

MANAJEMEN PENGELOMPOKAN SMP/SEDERAJAT NEGERI DI KOTA BANDUNG DENGAN *FUZZY C-MEANS CLUSTERING*

Nurul Qitshi¹, Dhita Diana Dewi², Susi Widyastuti³

¹Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Poltek Cirebon

²Program Studi Teknik Informatika, Universitas Nusa Putra

³Program Studi Komputerisasi Akuntansi, Sekolah Tinggi Ilmu Komputer Poltek Cirebon

¹nurul.qitshi@stikompoltek.ac.id,

²dhita.dianadewi@nusaputra.ac.id,³susi.widyastuti@stikompoltek.ac.id

ABSTRACT - *The National Examination (UN) is one of the efforts to promote the improvement of school education quality. In addition, the results of the UN are useful as considerations for admission to the next level of education. In relation to the efforts to improve the quality of education, grouping is carried out for public Junior High Schools (SMP) or equivalent in Bandung City. To address this issue, a case study is needed that can classify public Junior High Schools (SMP) or equivalent. This research involves studying cluster analysis and its implementation using the Fuzzy C-means algorithm for clustering public Junior High Schools (SMP) or equivalent in Bandung City based on UN scores. The data used consists of the average UN scores from 60 public Junior High Schools (SMP) or equivalent in Bandung City in 2018, taken from the Report on UN Results by the Ministry of Education and Culture's Assessment Center. After processing the data using the Fuzzy C-means method, the final results indicate that 7 groups are the best clusters with the smallest RMSD and MAD values. The highest average UN scores for all subjects are found in group 2, while the lowest average UN scores for all subjects are found in group 3.*

Keywords: *Cluster Analysis, Fuzzy Logic, Fuzzy C-Means, National Examination*

ABSTRAK - Ujian Nasional (UN) merupakan salah satu upaya dalam memacu peningkatan mutu pendidikan sekolah. Selain itu hasil dari UN berguna sebagai pertimbangan seleksi masuk ke jenjang pendidikan selanjutnya. Berkaitan dengan upaya peningkatan mutu pendidikan, dilakukan pengelompokan terhadap SMP/Sederajat Negeri di Kota Bandung. Untuk mengatasi hal tersebut, dibutuhkan sebuah studi kasus yang dapat mengelompokkan SMP/Sederajat Negeri. Dalam penelitian ini dilakukan penelitian mengenai *cluster analysis* dan implementasinya menggunakan algoritma *Fuzzy C-means* untuk *clusterisasi* SMP/Sederajat Negeri di Kota Bandung berdasarkan nilai UN. Data yang digunakan merupakan data rata-rata nilai UN dari 60

SMP/Sederajat Negeri di Kota Bandung pada tahun 2018 yang diambil dari Laporan Hasil UN Pusat Penilaian Pendidikan Kementerian Pendidikan dan Kebudayaan. Setelah data tersebut diolah dengan metode *Fuzzy C-means*, hasil akhir menunjukkan bahwa 7 kelompok merupakan *cluster* terbaik dengan nilai RMSD dan nilai MAD terkecil, dengan rata-rata nilai UN untuk semua mata pelajaran tertinggi berada di kelompok 2, sedangkan rata-rata nilai UN untuk semua mata pelajaran terendah berada pada kelompok 3.

Kata Kunci: *Analisis Cluster, Logika Fuzzy, Fuzzy C-means, Ujian Nasional*

PENDAHULUAN

Ujian Nasional atau biasa disingkat dengan UN merupakan alat ukur tercapainya kompetensi lulusan secara nasional dengan mengacu pada Standar Kompetensi Lulusan (SKL) (Damarjati, 2018). UN juga merupakan upaya pemicu peningkatan mutu pendidikan sekolah dan sebagai motivator agar mendapatkan hasil yang terbaik. Hasil dari UN juga biasanya digunakan sebagai bahan pertimbangan seleksi masuk ke jenjang pendidikan selanjutnya. Berkaitan dengan upaya peningkatan mutu pendidikan, pengelompokan SMP/Sederajat Negeri di Kota Bandung berdasarkan nilai rata-rata UN pada setiap mata pelajaran dianggap perlu dilakukan untuk menentukan SMP/Sederajat mana saja yang diprioritaskan perlu dilakukan pengayaan kepada siswa untuk meningkatkan nilai rata-rata UN.

Berdasarkan hal tersebut, dibutuhkan sebuah studi kasus yang dapat mengelompokkan SMP/Sederajat Negeri di Kota Bandung. Algoritma *Fuzzy C-means Clustering* dapat digunakan untuk mengatasi hal tersebut.

Clustering adalah metode bertujuan untuk membagi data menjadi beberapa kelompok berdasarkan kemiripan antara satu data dengan data lainnya sehingga objek-objek yang memiliki tingkat kemiripan yang hampir sama antara satu sama lain berada dalam satu *cluster* yang sama (Johnson & Wichern, 2017). *Fuzzy clustering* merupakan salah satu teknik *clustering* data, dimana metode ini akan menentukan *cluster* optimal dalam suatu ruang vector menggunakan jarak *euclidean* (Sutoyo, 2015).

Teknik *Fuzzy clustering* yang sering digunakan yaitu *Fuzzy C-means Clustering*. *Fuzzy C-means Clustering* mampu mendeteksi *cluster* tingkat tinggi. Selain itu, *fuzzy c-means clustering* juga dapat menunjukkan hubungan antar pola *cluster* yang berbeda, serta dapat memberikan hasil cukup efektif untuk meningkatkan homogenitas pada setiap kelompok.

Berdasarkan latar belakang yang telah dijelaskan, maka dilakukan penelitian mengenai analisis *cluster* dengan menggunakan algoritma *Fuzzy C-means* untuk mengelompokkan SMP/Sederajat Negeri di Kota Bandung berdasarkan nilai UN. Manfaat dari penelitian ini adalah dapat menentukan berapa banyak *cluster* yang dapat terbentuk, serta mencari nilai rata-rata UN tertinggi dan nilai rata-rata UN terendah untuk selanjutnya dilakukan pengayaan terutama untuk sekolah dengan nilai rata-rata UN terendah. Selain itu, manfaat yang diharapkan adalah penelitian

dapat menjadi bahan pertimbangan oleh instansi terkait dalam pengambilan keputusan mengenai peningkatan kualitas pendidikan.

METODE PENELITIAN

Pada bagian ini akan menjelaskan mengenai data yang akan digunakan dalam penelitian dan juga langkah-langkah untuk menentukan pengelompokan nilai UN SMP/Sederajat di Kota Bandung menggunakan metode *Fuzzy C-means Clustering*.

1. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data nilai UN dari empat pelajaran yang diujikan dari 60 SMP/Sederajat Negeri di Kota Bandung pada tahun 2018 yang diambil dari Laporan Hasil Ujian Nasional Pusat Penilaian Pendidikan Kementerian Pendidikan dan Kebudayaan (Kemdikbud, 2019). Variabel yang digunakan adalah rata-rata nilai UN untuk pelajaran Bahasa Indonesia (x_1), Bahasa Inggris (x_2), Matematika (x_3), dan IPA (x_4).

2. Variabel yang Saling Ortogonal

X harus ortogonal karena penggunaan jarak Mahalanobis dalam analisis *cluster* mengharuskan antar variabelnya saling ortogonal. Matriks X dikatakan ortogonal jika hasil kali sesuai dengan persamaan (1)

$$Y^t Y = I \quad (1)$$

Jika variabel tidak ortogonal, maka analisis komponen utama dapat digunakan untuk mendapatkan variabel baru yang saling ortogonal.

3. Multivariat Normal

Uji Mardia adalah pengujian yang dapat digunakan untuk menentukan normalitas pada data multivariat (Rencher, 2002).

$H_0 : F(x) = F_0(x)$, data berdistribusi normal multivariat

$H_1 : F(x) \neq F_0(x)$, data tidak berdistribusi normal multivariat

$F(x)$ adalah representasi distribusi kumulatif dari data jika data tersebut berdistribusi normal multivariat dan $F_0(x)$ adalah representasi kumulatif dari fungsi distribusi normal multivariat. Jika data berdistribusi normal, maka *skewness* akan berdistribusi χ^2 dan *kurtosis* akan berdistribusi Z .

Statistik uji :

$$skew = \frac{nb_{1,k}}{6} \quad (2)$$

$$kurt = \frac{b_{2,k} - k(k-2)}{\sqrt{\frac{sk(k+2)}{n}}} \quad (3)$$

dalam hal ini,

$$b_{1,k} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n g_{ij}^3 \quad (4)$$

$$b_{2,k} = \frac{1}{n} \sum_{i=1}^n g_{ii}^2 \quad (5)$$

dimana

$$g_{ij} = (X_i - \bar{X})^T S^{-1} (X_j - \bar{X}) \quad (6)$$

$$S = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})^T \quad (7)$$

Kriteria uji :

Tolak H_0 jika $skew \geq \chi_{\alpha; \frac{k(k+1)(k+2)}{6}}^2$ dan $kurt \geq Z$

4. Homogenitas Varians

Uji homogenitas varians digunakan untuk mengetahui kehomogenan varians antar variabel. Pengujian homogenitas varians pada kasus beberapa sampel dapat menggunakan uji Bartlett (Rencher, 2002).

$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$, semua kelompok mempunyai varians yang sama (homogen)

$H_1 : \exists \sigma_i^2 \neq \sigma_j^2$ untuk $i \neq j$, terdapat kelompok yang mempunyai varians berbeda (tidak homogen)

Statistik uji :

$$b = \frac{[(s_1^2)^{n_1-1} (s_2^2)^{n_2-1} \dots (s_k^2)^{n_k-1}]^{\frac{1}{N-k}}}{s_p^2} \quad (8)$$

dimana

$$s_p^2 = \frac{\sum_{i=1}^k (n_i-1) s_i^2}{N-k} \quad (9)$$

s_i^2 adalah varians masing-masing kelompok

Kriteria uji :

Tolak H_0 jika $b < b_k(\alpha, n)$ untuk $n_1 = n_2 = \dots = n_k$ dan tolak H_0 jika $b < b_k(\alpha; n_1, n_2, \dots, n_k)$ untuk ukuran sampel yang berbeda dimana

$$b_k(\alpha; n_1, n_2, \dots, n_k) \cong \frac{n_1 b_k(\alpha, n_1) + \dots + n_k b_k(\alpha, n_k)}{N} \quad (10)$$

5. Analisis Komponen Utama

Analisis komponen utama merupakan kombinasi linear dari variabel awal dengan cara mereduksi variabel awal menjadi variabel baru yang berbentuk ortogonal dan saling bebas (Rencher, 2002). variabel baru yang terbentuk ini disebut dengan komponen utama yang dapat menjelaskan sebagian besar variasi yang terdapat dalam data. Komponen utama yang pertama adalah kombinasi linear dengan varians maksimal. Komponen utama kedua merupakan kombinasi linear dari seluruh variabel yang diamati yang bersifat ortogonal terhadap komponen utama pertama dan memiliki varians kedua terbesar, dan seterusnya. Koefisien komponen utama dapat diperoleh secara bersamaan menggunakan eigen value decomposition atau dari singular value decomposition.

Penelitian ini menggunakan singular value decomposition (SVD). Metode SVD merupakan salah satu cara yang dapat digunakan untuk mendapatkan koefisien komponen utama secara bersamaan dari matriks data (Jolliffe, 2002).

Metode SVD mendekomposisikan suatu matriks menjadi tiga komponen matriks.

$$\mathbf{X}_{(n \times k)} = \mathbf{U}_{(n \times k)} \mathbf{P}_{(k \times k)} \mathbf{V}_{(k \times k)}^t \quad (11)$$

Metode SVD ini lebih baik digunakan untuk data yang tidak berdistribusi normal multivariat.

6. Jarak Euclidean

Jarak euclid adalah jarak yang biasa digunakan untuk analisis *cluster*. Jarak ini mengukur antara jarak objek data ke pusat *cluster* (Johnson & Wichern, 2017). Secara umum jarak euclid dapat ditulis seperti berikut

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (12)$$

di mana:

d_{ij} : jarak Euclid obyek data ke-i dan obyek data ke-j

m : banyaknya peubah/ parameter yang digunakan

x_{ik} : obyek data ke-i pada peubah ke-k

x_{ij} : obyek data ke-j pada peubah ke-k

7. Analisis Cluster

Analisis *cluster* digunakan untuk mengelompokkan objek penelitian yang memiliki kemiripan karakteristik dari objek-objek yang akan diteliti menjadi sebuah kelompok (Sitepu, 2011). Terdapat dua metode dalam metode *cluster*, yaitu metode hirarki dan metode non hirarki:

1. Metode Hirarki

Metode hirarki dilakukan dengan cara mengelompokkan data berdasarkan kemiripan karakteristik dari objek, dimulai dengan objek memiliki kemiripan yang paling dekat, kemudian dilanjutkan ke objek berikutnya yang mempunyai kedekatan kedua dan seterusnya, sehingga kelompok akan membentuk semacam “pohon”

yang menunjukkan adanya hirarki yang jelas (Sitepu, 2011).

Terdapat beberapa algoritma yang digunakan dalam metode ini, yaitu *Single Linkage*, *Complete linkage*, *Average Linkage*, *Centroid Method*, *Ward's Method*, dan *Mahalanobis Distance* (Johnson & Wichern, 2017).

2. Metode Non Hirarki

Metode non hirarki berbeda dengan metode hirarki. Metode ini menetapkan objek ke dalam *cluster* setelah jumlah *cluster* ditentukan. Ada dua cara metode non hirarki, yang pertama menentukan jumlah *cluster* dan yang kedua tetapkan setiap pengamatan ke salah satu *cluster* yang telah ditentukan berdasarkan jumlah *cluster*.

8. Fuzzy C-Means Clustering

Fuzzy C-means adalah suatu teknik *clustering* data berdasarkan derajat keanggotaan (Muslimatin, 2011). Adapun algoritma yang digunakan dalam *Fuzzy C-means* adalah sebagai berikut (Sutoyo, 2015):

1. Input data yang akan di *cluster* X, berupa matriks berukuran $n \times m$ (n =jumlah sampel data, m =atribut setiap data). X_{ij} = data sampel ke- i ($i = 1, 2, \dots, n$), atribut ke- j ($j = 1, 2, \dots, m$).

2. Tentukan:

- Jumlah *cluster* = c
- Pangkat = w
- Maksimum iterasi = $Maxlter$
- Error* terkecil yang diharapkan = α
- Fungsi objektif, $P_0 = 0$
- Iterasi awal, $t = 1$

3. Bangkitkan bilangan *random* μ_{ik} , $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$; sebagai elemen-elemen matriks partisi awal U . Hitung jumlah setiap kolom:

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (13)$$

dengan $j = 1, 2, \dots, n$. Hitung:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (14)$$

4. Hitung pusat *cluster* ke- k : V_{kj} , dengan $k = 1, 2, \dots, c$; dan $j = 1, 2, \dots, m$.

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * (X_{ij}))}{\sum_{i=1}^n (\mu_{ik})^w} \quad (15)$$

5. Hitung fungsi objektif pada iterasi ke- t , P_t :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c (|\sum_{j=1}^m (X_{ij} - V_{kj})^2| (\mu_{ik})^w) \quad (16)$$

6. Hitung perubahan matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad (17)$$

dengan: $i = 1, 2, \dots, n$; dan $k = 1, 2, \dots, c$

7. Cek kondisi berhenti:

- Jika: $(|P_t - P_{t-1}| < \alpha)$ atau $(t > MaxIter)$ maka berhenti.
- Jika tidak: $t = t + 1$, ulangi langkah ke-4.

9. Langkah Penelitian

Analisis yang dilakukan dalam penelitian ini menggunakan *software* Rstudio, dengan langkah-langkah penelitian adalah

1. Memeriksa ortogonalitas antar variabel, apabila tidak ortogonal dilakukan analisis komponen utama
2. Melakukan pengujian normal multivariat pada data, apabila tidak normal multivariat maka analisis dilanjutkan menggunakan *Singular Value Decomposition* (SVD)
3. Membentuk komponen utama untuk setiap observasi

4. Menghitung matriks jarak

5. Lakukan *pengclusteran* menggunakan metode *Fuzzy C-Means Clustering*

6. Membuat karakterisasi untuk mengetahui karakteristik dari setiap kelompok.

HASIL DAN PEMBAHASAN

Langkah awal yang harus dilakukan dalam penelitian ini adalah mengecek ortogonalitas antar variabel karena dalam analisis *cluster* diperlukan ukuran jarak, dimana antar variabel harus saling ortogonal. Dari data yang sudah dinormalisasi diperoleh hasil perhitungan dari $\mathbf{Y}^t\mathbf{Y}$ adalah sebagai berikut:

$$(\mathbf{Y}^t\mathbf{Y})_{(4 \times 4)} = \begin{pmatrix} 1 & 0.9103745 & 0.8829433 & 0.9281718 \\ 0.9103745 & 1 & 0.97337646 & 0.9795451 \\ 0.8829433 & 0.97337646 & 1 & 0.9746376 \\ 0.9281718 & 0.9795451 & 0.9746376 & 1 \end{pmatrix}$$

Hasil tersebut menunjukkan $\mathbf{Y}^t\mathbf{Y} \neq \mathbf{I}$ artinya antar variabel tidak saling ortogonal. Sehingga perlu dilakukan analisis komponen utama untuk mendapatkan variabel yang saling ortogonal.

Di dalam analisis komponen utama memiliki asumsi bahwa data berdistribusi normal multivariat. Untuk itu, dilakukan pengujian normal multivariat menggunakan uji mardia. Berdasarkan pengujian dengan taraf signifikansi sebesar $\alpha = 5\%$ diperoleh hasil bahwa nilai $skew = 1.67514231046001 \times 10^{-11} < \alpha$ dan nilai $kurt = 0.0320788908153267 < \alpha$, sehingga dapat ditarik kesimpulan bahwa H_0 ditolak, artinya data nilai UN SMP/Sederajat Negeri di Kota Bandung tidak berdistribusi normal multivariat.

Selanjutnya, menguraikan matriks data menggunakan SVD sehingga diperoleh matriks U ,

Tabel 1. Nilai komponen utama untuk setiap observasi

D , dan V yang dinyatakan sebagai berikut:

$$U_{(60 \times 4)} = \begin{bmatrix} -0.16152884 & 0.225537124 & 0.269866788 & -0.143087024 \\ -0.17276224 & 0.339120135 & -0.026330138 & -0.104297290 \\ \vdots & \vdots & \vdots & \vdots \\ -0.13331580 & 0.001932555 & 0.258487909 & -0.256380908 \end{bmatrix}$$

$$D_{(4 \times 4)} = \begin{bmatrix} 967.6481 & 0 & 0 & 0 \\ 0 & 50.53633 & 0 & 0 \\ 0 & 0 & 13.46686 & 0 \\ 0 & 0 & 0 & 11.43221 \end{bmatrix}$$

$$V_{(4 \times 4)} = \begin{bmatrix} -0.5971565 & -0.7704701 & 0.19270369 & -0.1124509 \\ -0.5006919 & 0.2410804 & -0.77025304 & -0.3128867 \\ -0.4293375 & 0.5418046 & 0.60755270 & -0.3911481 \\ -0.4564877 & 0.2338885 & 0.02133701 & 0.8581724 \end{bmatrix}$$

| Nama Sekolah | KU ₁ | KU ₂ | KU ₃ | KU ₄ |
|---------------------|-----------------|-----------------|-----------------|-----------------|
| SMPN 1 Bandung | -156,3 | 11,4 | 3,63 | -1,64 |
| SMPN 2 Bandung | -167,17 | 17,14 | -0,35 | -1,19 |
| SMPN 3 Bandung | -145,12 | 7,21 | 2,56 | 1,12 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| MTsN 2 Kota Bandung | -129,003 | 0,098 | 3,48 | -2,93 |

Dekomposisi tersebut akan digunakan untuk mendapatkan nilai koefisien komponen utama. Diperoleh nilai koefisien komponen utama yaitu matriks V , sehingga bentuk komponen utama dari vektor eigen seperti berikut:

Tabel 2. Nilai RSMD dan MAD

$$KU_1 = -0.5971565 X_1 - 0.5006919 X_2 - 0.4293375 X_3 - 0.4564877 X_4$$

$$KU_2 = -0.7704701 X_1 + 0.2410804 X_2 + 0.5418046 X_3 + 0.2338885 X_4$$

$$KU_3 = 0.19270369 X_1 - 0.77025304 X_2 + 0.60755270 X_3 + 0.02133701 X_4$$

$$KU_4 = -0.1124509 X_1 - 0.3128867 X_2 - 0.3911481 X_3 + 0.8581724 X_4$$

| Jumlah Kelompok | RSMD | MAD |
|-----------------|-------|-------|
| 2 | 14,26 | 72,22 |
| 3 | 12,92 | 60,05 |
| 4 | 12,49 | 56,08 |
| 5 | 17,67 | 87,20 |
| 6 | 14,15 | 73,60 |
| 7 | 4,57 | 25,30 |
| 8 | 6,09 | 33,51 |
| 9 | 12,81 | 54,01 |
| 10 | 12,07 | 60,71 |

Keempat komponen utama tersebut sudah saling ortogonal, sehingga dari hasil komponen utama untuk setiap observasi yang disajikan pada **Tabel 1** dapat dilakukan analisis *cluster*.

Dapat lihat bahwa nilai RMSD dan MAD terkecil adalah dengan 7 kelompok. Oleh karena itu, pengelompokkan SMP/Sederajat Negeri di Kota Bandung berdasarkan nilai UN terbaik adalah menggunakan 7 kelompok.

Kemudian dilakukan pengelompokkan menggunakan analisis *Fuzzy C-Means Clustering* dengan ukuran jarak yang digunakan adalah jarak *euclidean*. Dalam penelitian ini akan dihitung nilai *Root Mean Squared Deviations* (RMSD) dan *Mean Absolute Deviation* (MAD) untuk masing-masing jumlah kelompok yang disajikan pada **Tabel 2**.

Data hasil pengelompokkan SMP/Sederajat di Kota Bandung dapat dilihat pada **Tabel 3**.

Tabel 3. Hasil pengelompokkan

| Kelompok | Anggota |
|----------|--|
| 1 | SMPN 10, SMPN 23, SMPN 25, SMPN 37, SMPN 43, SMPN 47, SMPN 49, MTSN 1 |
| 2 | SMP Negeri 1 Bandung, SMP Negeri 2 Bandung, SMP Negeri 5 Bandung, SMP Negeri 7 Bandung, SMP Negeri 8 Bandung, SMP Negeri 13 Bandung |
| 3 | SMPN Terbuka 12, SMPN Terbuka 40, SMPN Terbuka 27, SMPN Terbuka 8, SMPN Terbuka 51, SMPN Terbuka 36 |
| 4 | SMPN 3, SMPN 12, SMPN 14, SMPN 15, SMPN 28, SMPN 34, SMPN 44 |
| 5 | SMPN 4, SMPN 9, SMPN 16, SMPN 17, SMPN 18, SMPN 30, SMPN 45, MTSN 2 |
| 6 | SMPN 11, SMPN 19, SMPN 20, SMPN 22, SMPN 26, SMPN 27, SMPN 36, SMPN 46, SMPN 48, SMPN 50 |
| 7 | SMPN 6, SMPN 21, SMPN 24, SMPN 29, SMPN 31, SMPN 32, SMPN 33, SMPN 35, SMPN 38, SMPN 39, SMPN 40, SMPN 41, SMPN 42, SMPN 51, SMPN 52 |

Karakteristik pada masing-masing kelompok yang terbentuk berdasarkan nilai rata-rata UN pada masing-masing pelajaran dapat dilihat pada **Tabel 4**.

Tabel 4. Kelompok yang terbentuk berdasarkan nilai rata-rata

| Kelompok | Bahasa Indonesia | Bahasa Inggris | Matematika | IPA |
|----------|------------------|----------------|------------|-------|
| 1 | 73.24 | 56.45 | 47.58 | 52.66 |
| 2 | 86.13 | 84.25 | 77.88 | 77.61 |
| 3 | 57.16 | 45.41 | 37.44 | 39.44 |
| 4 | 81.27 | 73.86 | 64.47 | 67.17 |
| 5 | 78.28 | 66.15 | 56.93 | 59.97 |
| 6 | 76.26 | 61.85 | 51.03 | 56.23 |
| 7 | 70.62 | 53.36 | 44.11 | 48.64 |

Dari tabel di atas diketahui bahwa rata-rata nilai UN untuk semua mata pelajaran tertinggi berada di kelompok 2, sedangkan rata-rata nilai UN terendah untuk semua mata pelajaran berada pada kelompok 3.

KESIMPULAN

Hasil analisis diperoleh bahwa metode *Fuzzy C-Means Clustering* dengan 7 kelompok merupakan *cluster* terbaik dengan nilai RMSD dan nilai MAD terkecil. Berdasarkan analisis, rata-rata nilai UN untuk semua mata pelajaran tertinggi berada di kelompok 2, sedangkan rata-rata nilai UN untuk semua mata pelajaran terendah berada pada kelompok 3.

Adapun saran yang dapat penulis berikan kepada pemerintah yaitu:

1. Perlu dilakukan pengayaan kepada siswa untuk meningkatkan nilai rata-rata UN, terutama untuk sekolah-sekolah yang berada pada kelompok 3 yaitu SMPN Terbuka 12, SMPN Terbuka 40, SMPN Terbuka 27, SMPN Terbuka 8, SMPN Terbuka 51, dan SMPN Terbuka 36.
2. Mengadakan pelatihan dan dukungan tambahan untuk guru di sekolah-sekolah pada kelompok 3 guna membantu para guru dalam meningkatkan pengetahuan, keterampilan, dan pemahaman guru dalam strategi pengajaran yang efektif.

DAFTAR PUSTAKA

Damarjati. (2018, Oktober 17). *Ujian Nasional*. Retrieved from <http://psmk.kemdikbud.go.id/konten/2142/ujian-nasional>

- Johnson, R., & Wichern, D. (2017). *Applied Multivariate Statistical Analysis Sixth Edition*. Upper Saddle River, New Jersey: Prentice Hall International, Inc.
- Jolliffe, I. (2002). *Principal Component Analysis : Second Edition*. New York: Springer.
- Kemdikbud. (2019). *Laporan Hasil Ujian Nasional*. Kementerian Pendidikan dan Kebudayaan.
- Muslimatin. (2011). *Perbandingan Metode K-Means dan Metode Fuzzy C-Means (FCM untuk Clustering Data (Studi Kasus pada Data Saham Harian PT Astra TBK)*. Malang: UIN Maulana Malik Ibrahim.
- Rencher, A. (2002). *Methods of Multivariate Analysis : Second Edition*. New York: John Wiley and Sons, Inc.
- Sitepu. (2011). Analisis Cluster terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan. *Jurnal Universitas Sriwijaya*.
- Sutoyo. (2015). *Penerapan Fuzzy C-means untuk Deteksi Kemampuan Penalaran Matematis*. Semarang: Unnes.