

# Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Pendidikan SMA/SMK/MA dengan Metode *C-Means* dan *Fuzzy C-Means*

Hanna Silia Karti, Irhamah

Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

*e-mail*: irhamah@statistika.its.ac.id

**Abstrak**—Di Provinsi Jawa Timur memiliki beberapa komponen indikator pendidikan. Beberapa komponen yang masuk dalam indikator pendidikan di Jawa Timur yaitu Angka Partisipasi Murni, Angka Partisipasi Kasar, Angka Transisi, Angka Putus Sekolah, Angka Murid Mengulang, Angka Lulusan, Rasio Murid/Ruang Belajar, Rasio Kelas/Ruang Belajar, Rasio Murid/Guru dan Rasio Murid/Sekolah. Salah satu parameter keberhasilan pendidikan adalah menuntaskan APK dan APM mutu pendidikan hingga minimal mencapai 95%. Berdasarkan parameter tersebut pendidikan di Jawa Timur masih belum maksimal apabila ditinjau berdasarkan jenjang pendidikan formal khususnya pada jenjang pendidikan SMA/SMK/MA. Berdasarkan informasi yang telah didapat, penelitian ini dilakukan untuk mengetahui kabupaten/kota mana sajakah yang dapat dikelompokkan berdasarkan tingkat kemiripan berdasarkan indikator pendidikan SMA/SMK/MA di Provinsi Jawa Timur dengan menggunakan metode *c-means* dan *fuzzy c-means*. Pada kasus ini metode *c-means* memiliki kinerja yang hampir sama dengan metode *fuzzy c-means*. Hal ini berdasarkan perbandingan nilai *icdrate* yang hanya memiliki nilai selisih sebesar 0,001.

**Kata Kunci**—*C-means*, *Fuzzy C-means*, *Icdrate*, Indikator Pendidikan.

## I. PENDAHULUAN

Pendidikan adalah proses perubahan sikap dan tatalaku seseorang atau kelompok orang dalam usaha mendewasakan manusia melalui upaya pengajaran dan pelatihan, proses, cara, perbuatan mendidik [1]. Salah satu parameter keberhasilan pendidikan dapat dilihat dari indikator pendidikan di suatu daerah. Indikator pendidikan di Indonesia memiliki komponen yang berbeda-beda. Di Provinsi Jawa Timur memiliki beberapa komponen indikator pendidikan. Salah satu parameter keberhasilan pendidikan adalah menuntaskan Angka Partisipasi Kasar (APK) dan Angka Partisipasi Murni (APM) mutu pendidikan hingga minimal mencapai 95%. Berdasarkan parameter tersebut pendidikan di Jawa Timur masih belum maksimal apabila ditinjau berdasarkan jenjang pendidikan formal khususnya

pada jenjang pendidikan SMA/SMK/MA. Hal ini dapat dilihat dari data APK Dinas Pendidikan Provinsi Jawa Timur pada tahun 2011/2012 untuk usia SMA sebesar 74,21% sehingga dapat dikatakan bahwa persentase APK usia SMA/SMK/MA di Jawa Timur masih rendah karena belum mencapai 95%. Beberapa penelitian tentang pendidikan sudah banyak dilakukan sebelumnya. Perbandingan pengelompokan kabupaten/kota di Jawa Timur berdasarkan indikator pencapaian strategi T3 untuk Sekolah Menengah Kejuruan dengan metode *c-means* dan *fuzzy c-means*[2]. Kemudian, pengelompokan kabupaten dan kota di Jawa Timur berdasarkan tingkat partisipasi pendidikan [3]. Berdasarkan informasi yang telah didapat, penelitian ini dilakukan untuk mengetahui kabupaten/kota mana sajakah yang dapat dikelompokkan berdasarkan tingkat kemiripan atau kedekatan berdasarkan indikator pendidikan SMA/SMK/MA di Provinsi Jawa Timur dengan menggunakan metode *c-means* dan *fuzzy c-means*.

## II. TINJAUAN PUSTAKA

### A. Statistika *Statistika Deskriptif*

Statistika deskriptif adalah suatu metode yang berkaitan dengan pengumpulan dan penyajian data sehingga memberikan informasi yang berguna [4].

### B. *Distribusi Normal Multivariat*

Untuk mengetahui apakah data mengikuti distribusi normal, dilakukan uji distribusi normal multivariat. Pengujian distribusi normal multivariat dilakukan untuk memperkuat dugaan bahwa data sudah berdistribusi normal multivariat dan sebagai asumsi dasar yang harus dipenuhi sebelum menguji lainnya [5]. Pemeriksaan distribusi multinormal dapat dilakukan dengan cara melihat dari *q-q plot* dan dari nilai

$$d_j^2 = (x_j - \bar{x})' S^{-1} (x_j - \bar{x}), j=1,2,3,\dots,n \quad (1)$$

jika *plot* ini cenderung membentuk garis lurus lebih dari 50 persen dan nilai  $d_j^2 \leq \chi_{p,0.50}^2$ , maka data tidak berdistribusi normal multivariat. Dimana:  $d_j^2$  adalah nilai jarak data ke- $j$ ,  $x_j$  adalah variabel ke- $j$ ,  $\bar{x}$  adalah nilai vektor rata-rata dan  $S$  adalah nilai matriks varians-kovarians.

C. Uji Kaizer-Meyer-Olkin (KMO)

Untuk kecukupan sampel dapat diukur berdasarkan nilai Kaizer-Meyer-Olkin (KMO) yang ditunjukkan oleh persamaan

$$\frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (2)$$

dimana:  $i = 1, 2, 3, \dots, p$  dan  $j = 1, 2, \dots, p$ ,  $r_{ij}$  = korelasi

antara variabel  $i$  dan  $j$  dan  $a_{ij}$  = korelasi parsial antara variabel  $i$  dan  $j$ . Dengan hipotesis sebagai berikut.

$H_0$  :Jumlah data sudah cukup untuk dianalisis

$H_1$  :Jumlah data tidak cukup untuk dianalisis

Sampel akan dikatakan layak untuk dilakukan analisis faktor bilamana  $KMO > 0,5$  [5].

D. Uji Bartlett

Uji Bartlett digunakan untuk menguji persamaan korelasi antara dua atau lebih data variabel independen untuk dilihat nilai korelasinya [5]. Pengujian ini ditentukan hipotesisnya sebagai berikut.

$H_0 : \rho = I$

$H_1 : \rho \neq I$

Statistik Uji

$$\chi^2_{hitung} = - \left[ (n - 1) - \frac{(2p+5)}{6} \right] \ln|R| \quad (3)$$

dimana:

$n$  = Jumlah observasi

$p$  = Jumlah variabel

$|R|$  = Determinan dari matriks korelasi

Keputusan Tolak  $H_0$  jika nilai  $\chi^2_{hitung} > \chi^2_{\alpha, \frac{p(p-1)}{2}}$ , suatu

matriks korelasi dikatakan menyerupai matriks identitas bilamana nilai determinannya mendekati 1.

E. Analisis Pengelompokan

Analisis kelompok merupakan suatu teknik statistik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan kesamaan karakteristik yang dimilikinya. Terdapat dua metode dalam analisis kelompok yaitu metode hierarki dan metode nonhierarki[5].

F. C-Means

C-Means (CM) merupakan salah satu metode data clustering nonhirarki yang berusaha mempartisi data kedalam bentuk satu atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu kelompok yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan kedalam kelompok lain[6].

Algoritma C-Mean.

Metode c-mean merupakan metode yang algoritmanya mendeskripsikan bahwa tiap-tiap item yang dikelompokkan memiliki centroid atau rata-rata yang terdekat. Adapun langkah-langkah dalam c-means adalah berikut [5].

1. Membagi item-item ke dalam  $k$ -kelompok.
2. Menghitung nilai centroid dengan rumus sebagai berikut.

$$v_{ij} = \frac{\sum_{k=1}^{N_i} x_{kj}}{N_i} \quad (4)$$

dimana:

$v_{ij}$  adalah centroid klaster ke- $i$  untuk variabel ke- $j$ .

$N_i$  adalah jumlah data yang menjadi anggota klaster ke- $i$ .

$i, k$  adalah indeks dari klaster.

$j$  adalah indeks dari variabel.

$x_{kj}$  adalah nilai data ke- $k$  yang ada di dalam klaster tersebut untuk variabel ke- $j$ .

3. Kemudian mengelompokkan item berdasarkan centroid terdekat (jarak yang digunakan adalah jarak euclidean), dengan rumus sebagai berikut.

$$D_e = \sqrt{(\sum_{i=1}^p (x_i - s_i)^2)} \quad (5)$$

dimana:

$D_e$  adalah jarak Euclidean.

$i$  adalah banyaknya objek.

$p$  adalah banyak variabel.

$x$  merupakan koordinat objek.

$s$  merupakan koordinat centroid.

4. Menghitung kembali centroid kelompok ketika menerima item baru maupun item yang keluar.

G. Fuzzy C-Means

Metode Fuzzy C-Means (FCM) merupakan salah satu metode pengelompokan yang dikembangkan dari C-Means dengan menerapkan sifat fuzzy keanggotaannya. Metode FCM mengalokasikan kembali data ke dalam masing-masing kelompok memanfaatkan teori fuzzy[14]. Dalam metode FCM dipergunakan variabel membership function ( $u_{ik}$ ), yang merujuk pada seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu kelompok. FCM memperkenalkan suatu variabel  $m$  yang merupakan weighting exponent dari membership function. Variabel ini dapat mengubah besar pengaruh dari membership function, dalam proses clustering menggunakan metode FCM,  $m$  mempunyai wilayah nilai lebih besar dari 1 ( $m > 1$ ). Membership function untuk suatu data ke suatu kelompok tertentu[12]. Membership function mempunyai jangkauan nilai  $0 \leq u_{ik} \leq 1$ . Untuk metode FCM, objective function yang digunakan adalah sebagai berikut.

$$J(X, U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m D(x_k, v_i)^2 \quad (6)$$

dimana:

$n$  = Banyaknya data

$k$  = Variabel ke- $k$

$i$  = Kelompok ke- $i$

$c$  = Banyaknya cluster

$u_{ik}$  = Keanggotaan variabel ke- $k$  ke kelompok ke- $i$   $0 \leq u_{ik} \leq 1$

$$\sum_{i=1}^c u_{ik} = 1$$

$v_i$  = Nilai centroid kelompok ke- $i$

$m$  = Weighting exponent

$D(x_k, v_i)^2$  = Distance space yang digunakan

Algoritma Fuzzy C-Mean [8].

1. Menentukan jumlah cluster, misal  $c$ .
2. Menentukan inisiasi awal matriks partisi  $U$ .
3. Menghitung centroid dengan rumus sebagai berikut.

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m} \quad (7)$$

4. Menghitung distance space, yang merupakan jarak euclidean kuadrat dengan rumus sebagai berikut.

$$D_{ik} = D(x_k, v_i) = \sum_{i=1}^c \|x_k - v_i\|^2 = \sqrt{\sum_{i=1}^c (x_k - v_i)^2} \quad (8)$$

dengan:

$D_{ik}$  = Jarak antara objek dengan pusat kelompok

$x_k$  = Variabel ke- $k$

$v_i$  = Nilai *centroid* (pusat kelompok) kelompok ke- $i$   
 $c$  = Banyak *cluster*

- Menghitung nilai *membership function* masing-masing data ke masing-masing kelompok dengan persamaan

$$u_{ik} = \sum_{j=1}^c \left[ \frac{(D(x_k, v_i))^2}{(D(x_k, v_j))^2} \right]^{-1} \quad (9)$$

dimana:

$u_{ik}$  = *Membership function* data ke- $k$  ke kelompok ke- $i$   
 $v_i$  = Nilai *centroid* kelompok ke- $i$   
 $v_j$  = Nilai *centroid* kelompok ke- $j$   
 $m$  = *Weighting exponent*  
 $c$  = Banyaknya *cluster*

- Apabila  $|U^i - U^{(i-1)}| < \epsilon$ , maka proses berhenti. Namun apabila perubahan nilai *membership function* masih di atas nilai *threshold* ( $\epsilon$ ), maka kembali ke langkah 3.

#### H. Calinski – Harabasz Pseudo F-statistic

Metode yang digunakan untuk menentukan banyaknya kelompok yang optimum adalah Pseudo F-statistic. Pseudo F tertinggi menunjukkan bahwa kelompok tersebut menunjukkan hasil yang optimal, dimana keragaman dalam kelompok sangat homogen sedangkan antar kelompok sangat heterogen. Berikut rumus yang digunakan untuk mencari Pseudo F [9].

$$PseudoF = \frac{\left(\frac{R^2}{c-1}\right)}{\left(\frac{1-R^2}{n-c}\right)} \quad (10)$$

Dimana  $R^2 = \frac{(SST-SSW)}{SST}$  (11)

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ij}^k - \bar{x}^k)^2 \quad (12)$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ij}^k - \bar{x}_j^k)^2 \quad (13)$$

Keterangan:

$SST$  = (*Sum Square Total*) total jumlah dari kuadrat jarak terhadap rata-rata keseluruhan.

$SSW$  = (*Sum Square Within*) total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya.

$n$  = banyaknya sampel.

$c$  = banyaknya kelompok.

$p$  = banyaknya variabel.

$x_{ij}^k$  = sampel ke- $i$  pada kelompok ke- $j$  dan variabel ke- $k$ .

$\bar{x}^k$  = rata-rata seluruh sampel pada variabel- $k$ .

$\bar{x}_j^k$  = rata-rata sampel pada kelompok ke- $j$  dan variabel ke- $k$  [10].

#### I. Internal Cluster Dispersion Rate (Icdrate)

Beberapa macam metode untuk membandingkan hasil pengelompokan dapat dilakukan berbagai cara dan rumusan. Salah satunya dengan menghitung performansi klaster dengan menghitung nilai SSE dari hasil pengolahan data dan menghitung persebaran (*internal cluster dispersion rate*) dalam masing-masing klaster yang telah terbentuk. Semakin kecil nilai *icdrate* maka semakin baik hasil pengelompokannya [16]. Membandingkan metode klaster yang terbaik dengan mengevaluasi performansi algoritma dengan menggunakan prosentase rata-rata dari klasifikasi yang benar (*recovery rate*) dan nilai persebaran data-data dalam klaster (*internal cluster dispersion rate*) dari hasil akhir

pengelompokan yang didefinisikan dengan persamaan berikut [11].

$$icdrate = 1 - \frac{SSB}{SST} = 1 - \frac{SST-SSW}{SST} = 1 \quad (14)$$

Keterangan:

$SSB$  = (*Sum Square Between*)  $SST-SSW$

$SST$  = total jumlah dari kuadrat jarak terhadap rata-rata keseluruhan

$SSW$  = total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya

$R^2$  = (*Recovery Rate*)  $SSB/SST$

$k$  = banyaknya kelompok

$n$  = banyaknya observasi

#### J. Pengertian Pendidikan

Pendidikan adalah proses perubahan sikap dan tatalaku seseorang atau kelompok orang dalam usaha mendewasakan manusia melalui upaya pengajaran dan pelatihan, proses, cara, perbuatan mendidik [1].

#### K. Indikator Pendidikan

Berikut ini merupakan indikator pendidikan di Provinsi Jawa Timur [21].

##### 1. Angka Partisipasi Murni (APM)

Perbandingan antara jumlah siswa usia sekolah pada jenjang pendidikan tertentu dengan penduduk kelompok usia sekolah yang sesuai dan dinyatakan dalam persentase. Kriteria makin tinggi APM berarti makin banyak dan tepat anak usia sekolah yang bersekolah di tingkat pendidikan tertentu di suatu daerah. Idealnya=100 persen, bila lebih besar dari 100 persen karena adanya siswa usia sekolah dari luar daerah, daerah kota, atau daerah perbatasan.

##### 2. Angka Partisipasi Kasar (APK)

Perbandingan antara jumlah siswa dengan penduduk usia sekolah yang sesuai dan dinyatakan dalam persentase. Kriteria makin tinggi APK berarti makin banyak penduduk usia sekolah yang bersekolah di satuan pendidikan. Nilai APK yang baik mendekati 100 persen.

##### 3. Angka Melanjutkan (Angka Transisi)

Perbandingan antara jumlah siswa baru tingkat I pada jenjang pendidikan tertentu dengan jumlah lulusan pada jenjang yang lebih rendah dan dinyatakan dalam persentase. Kriteria makin tinggi angkanya makin baik. Idealnya = 100 persen berarti semua lulusan dapat ditampung di jenjang pendidikan yang lebih tinggi. Bila angkanya lebih dari 100 persen karena ada siswa/mahasiswa baru yang berasal dari daerah lain seperti di daerah kota dan perbatasan.

##### 4. Angka Putus Sekolah

Perbandingan antara jumlah putus sekolah pada tingkat dan jenjang tertentu dengan jumlah siswa pada tingkat dan jenjang yang sesuai pada tahun ajaran sebelumnya dan dinyatakan dalam persentase. Kriteria makin rendah nilainya, berarti makin baik, idealnya = 0 persen berarti tidak ada siswa yang putus sekolah.

##### 5. Angka Mengulang (AU)

Perbandingan antara jumlah siswa mengulang pada tingkat dan jenjang pendidikan tertentu dengan jumlah siswa pada tingkat dan jenjang yang sesuai tahun ajaran sebelumnya dan dinyatakan dalam persentase. Kriteria makin rendah nilainya,

berarti makin baik, idealnya = 0 persen berarti tidak ada siswa yang mengulang.

#### 6. Angka Lulusan (AL)

Perbandingan antara jumlah lulusan pada jenjang tertentu dengan jumlah siswa tingkat tertinggi dari jenjang pendidikan yang sesuai dan dinyatakan dalam persentase. Kriteria makin tinggi nilainya, berarti makin baik, idealnya untuk persekolahan = 100 persen, berarti siswa tingkat tertinggi lulus seluruhnya.

#### 7. Rasio Siswa dan Ruang Belajar (R-S/RB)

Perbandingan antara jumlah siswa dengan jumlah ruang belajar pada jenjang pendidikan tertentu. Kriteria makin tinggi nilai rasio berarti makin padat siswa di kelas atau makin kurang jumlah ruang belajar di daerah.

#### 8. Rasio Kelas dan Ruang Belajar (R-K/RB)

Perbandingan antara jumlah kelas dengan jumlah ruang belajar pada jenjang pendidikan tertentu. Kriteria idealnya = 1, berarti setiap ruang belajar hanya digunakan sekali, kurang dari 1 berarti terdapat ruang belajar yang tidak digunakan dan lebih dari 1 berarti terdapat ruang belajar yang digunakan lebih dari sekali.

#### 9. Rasio Murid dan Guru (R-M/G)

Perbandingan antara jumlah murid dengan jumlah guru pada jenjang pendidikan tertentu.

Rumus yang digunakan:

#### 10. Rasio Murid dan Sekolah (R-M/S)

Perbandingan antara jumlah murid dengan jumlah sekolah pada jenjang pendidikan tertentu.

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Dinas Pendidikan Provinsi Jawa Timur. Data merupakan indikator pendidikan yang ada di buku Data Pokok Pendidikan Provinsi Jawa Timur Tahun 2011/2012.

#### B. Variabel Penelitian

Dalam penelitian ini variabel yang digunakan adalah data kabupaten/kota di Provinsi Jawa Timur yang mendukung indikator pendidikan SMA/SMK/MA di Provinsi Jawa Timur tahun 2011/2012. Variabel-variabel yang digunakan yaitu Angka Partisipasi Murni ( $X_1$ ), Angka Partisipasi Kasar ( $X_2$ ), Angka Transisi ( $X_3$ ), Angka Putus Sekolah ( $X_4$ ), Angkat Murid Mengulang ( $X_5$ ), Angka Lulusan ( $X_6$ ), Rasio Murid/Ruang Belajar ( $X_7$ ), Rasio Kelas/Ruang Belajar ( $X_8$ ), Rasio Murid/Guru ( $X_9$ ), Rasio Murid/Sekolah ( $X_{10}$ ).

#### C. Langkah Analisis

Berikut ini merupakan beberapa langkah analisis yang digunakan pada penelitian ini.

- 1) Melakukan kajian tentang *CM Clustering* dan *FCM Clustering*.
- 2) Menyusun matriks ukuran  $N \times p$ , dimana  $N$  merupakan banyaknya observasi (kabupaten/kota di Provinsi Jawa Timur) dan  $p$  adalah banyaknya variabel (indikator pendidikan SMA/SMK/MA).
- 3) Melakukan analisis statistika deskriptif pada semua variabel.
- 4) Melakukan uji normal multivariat pada semua variabel.

- 5) Melakukan uji *Kaiser Meyer Olkin* dan uji *Bartlett* pada semua variabel untuk mengetahui kecukupan data dan ada tidaknya korelasi antar variabel.
- 6) Melakukan standardisasi/transformasi pada semua variabel.
- 7) Melakukan pengelompokan dengan metode *CM Clustering* dan *FCM Clustering*.
- 8) Menggambarkan peta pengelompokan dari metode *CM Clustering* dan *FCM Clustering*.
- 9) Melakukan interpretasi hasil *CM Clustering* dan *FCM Clustering* pada kasus data indikator pendidikan SMA/SMK/MA kabupaten/kota di Provinsi Jawa Timur.
- 10) Melakukan perbandingan hasil pengelompokan antara metode *CM Clustering* dan *FCM Clustering* dengan metode *icdrate*. Metode *icdrate* merupakan metode yang sering digunakan dan sederhana.

### IV. ANALISIS DAN PEMBAHASAN

#### A. Analisis Deskriptif

Kabupaten Madiun memiliki persentase angka partisipasi murni tertinggi yaitu sebesar 88,57 persen. Daerah-daerah yang memiliki angka partisipasi murni 3 terendah adalah Kabupaten Probolinggo sebesar 33,48 persen, Kabupaten Bangkalan sebesar 28,92 persen dan Kabupaten Sampang sebesar 26,44 persen. Persentase angka partisipasi kasar di Jawa Timur terdapat 6 kabupaten/kota yang diatas persentase ideal, yaitu Kota Surabaya, Kota Malang, Kota Madiun, Kota Kediri, Kota Pasuruan dan Kabupaten Madiun. Kota Kediri memiliki persentase angka partisipasi kasar tertinggi yaitu sebesar 115,9 persen. Daerah-daerah yang memiliki angka partisipasi kasar 3 terendah adalah Kabupaten Lumajang sebesar 46,98 persen, Kabupaten Bangkalan sebesar 37,6 persen dan Kabupaten Sampang sebesar 35,26 persen.

Persentase angka transisi di Provinsi Jawa Timur pada jenjang SMA/SMK/MA paling tinggi terdapat pada Kota Kediri yaitu sebesar 180,71 persen. Daerah-daerah yang memiliki angka transisi 3 terendah adalah Kabupaten Jombang sebesar 43,69 persen, Kabupaten Blitar sebesar 35,09 persen dan Kota Malang sebesar 31,83 persen. Persentase angka putus sekolah 3 tertinggi terdapat pada Kabupaten Trenggalek yaitu sebesar 2,6 persen, Kabupaten Nganjuk sebesar 1,97 persen dan Kabupaten Blitar sebesar 1,9 persen. Daerah yang memiliki angka putus sekolah terendah adalah Kabupaten Sidoarjo, Kabupaten Madiun dan Kabupaten Tulungagung yaitu sebesar 0,01 persen.

Persentase angka mengulang di Provinsi Jawa Timur dilihat dari 3 kabupaten dengan persentase angka mengulang tertinggi adalah Kabupaten Nganjuk, Kota Pasuruan dan Kabupaten Bondowoso dengan persentase masing-masing sebesar 3,85 persen, 1,07 persen dan 0,97 persen. Untuk daerah di Jawa Timur yang memiliki persentase angka mengulang terendah adalah Kabupaten Tulungagung. Persentase angka lulusan di Provinsi Jawa Timur, daerah yang lulus 100 persen adalah Kabupaten Gresik dan Kabupaten Bangkalan. Untuk daerah yang memiliki persentase angka kelulusan 3 terendah adalah Kota Batu sebesar 90,98 persen, Kota Blitar sebesar 90,71 persen dan Kabupaten Ponorogo sebesar 90,14 persen.

Rasio murid/ruang belajar di Provinsi Jawa Timur tiga daerah yang memiliki rasio murid/ruang belajar tertinggi adalah Kota Probolinggo, Kota Mojokerto dan Kabupaten Probolinggo, dengan nilai rasio masing-masing sebesar 71:1, 59:1, dan 45:1. Perlu adanya penambahan ruang belajar juga pada kedua daerah tersebut, agar siswa didalam ruang belajar tidak terlalu banyak, sehingga proses belajar mengajar bisa lebih baik. Daerah yang memiliki rasio murid/ruang kelas terendah adalah Kabuapten Madiun sebesar 18:1, artinya dalam 1 ruang belajar di Kabupaten Madiun hanya berisi 18 siswa.

Rasio kelas/ruang belajar di Provinsi Jawa Timur. Tiga daerah yang memiliki rasio paling tinggi adalah Kota Probolinggo sebesar 3 (pembulatan dari 2,08) artinya dalam 1 ruang belajar digunakan 2 hingga 3 kali proses belajar mengajar. Idealnya adalah 1, berarti setiap ruang belajar hanya digunakan sekali. Untuk daerah yang memiliki rasio kelas/ruang belajar terendah adalah Kabupaten Sampang sebesar 1 (pembulatan dari 0,54) artinya dalam 1 ruang belajar tidak digunakan atau digunakan hanya 1 kali proses belajar mengajar.

Rasio murid/guru, dapat dilihat rasio tertinggi terdapat pada Kabupaten Probolinggo sebesar 23 artinya 1 guru mengajar 23 siswa. Untuk daerah dengan rasio murid/guru terendah adalah Kota Batu, Kabupaten Gresik dan Kabupaten Sampang sebesar 9, artinya 1 guru megajar 9 siswa. Rasio murid/sekolah di Provinsi Jawa Timur tertinggi terdapat pada Kota Probolinggo sebesar 726 artinya di Kota Probolinggo rata-rata 1 sekolah memiliki murid sebanyak 726 siswa. Untuk rasio terendah terdapat pada Kabupaten Madiun sebesar 116 artinya di Kabupaten Madiun rata-rata 1 sekolah terdapat 116 siswa.

*B. Analisis Pengelompokan C-Means dan Fuzzy C-means*

Tabel 1 menjelaskan uji *Kaiser Meyer Olkin* dan uji *Bartlett*. Berikut adalah hasil analisis dari *software* SPSS.

Berdasarkan output diatas diperoleh nilai KMO sebesar 0,652 yang menunjukkan lebih besar dari 0,5 ( $KMO > 0,5$ ). Hal ini berarti bahwa gagal tolak  $H_0$  yang artinya bahwa data sudah cukup baik untuk dianalisis. Selain itu juga diketahui hasil uji *Bartlett*. Berdasarkan Tabel 2 dapat dilihat bahwa nilai *Chi-Square* adalah 192,59, dengan derajat bebas sebesar 45, dan *P-value* (sig.) sebesar 0,000. Karena *P-value*  $0,000 < 0,05$  maka tolak  $H_0$ . Artinya ada korelasi antar variabel bebas.

Untuk melanjutkan pada analisis pengelompokan, perlu adanya standardisasi/transformasi variabel. Hal ini dikarenakan data yang digunakan mempunyai variabilitas satuan, maka perlu dilakukan langkah standardisasi atau transformasi terhadap variabel yang relevan ke bentuk zscore. Hasil pengelompokan dapat dilihat dalam Tabel 2, yang menunjukkan pengelompokan paling optimum pada metode *c-means* adalah 2 kelompok dan pada metode *fuzzy c-means* adalah 2 kelompok. Hal ini dikarenakan pada kedua metode tersebut memiliki nilai PseudoF yang paling besar diantara kelompok yang lain. Nilai dari metode metode *c-means* dan *fuzzy c-means* masing-masing sebesar 9,24 dan 9,19.

Pengelompokan berikut ini dapat dilihat daerah mana saja yang memiliki persamaan karakteristik berdasarkan indikator pendidikan SMA/SMK/MA di Provinsi Jawa Timur.

Tabel 1.  
Uji KMO dan Uji *Bartlett*

|  |                    |
|--|--------------------|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | 0,652              |
| Bartlett's Test of Sphericity                    | Approx. Chi-Square |
|  | Df                 |
|  | Sig.               |
|  | 192,59             |
|  | 45                 |
|  | 0                  |

Tabel 2.  
Perbandingan Nilai Pseudo F

| Metode   | KELOMPOK |      |      |      |      |      |      |      |      |      |
|----------|----------|------|------|------|------|------|------|------|------|------|
|          | 2        | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |      |
| Pseudo F | C        | 9,24 | 8,59 | 6,22 | 6,37 | 5,86 | 4,60 | 4,15 | 4,42 | 3,76 |
|          | M        |      |      |      |      |      |      |      |      |      |
|          | F        |      |      |      |      |      |      |      |      |      |
|          | C        | 9,19 | 6,59 | 5,72 | 4,61 | 3,81 | 3,00 | 3,15 | 2,05 | 2,50 |
|          | M        |      |      |      |      |      |      |      |      |      |

Berdasarkan hasil dari analisis *c-means* dengan membagi menjadi 2 kelompok didapatkan masing-masing anggota kelompok yang memiliki kesamaan adalah sebagai berikut. Kelompok 1: Kota Surabaya, Kota Malang, Kota Madiun, Kota Kediri, Kota Mojokerto, Kota Blitar, Kota Pasuruan, Kota Probolinggo, Kota Batu, Kab. Gresik, Kab. Sidoarjo, Kab. Mojokerto, Kab. Madiun. Kelompok 2: Kab. Jombang, Kab. Bojonegoro, Kab. Tuban, Kab. Lamongan, Kab. Ngawi, Kab. Magetan, Kab. Ponorogo, Kab. Pacitan, Kab. Kediri, Kab. Nganjuk, Kab. Blitar, Kab. Tulungagung, Kab. Trenggalek, Kab. Malang, Kab. Pasuruan, Kab. Probolinggo, Kab. Lumajang, Kab. Bondowoso, Kab. Situbondo, Kab. Jember, Kab. Banyuwangi, Kab. Pamekasan, Kab. Sampang, Kab. Sumenep, Kab. Bangkalan.

Hasil dari analisis dengan metode *fuzzy c-means* didapatkan pengelompokan terbaik sebanyak 2 kelompok masing-masing anggota kelompok adalah sebagai berikut. Kelompok 1: Kota Surabaya, Kota Malang, Kota Madiun, Kota Kediri, Kota Mojokerto, Kota Blitar, Kota Pasuruan, Kota Probolinggo, Kota Batu, Kab. Gresik, Kab. Sidoarjo, Kab. Mojokerto, Kab. Jombang, Kab. Madiun, Kab. Magetan, Kab. Ponorogo, Kab. Tulungagung, Kab. Pasuruan, Kab. Bondowoso. Kelompok 2:

Kab. Bojonegoro, Kab. Tuban, Kab. Lamongan, Kab. Ngawi, Kab. Pacitan, Kab. Kediri, Kab. Nganjuk, Kab. Blitar, Kab. Trenggalek, Kab. Malang, Kab. Probolinggo, Kab. Lumajang, Kab. Situbondo, Kab. Jember, Kab. Banyuwangi Kab. Pamekasan, Kab. Sampang, Kab. Sumenep, Kab. Bangkalan.

Karakteristik yang terdapat pada kelompok 1 dan kelompok 2 pada metode *c-means* dan *fuzzy c-means* memiliki kesamaan. Pada kelompok 1 memiliki karakteristik sebagai berikut: angka partisipasi murni, angka partisipasi kasar, angka transisi dan angka lulusan yang tinggi; angka putus sekolah dan angka murid mengulang yang rendah, serta rasio murid/ruang belajar, rasio kelas/ruang belajar, rasio murid/gurudan rasio murid/sekolah yang cukup. Pada kelompok 2 memiliki karakteristik angka partisipasi murni, angka partisipasi kasar, angka transisi dan angka lulusan yang rendah; angka putus sekolah dan angka murid mengulang yang tinggi; serta rasio murid/ruang, rasio kelas/ruang belajar, rasio murid/gurudan rasio murid/sekolah yang cukup.

*C. Perbandingan Metode C-Means dan Fuzzy C-means*

Pada Tabel 3 merupakan hasil pengelompokan kab/kota berdasarkan indikator pendidikan SMA/SMK/MA di Provinsi Jawa Timur dengan metode *c-means* adalah 2 kelompok dan pada metode *fuzzy c-means* adalah 2 kelompok. Berdasarkan

perhitungan *icdrate*, pada kasus ini pengelompokan dengan metode *c-means* dan metode *fuzzy c-means* memiliki kinerja yang hampir sama. Hal ini dapat dilihat dari nilai *icdrate* untuk *c-means* yaitu sebesar 0,796. Pada metode *fuzzy c-means* nilai *icdrate* nilainya sebesar 0,797. Jadi, pada kasus ini metode *c-means* memiliki kinerja yang hampir sama dengan metode *fuzzy c-means*. Hal ini berdasarkan perbandingan nilai *icdrate* yang hanya memiliki nilai selisih sebesar 0,001.

Pada metode *c-means* terdapat 13 kabupaten/kota yang masuk pada kelompok 1 yaitu Kota Surabaya, Kota Malang, Kota Madiun, Kota Kediri, Kota Mojokerto, Kota Blitar, Kota Pasuruan, Kota Probolinggo, Kota Batu, Kab. Gresik, Kab. Sidoarjo, Kab. Mojokerto, Kab. Madiun. Sedangkan pada metode *fuzzy c-means* terdapat 19 kabupaten kota yang masuk pada kelompok 1. Terdapat tambahan 6 kabupaten/kota yang pada metode *c-means* terdapat pada kelompok 2, namun pada metode *fuzzy c-means* masuk kekelompok 1. Kabupaten/kota tersebut adalah Kab. Jombang, Kab. Magetan, Kab. Ponorogo, Kab. Tulungagung, Kab. Pasuruan, dan Kab. Bondowoso. Perbedaan pada metode *c-means* dan metode *fuzzy c-means* terdapat pada keenam daerah tersebut, dikarenakan memiliki karakteristik yang hampir sama dengan metode *c-means* yang masuk pada kelompok 2, sedangkan pada metode *fuzzy c-means* masuk pada kelompok 1. Terdapat dua variabel yang berpengaruh yaitu angka lulusan dan rasio kelas/ruang belajar.

## V. KESIMPULAN DAN SARAN

Kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut.

1. Karakteristik kabupaten/kota di Jawa Timur berdasarkan indikator pendidikan SMA/SMK/MA. APM dan APK terendah terdapat pada Kabupaten Sampang masing-masing sebesar 26,44 persen dan 35,26 persen. Angka transisi paling rendah terdapat pada Kota Malang sebesar 31,83 persen. Angka putus sekolah tertinggi terdapat pada Kabupaten Trenggalek yaitu sebesar 2,6 persen. Angka mengulang tertinggi adalah Kabupaten Nganjuk sebesar 3,85 persen. Angka lulusan paling rendah terdapat pada Kabupaten Ponorogo sebesar 90,14 persen. Rasio murid/ruang belajar tertinggi adalah Kota Probolinggo sebesar 71:1. Rasio kelas/ruang belajar paling tinggi adalah Kota Probolinggo sebesar 3:1. Rasio murid/guru tertinggi terdapat pada Kabupaten Probolinggo sebesar 23:1. Rasio murid/sekolah tertinggi terdapat pada Kota Probolinggo sebesar 726:1.
2. Hasil pengelompokan paling optimum sebanyak 2 kelompok. Karakteristik yang terdapat pada kelompok 1 dan kelompok 2 pada metode *c-means* dan *fuzzy c-means* memiliki kesamaan. Pada kelompok 1 memiliki karakteristik sebagai berikut: angka partisipasi murni, angka partisipasi kasar, angka transisi dan angka lulusan yang tinggi; angka putus sekolah dan angka murid mengulang yang rendah, serta rasio murid/ruang belajar, rasio kelas/ruang belajar, rasio murid/gurudan rasio murid/sekolah yang cukup. Pada kelompok 2 memiliki karakteristik angka partisipasi murni, angka partisipasi kasar, angka transisi dan angka lulusan yang rendah; angka putus sekolah dan angka murid mengulang yang tinggi;

Tabel 3.

| Perbandingan Nilai <i>icdrate</i> |         |
|-----------------------------------|---------|
| Metode                            | ICDrate |
| CM                                | 0,796   |
| FCM                               | 0,797   |

3. serta rasio murid/ruang, rasio kelas/ruang belajar, rasio murid/gurudan rasio murid/sekolah yang cukup.
4. Perbandingan hasil pengelompokan, pada kasus ini metode *c-means* memiliki kinerja yang hampir sama dengan metode *fuzzy c-means*.

Pada penelitian berikutnya diharapkan agar lebih optimal maka perbandingan yang dilakukan tidak hanya menggunakan metode *c-mean* dan *fuzzy c-means*. Penentuan metode terbaik dapat menggunakan selain *icdrate*.

## DAFTAR PUSTAKA

- [1] Pusat Bahasa Departemen Pendidikan Nasional. 2013. *Pendidikan*. Diakses dari <http://www.sarjanaku.com/2012/12/pengertian-pendidikan-menurut-para-ahli.html> Kamis, 28 Februari 2013. Pukul 07.09 WIB.
- [2] F Qori'ah, 2012. *Perbandingan Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pencapaian Strategi T3 untuk Sekolah Menengah Kejuruan dengan Metode C-Means dan Fuzzy C-Means*. Tugas Akhir, Jurusan Statistika FMIPA-ITS, Surabaya. Wahyudi, I., 2009, *Analisis Faktor-Faktor Yang Mempengaruhi Korupsi Anggaran Pendapatan Belanja Daerah (APBD) Di Malang Raya*, Universitas Muhammadiyah, Gresik.
- [3] P Nugrahani, 2009. *Pengelompokan Kabupaten/Kota Jawa Timur berdasar-kan Tingkat Partisipasi Pendidikan*. Tugas Akhir, Jurusan Statistika FMIPA-ITS, Surabaya
- [4] R.E Walpole, dan Myer. 1995. *Ilmu Peluang dan Statistik untuk Insinyur dan Ilmuwan* (terjemahan RK Sembiring) Edisi keempat. Bandung: Institut Teknologi Bandung.
- [5] R.A Johnson dan D.W Wichern, 2007. *Applied Multivariate Statistical Analysis, 6<sup>th</sup> ed.* Prentice Hall International Inc, New Jersey.
- [6] Y. Agusta, 2007. *K-Means Penerapan, Permasalahan dan Metode Terkait*. *Jurnal Sistem dan Informatika*. Vol.3: 47-60.
- [7] R.R Syoer, 2011. *Analisis Kelompok dengan Algoritma Fuzzy C-Means dan Gath-Geva Clustering (Studi Kasus Pengelompokan Desa/Kelurahan di Kabupaten Kutai Kertanegara)*. Thesis, Jurusan Statistika FMIPA-ITS, Surabaya.
- [8] J.C Bezdek, R. Ehrlich, dan W. Full, 1984. *FCM: Fuzzy C-Means Clustering Algorithm*. USA: Computers & Geosciences Vol. 10, No. 2-3, pp. 191-203.
- [9] A.R Orpin dan V.E Kostylev, 2006. Towards a statistically valid method of textural sea floor characterization of benthic habitats. *Marine Geology* 225 : 209-222.
- [10] A. Hinde, T. Whiteway, R. Ruddick, dan A.D Heap, 2007. *Seascape of the Australian Margin and adjacent sea floor: Keystroke Methodology*. Canberra: Geoscience Australia.
- [11] S.A Mingotidan O.J Lima, 2006. *Comparing SOM Neural Network with Fuzzy C-Means, C-Means and Traditional Hierarchical Clustering Algorithms*. *European Journal of Operational Research* 174: 1742-1759.