diisi oleh wilayah yang mayoritas PDRBnya paling rendah. Rata-rata persentase kemiskinan diatas angka provinsi dan nasional.
3. Kaster 3 hanya beranggotakan 1 klaster yakni Kota Semarang dengan persentase kemiskinan terendah sebesar 4,85%. Memiliki PDRB, PAD, UMK, IPM dan rata-rata jumlah penduduk tertinggi dibandingkan klaster lainnya. Hal ini wajar jika Kota Semarang memiliki PDRB, PAD, UMK dan IPM tertinggi sebab Kota Semarang merupakan pusat pemerintahan Provinsi Jawa Tengah. Sekalipun demikian, angka pengangguran di kota ini masih terbilang cukup tinggi dengan rata-rata kedua tertinggi setelah klaster 4 sebanyak 51.229 orang.
4. Klaster 4 memiliki persentase penduduk miskin tertinggi, artinya wilayah yang tergabung dalam klaster ini merupakan wilayah-wilayah yang memiliki angka kemiskinan yang tinggi dimana kesejahteraan rakyat masih sangat rendah. Rata-rata angka kemiskinan

sebesar 15,23%, angka tersebut terpaut jauh dari angka provinsi dan nasional. Sehingga wilayah-wilayah yang tergabung dalam kluster ini perlu mendapatkan perhatian khusus dari pemerintah. Selain itu, rata-rata angka pengangguran tertinggi dibandingkan kluster yang lain, kasus gizi buruk terbanyak dengan rata-rata IPM terendah.

4. SIMPULAN

Berdasarkan hasil penelitian dapat ditarik beberapa kesimpulan:

1. Objek penelitian dikelompokkan kedalam 4 kelompok dengan hasil pengelompokan sebagai berikut:
 - a. Hasil pengelompokan metode K-Means: kluster 1 beranggotakan 13 kabupaten/kota, kluster 2 beranggotakan 17 kabupaten, kluster 3 beranggotakan 1 kota dan kluster 4 beranggotakan 4 kabupaten.
 - b. Hasil pengelompokan metode *Fuzzy C-Means* : kluster 1 beranggotakan 18 kabupaten, kluster 2 beranggotakan 1 kota dan 3 kabupaten, kluster 3 beranggotakan 8 kabupaten/kota dan kluster 4 beranggotakan 4 kabupaten.
2. Setelah dilakukan perbandingan antara metode *K-Means* dan metode *Fuzzy C-means* dengan menggunakan rasio simpangan baku dalam kelompok dengan antar kelompok diperoleh bahwa K-Means merupakan metode terbaik

5. REFERENSI

- BPPD Jateng. 2016. *Target dan Realisasi Pendapatan Daerah Tahun Anggaran 2016*. Diunduh di <http://pad-dppad.jatengprov.go.id:8080/jateng1/>. (14 Agustus 2017)
- BPS. 2017. *Data dan Informasi Kemiskinan Jawa Tengah Tahun 2011-2015*. Semarang: @Badan Pusat Statistika Provinsi Jawa Tengah.
- Everitt, B.S., Leese, M., Landau, S., dan Stahl, D. 2011. *Cluster Analysis*, 5th Edition. United Kindom: John Wiley & Sons.
- Hair, et. al. 2006. *Multivariate Data Analysis (6th ed.)*. Upper Saddle River, New Jersey: Prantice Hall, Inc.
- Hand, D., Mannila, H., dan Smyth, P. 2001. *Principles of data mining*, A Bradford book The MIT Press, Cambridge, Massachusetts London England. Diunduh di <http://box.cs.istu.ru/.../20Mining.pdf>. (2 Agustus 2017).
- Johnson, R.A dan Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. USA: Pearson Education, Inc.
- Karti dan Irhamah. 2013. Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Pendidikan SMA/SMK/MA dengan Metode *C-Means* dan *Fuzzy C-Means*. *Jurnal Sains dan Seni POMITS*. 2(2): 282-287.
- Komariyah, N dan Akbar, M.S. 2011. Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Kemiskinan dengan Metode Cluster Analysis. ITS.
- Kusumadewi, S., Hartati, S., Harjoko, A., dan Wardoyo, R. 2006. *Fuzzy Multi-Attribute Decision Making (Fuzzy MADM)*. Yogyakarta: Graha Ilmu.
- Machfudhoh, S dan Wahyuningsih, N. 2013. Analisis Cluster Kabupaten/Kota Berdasarkan Pertumbuhan Ekonomi Jawa Timur. *Jurnal Sains dan Seni POMITS*. 2(1).
- MacQueen, J.B. 1967. *Some Methods for Classification and Analysis of Multivariate Observations Hierarchical Grouping to Optimize an Objective Function*. Berkeley: University of California Press A967.

- Ningrat, D.S., Maruddani, D.A.I dan Wuryandari, T. 2016. Analisis Cluster dengan Algoritma *K-Means* Clustering untuk Pengelompokan Data Obligasi Korporasi. *Gaussian*.5(4):641-650.
- Prasetyo, Eko. 2012. Data Mining; Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta: Andi Offset.
- Puspitasari, M.W dan Susanti, M. 2016. Pengelompokan Kabupaten/Kota Berdasarkan Faktor-faktor yang Mempengaruhi Kemiskinan di Jawa Tengah Menggunakan Metode Ward dan Average Linkage. *Jurnal Matematika UNY*. 5(6).
- Rachmatin, D dan Sawitri, K. 2014. Perbandingan Antara Metode Agglomeratif, Metode Divisif, dan Metode K-Means dalam Analisis Klaster. Seminar Nasional Matematika UNPAR.
- Soemartini dan Supartini, E. 2015. Prosiding Nasional Penelitian Matematika dan Pembelajarannya II (KNPMP II) UMS. Surakarta, 18 Maret 2018.