

## PENGGEROMBOLAN SMA/MA DI KOTA PADANG BERDASARKAN INDIKATOR MUTU PENDIDIKAN DENGAN MENGGUNAKAN METODE *CLUSTER ENSEMBLE*

FITTRI RAHMI YETTI, HAZMIRA YOZZA, IZZATI RAHMI HG  
*Program Studi Matematika,  
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,  
Kampus UNAND Limau Manis Padang, Indonesia,  
fittrirahmiyeti@yahoo.com*

**Abstrak.** Penelitian ini dilakukan untuk menggerombolkan SMA/MA di Kota Padang berdasarkan standar mutu pendidikan. Metode yang digunakan adalah metode Cluster Ensemble dan data yang digunakan skor mutu pendidikan berstandar nasional hasil penilaian BAN-SM. Penelitian ini menghasilkan solusi empat gerombol, dimana gerombol 1 terdiri dari 9 SMA/MA, gerombol 2 terdiri dari 16 SMA/MA, gerombol 3 terdiri dari 4 SMA/MA dan gerombol 4 terdiri dari 8 SMA/MA. Berdasarkan penilaian dari BAN-SM, gerombol 1 merupakan SMA/MA yang berakreditasi C, gerombol 2 berakreditasi A dan B, gerombol 3 berakreditasi B dan gerombol 4 berakreditasi A.

*Kata Kunci:* Analisis gerombol, nilai reproducibility, nilai reproducibility adjusted, metode Cluster Ensemble

### 1. Pendahuluan

Standar Nasional Pendidikan adalah kriteria minimal mutu pendidikan yang harus dipenuhi oleh setiap satuan pendidikan di seluruh wilayah hukum Negara Kesatuan Republik Indonesia. Terdapat 8 standar mutu yang diatur, yaitu standar kompetensi lulusan, standar isi, standar proses, standar pendidik dan tenaga kependidikan, standar sarana dan prasarana, standar pengelolaan, standar pembiayaan dan standar penilaian pendidikan.

Mutu pendidikan setiap sekolah tidak seragam, sehingga pemerintah perlu mempertimbangkan kebijakan peningkatan mutu yang berbeda untuk tiap-tiap sekolah. Namun, membedakan kebijakan untuk tiap-tiap sekolah tidaklah mungkin. Suatu alternatif yang dapat dilakukan adalah dengan menggerombolkan sekolah secara acak menjadi beberapa gerombol dan kebijakan yang sama diberikan kepada sekolah pada gerombol yang sama. Penggerombolan tersebut dapat memudahkan pemerintah dalam menyamakan kebijakan yang akan diambil.

Analisis gerombol adalah salah satu metode statistika yang digunakan untuk menganalisis data peubah ganda. Tujuan analisis ini adalah untuk mengelompokkan  $n$  objek berdasarkan kemiripan atau ketidakmiripan karakteristiknya, sehingga objek-objek yang mirip akan berada pada gerombol yang sama dan objek-objek yang tidak mirip akan masuk ke dalam gerombol yang berbeda [1]. Penggerom-

bolan objek tersebut dilakukan berdasarkan  $k$  peubah yang diamati pada  $n$  objek.

Teknik yang sering digunakan adalah analisis gerombol berhirarki dengan berbagai metode perbaikan jarak dan analisis gerombol tak berhirarki  $k$ -means dengan berbagai pemilihan pusat gerombol awal. Kelemahan dari kedua metode ini adalah solusi yang dihasilkan oleh setiap metode berbeda-beda. Dengan kondisi tersebut, akan sulit ditentukan mana solusi terbaik yang dapat diambil dalam menggerombolkan objek.

Metode analisis yang dapat mengatasi kelemahan tersebut adalah *Cluster Ensemble*. Metode ini dilakukan dengan mengkombinasikan beberapa hasil penggerombolan yang diperoleh dari metode penggerombolan berhirarki dan tak berhirarki. Metode ini mampu meningkatkan kestabilan, kualitas dan kekekaran solusi gerombol. Suatu hal yang menarik untuk diketahui adalah bagaimana objek-objek pengamatan dapat digerombolkan dengan menggunakan metode *Cluster Ensemble* ini.

Pada penelitian ini, metode *Cluster Ensemble* menggerombolkan SMA/MA di Kota Padang berdasarkan standar mutu pendidikan hasil penilaian dari BAN-SM dan solusi gerombol yang dihasilkan akan dibandingkan dengan hasil penilaian akreditasi SMA/MA yang dikeluarkan oleh BAN-SM.

## 2. Analisis Gerombol

Analisis gerombol didasarkan pada ukuran kemiripan atau ketakmiripan antar objek. Salah satu ukuran ketakmiripan yang paling sering digunakan adalah ukuran jarak dan ukuran jarak yang sering digunakan adalah jarak *Euclid*. Penggunaan jarak *Euclid* dilakukan jika tidak ada korelasi antar peubah. Jarak ini didefinisikan sebagai berikut:

$$d_{ij} = [(X_i - X_j)^T (X_i - X_j)]^{1/2}. \quad (2.1)$$

### 2.1. Analisis Gerombol Berhirarki

Metode berhirarki digunakan jika peneliti belum mengetahui banyaknya gerombol yang akan dibentuk dan ukuran contoh relatif kecil. Terdapat dua prosedur pada metode berhirarki yaitu *agglomerative hierarchical clustering* dan *divisive hierarchical clustering*. Pada prosedur *agglomerative* proses penggerombolan dimulai dari  $N$  gerombol yang masing-masingnya terdiri dari satu objek. Selanjutnya secara bertahap, gerombol yang paling mirip akan digabungkan sampai didapatkan satu kesatuan gerombol yang terdiri dari  $N$  objek. Pada prosedur *divisive* dilakukan prosedur sebaliknya dimana proses penggerombolan dimulai dari satu gerombol yang terdiri dari  $N$  objek. Secara bertahap, gerombol dipisahkan menjadi dua gerombol sampai diperoleh  $N$  gerombol yang masing-masingnya terdiri dari satu objek.

Dalam metode *agglomerative* terdapat lima metode perbaikan jarak yang dapat digunakan, yaitu *Single Linkage*, *Complete Linkage*, *Centroid Method*, *Ward's Method*, *Average Linkage*.

### 2.2. Analisis Gerombol Tak Berhirarki

Metode tak berhirarki digunakan bila banyaknya gerombol ( $k$ ) yang akan dibentuk telah ditentukan dan ukuran contoh relatif besar. Metode tak berhirarki bertujuan untuk mengelompokkan  $n$  objek ke dalam  $k$  kelompok ( $k < n$ ). Metode tak berhirarki yang umum digunakan adalah  $k$ -rataaan ( $k$ -means). Penentuan pusat gerombol pada langkah awal sangat mempengaruhi hasil penggerombolan yang diperoleh. Beberapa metode yang dapat digunakan untuk menentukan pusat gerombol awal adalah  $k$  data awal, *Distance Based Starting Point*, *Density Based Starting Point*, *Hierarchical Starting Point*.

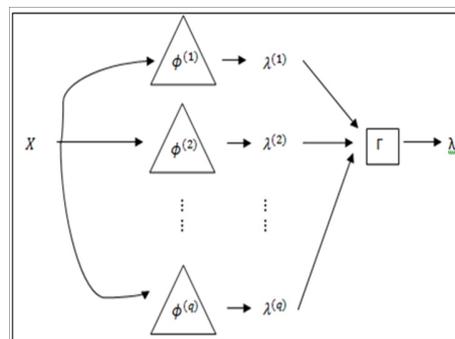
### 2.3. Cluster Ensemble

*Cluster Ensemble* adalah salah satu metode yang dapat digunakan untuk menggerombolan objek. Metode ini pertama kali diperkenalkan oleh Strehl dan Gosh pada tahun 2002. Pada *cluster ensemble*, penggerombolan dilakukan dengan mengkombinasikan berbagai solusi dari berbagai metode penggerombolan hingga diperoleh satu penggerombolan akhir yang lebih baik. Input yang dibutuhkan dalam *cluster ensemble* adalah solusi penggerombolan yang telah diperoleh dengan menggunakan berbagai hasil penggerombolan tanpa melihat karakteristik data awal. Kekekaran (*robustness*) yang tinggi, keakuratan, dan kestabilan merupakan karakteristik dari *cluster ensemble*.

Secara umum, penggerombolan objek dengan metode *cluster ensemble* dilakukan dalam dua tahap yaitu:

- (1) Membentuk anggota *ensemble* yang anggotanya adalah solusi dari berbagai metode penggerombolan yang berbeda.
- (2) Mengkombinasikan seluruh anggota ensemble untuk memperoleh satu solusi akhir dengan menggunakan suatu fungsi *consensus*. Solusi akhir tersebut dinamakan solusi *consensus*.

Kedua tahap pada *cluster ensemble* diilustrasikan pada Gambar 1, dimana



Gambar 1. Kerangka Ilustrasi *Cluster Ensemble*

$X$  : data pengamatan,  
 $\theta^{(q)}$  : metode penggerombolan ke- $q$ ,  
 $\lambda^{(q)}$  : solusi gerombol metode ke- $q$   
 $\Gamma$  : fungsi *consensus*,  
 $\lambda$  : solusi akhir.

Fungsi *consensus*  $\Gamma$  didefinisikan sebagai fungsi  $\Gamma = \{\Lambda^{(q)} | q \in 1, 2, \dots, r\} \longrightarrow \{\Lambda\}$  yang memetakan sekumpulan hasil gerombol  $\lambda^{(q)}$  yang beragam, tanpa melihat kembali  $X$  (data awal) atau  $\theta^{(q)}$  (metode penggerombolan ke- $q$ ) hingga diperoleh satu hasil penggerombolan akhir ( $\lambda$ ) yang disebut solusi *consensus*. Tujuan dari fungsi *consensus* adalah mengkombinasikan hasil gerombol dan mencari solusi gabungan yang memberikan informasi terbanyak.

Terdapat tiga algoritma yang dapat digunakan pada fungsi *consensus* yaitu [2]:

- (1) *Cluster-based Similarity Partitioning Algorithm (CSPA)*.
- (2) *HyperGraph-Partitioning Algorithm (HGPA)*.
- (3) *Meta-Clustering Algorithm (MCLA)*.

Penelitian ini akan menggunakan algoritma *meta-clustering* pada fungsi *consensus*. Berikut adalah algoritma *meta-clustering* yang dikembangkan oleh [3]:

- (1) Mentransformasi sekumpulan solusi gerombol atau anggota *ensemble* menjadi sebuah matriks indikator. Matriks indikator adalah matriks yang kolom-kolomnya menggambarkan gerombol dari setiap solusi sedangkan baris-baris matriks indikator menggambarkan objek pengamatan. Matriks ini terdiri dari angka biner 1 dan 0. Objek bernilai 1 pada kolom tertentu jika merupakan anggota gerombol yang bersesuaian dengan kolom tersebut dan bernilai 0 jika sebaliknya.
- (2) Menggerombolkan kembali objek pengamatan dengan menganggap kolom-kolom matriks indikator sebagai peubah baru yang digunakan dalam analisis gerombol. Tahap ini disebut dengan *clustering on cluster (CC)*. Tahap *clustering on cluster (CC)* menggunakan metode  $k$ -rataan sebagai analisis gerombol. Penggerombolan pada tahap ini dilakukan secara berulang dengan inisialisasi acak. Pengulangan ini dilakukan untuk mengatasi adanya pengaruh inisialisasi yang berbeda terhadap solusi yang dihasilkan. Kedua langkah ini dilakukan berulang kali hingga tidak terjadi perubahan keanggotaan gerombol yang dihasilkan.

Berikut akan dijelaskan dengan lebih rinci mengenai langkah mentransformasi hasil penggerombolan ke dalam bentuk matriks indikator, sebagaimana yang telah disebutkan pada langkah pertama dari algoritma *meta clustering*. Misalkan penggerombolan akan dilakukan terhadap 10 objek pengamatan. Misalkan pembentukan *ensemble* dilakukan dengan menggunakan tiga metode penggerombolan yang berbeda, dan diperoleh solusi sebagai berikut:

**Tabel 2.4.2** Solusi gerombol metode  $\lambda^{(i)}$

| Objek    | $\lambda^{(1)}$ | $\lambda^{(2)}$ | $\lambda^{(3)}$ |
|----------|-----------------|-----------------|-----------------|
| $x_1$    | 1               | 2               | 1               |
| $x_2$    | 1               | 2               | 1               |
| $x_3$    | 1               | 2               | 4               |
| $x_4$    | 2               | 3               | 4               |
| $x_5$    | 2               | 3               | 2               |
| $x_6$    | 3               | 1               | 2               |
| $x_7$    | 3               | 1               | 2               |
| $x_8$    | 3               | 4               | 3               |
| $x_9$    | 4               | 4               | 3               |
| $x_{10}$ | 4               | 4               | 3               |

Pembentukan matriks indikator dilakukan dengan cara sebagai berikut:

- (1) Untuk setiap solusi gerombol ke- $i$   $\Lambda^{(i)}$ , buat matriks  $H_i$  berukuran  $n \times y_i$ , di mana  $n$  adalah jumlah objek pengamatan dan  $y_i$  adalah jumlah gerombol pada solusi ke- $i$ .
- (2) Kolom ke- $j$  pada matriks  $H_i$  menunjukkan gerombol ke- $j$  pada  $\Lambda^{(i)}$ .
- (3) Pengamatan akan bernilai 1 pada kolom tertentu jika pengamatan tersebut merupakan anggota gerombol yang bersesuaian dengan kolom dan nilai 0 jika sebaliknya.
- (4) Lakukan langkah 2 – 4 untuk setiap  $\Lambda^{(i)}$ ,  $i = 1, 2, \dots, k$

Dengan langkah-langkah tersebut, matriks indikator untuk anggota *ensemble* pada tabel 2.4.2 adalah:

**Tabel 2.4.3** Matriks Indikator

| Objek    | H1 |    |    |    | H2 |    |    |    | H3 |    |    |     |
|----------|----|----|----|----|----|----|----|----|----|----|----|-----|
|          | h1 | h2 | h3 | h4 | h4 | h5 | h6 | h7 | h7 | h8 | h9 | h10 |
| $x_1$    | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0  | 0  | 0   |
| $x_2$    | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0  | 0  | 0   |
| $x_3$    | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 1   |
| $x_4$    | 0  | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 1   |
| $x_5$    | 0  | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0  | 0   |
| $x_6$    | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0   |
| $x_7$    | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0   |
| $x_8$    | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0   |
| $x_9$    | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0   |
| $x_{10}$ | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0   |

**2.4. Nilai Reproducibility**

Untuk menentukan solusi terbaik dari beragam solusi gerombol yang dihasilkan pada setiap tahap *meta-clustering*, digunakan sebuah nilai yang disebut *reproducibility*. Nilai *reproducibility* menggambarkan kekonsistenan keanggotaan objek dalam gerombol tertentu. Nilai *reproducibility* dapat digunakan sebagai kriteria untuk menentukan solusi gerombol terbaik. Nilai *reproducibility* tertinggi pada suatu gerombol tertentu menunjukkan bahwa solusi gerombol tersebut merupakan solusi gerombol terbaik.

Nilai *reproducibility* dapat diperoleh, dengan cara berikut:

- Tabulasi silang solusi ke- $i$  dengan solusi gerombol ke- $j$  untuk  $i \neq j$

**Tabel 2.4.4** Tabulasi Silang Solusi Gerombol

| Solusi 1 | Solusi 2 |         |         |         | Total |
|----------|----------|---------|---------|---------|-------|
|          | Group 1  | Group 2 | Group 3 | Group 4 |       |
| Group 1  | 0        | 3       | 0       | 0       | 3     |
| Group 2  | 0        | 0       | 2       | 0       | 2     |
| Group 3  | 2        | 0       | 0       | 1       | 3     |
| Group 4  | 0        | 0       | 0       | 2       | 2     |
| Total    | 2        | 3       | 2       | 3       | 10    |

Tabel 2.4.4 menunjukkan tabulasi silang antara solusi  $\lambda^{(1)}$  dan  $\lambda^{(2)}$ . Penamaan gerombol yang bersifat *arbitrary* (tidak ada aturan pasti) sehingga penamaan gerombol pada solusi yang satu belum tentu sama dengan yang lain. Misal solusi  $\lambda^{(1)}$  dan  $\lambda^{(2)}$  yang menghasilkan solusi yang sama namun tampak berbeda. Perbedaan tersebut menimbulkan masalah ketika dilakukan tabulasi silang antar solusi sehingga mempengaruhi nilai *reproducibility*. Untuk mengatasi hal tersebut perlu dilakukan penukaran posisi kolom ke posisi yang bersesuaian dengan posisi baris yang mempunyai nilai terbesar pada kolom tersebut sehingga dapat memaksimalkan jumlah diagonal. Setelah dilakukan penukaran posisi kolom maka tabulasi silang antara solusi  $\lambda^{(1)}$  dan  $\lambda^{(2)}$  berubah seperti Tabel 2.4.5.

**Tabel 2.4.5** Tabulasi Silang Solusi Gerombol Setelah Penukaran Posisi Kolom

| Solusi 1 | Solusi 2 |         |         |         | Total |
|----------|----------|---------|---------|---------|-------|
|          | Group 1  | Group 2 | Group 3 | Group 4 |       |
| Group 1  | 3        | 0       | 0       | 0       | 3     |
| Group 2  | 0        | 2       | 0       | 0       | 2     |
| Group 3  | 0        | 0       | 2       | 1       | 3     |
| Group 4  | 0        | 0       | 0       | 2       | 2     |
| Total    | 3        | 2       | 2       | 3       | 10    |

- Selanjutnya dihitung  $R_{ij}$  = jumlah diagonal/jumlah objek Sebagai contoh jumlah objek yang konsisten berada pada gerombol tertentu pada solusi  $\lambda^{(1)}$  dan  $\lambda^{(2)}$  adalah  $(3+2+2+2=9)$  dan nilai *reproducibility* berpasangan solusi  $\lambda^{(1)}$  dan  $\lambda^{(2)}$  adalah 0,9.
- Nilai *reproducibility* total solusi gerombol ke- $i$  adalah

$$R_i = \frac{(R_{i1} + R_{i2} + \dots + R_{ij})}{j} \quad (2.2)$$

Sebagai contoh untuk mengetahui nilai *reproducibility* total solusi maka perlu dihitung nilai *reproducibility* solusi  $\lambda^{(1)}$  dan  $\lambda^{(2)}$  sehingga diperoleh nilai *reproducibility* total solusi  $\lambda^{(1)}$  sebesar  $R_1 = \frac{(R_{12}+R_{13})}{2}$ .

Solusi yang memiliki nilai *reproducibility* total tertinggi dipilih sebagai solusi terbaik. Nilai *reproducibility* dihitung untuk membandingkan solusi dengan jumlah gerombol yang sama. Untuk membandingkan solusi dengan jumlah gerombol yang berbeda, dihitung menggunakan nilai *reproducibility adjusted*.

Nilai *reproducibility adjusted* dapat diperoleh dengan cara berikut:

$$RA = \frac{[(kR) - 1]}{k - 1} \quad (2.3)$$

dengan RA adalah nilai *reproducibility adjusted*,  $k$  adalah banyak gerombol dan  $R$  adalah nilai *reproducibility*. Nilai RA tertinggi pada jumlah gerombol tertentu menunjukkan bahwa jumlah gerombol tersebut mampu mencirikan struktur data terbaik.

Selanjutnya  $F$  rasio digunakan untuk membandingkan nilai kuadrat tengah antara kelompok dengan nilai kuadrat tengah dalam kelompok dari solusi gerombol yang diperoleh.  $F$  rasio dapat dihitung dengan rumus

$$F = \frac{MSB}{MSW} \quad (2.4)$$

dengan

$$MSB = \frac{SSB}{dfb} \quad \text{dan} \quad MSW = \frac{SSW}{dfw} \quad (2.5)$$

dimana

$$dfb = k - 1 \quad (2.6)$$

$$dfw = n - k \quad (2.7)$$

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad (2.8)$$

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 \quad (2.9)$$

$$SSB = SST - SSW \quad (2.10)$$

### 3. Metode Penelitian

Data yang digunakan pada penelitian ini adalah skor mutu pendidikan berstandar nasional hasil penilaian BAN-SM terhadap SMA/MA di Kota Padang. Tahapan yang dilakukan dalam penelitian ini adalah:

- (1) Membentuk anggota *ensemble* dengan cara menggerombolkan data dengan menggunakan berbagai metode penggerombolan (metode hirarki dan tak berhirarki), dengan berbagai metode perbaikan jarak dan penentuan pusat gerombol awal serta dengan berbagai jumlah gerombol.
- (2) Dari berbagai solusi yang diperoleh, dihitung nilai *reproducibility* dan ditentukan solusi terbaik sementara dengan 2,3 dan 4 gerombol. Solusi terbaik inilah yang akan menjadi acuan pemberhentian metode penggerombolan pada tahap sebelumnya.
- (3) Mentransformasikan anggota *ensemble* ke dalam bentuk matriks indikator.
- (4) *Cluster on clustering*, yakni menggerombolkan kembali objek berdasarkan kolom-kolom pada matriks indikator. Penggerombolan dilakukan dengan menggunakan metode  $k$ -rataaan dengan berbagai metode pemilihan titik pusat gerombol awal dan berbagai jumlah gerombol.

- (5) Hitung kembali nilai *reproducibility* dari setiap solusi yang diperoleh, solusi dengan nilai *reproducibility* tertinggi adalah solusi terbaik sementara yang dapat diambil.
- (6) Bila solusi yang diperoleh pada tahap lima ini sama dengan solusi terbaik sementara yang diperoleh pada tahap sebelumnya, maka metode penggerombolan *cluster ensemble* dihentikan. Jika tidak, lakukan kembali langkah 4 – 5 hingga tidak terjadi perubahan keanggotaan gerombol dan solusi sebelumnya dijadikan solusi akhir gerombol.
- (7) Lakukan perhitungan *reproducibility adjusted* untuk memilih solusi gerombol mana yang paling baik sebagai solusi akhir *consensus*.

#### 4. Hasil dan Pembahasan

Dalam penelitian ini, data yang digunakan data sekunder yang bersumber dari BAN-SM. Objek yang digunakan pada penelitian ini adalah SMA/MA di Kota Padang, daftar objek dapat dilihat pada Tabel 2.4.5.

**Tabel 2.4.5** Daftar Objek Pengamatan

| No | Nama Sekolah | No | Nama Sekolah    | No | Nama Sekolah      | No | Nama Sekolah            |
|----|--------------|----|-----------------|----|-------------------|----|-------------------------|
| 1  | SMAN 1       | 11 | SMAN 11         | 21 | SMA Don Bosco     | 31 | SMA dr. ahmad PGAI      |
| 2  | SMAN 2       | 12 | SMAN 12         | 22 | SMA Yapi          | 32 | SMA PGRI 4              |
| 3  | SMAN 3       | 13 | SMAN 13         | 23 | SMA SEMEN         | 33 | SMA Kristen Kalam kudus |
| 4  | SMAN 4       | 14 | SMAN 14         | 24 | SMA murni         | 34 | Man 1                   |
| 5  | SMAN 5       | 15 | SMAN 15         | 25 | SMA Bunda         | 35 | Man 2                   |
| 6  | SMAN 6       | 16 | SMAN 16         | 26 | SMA PGRI 6        | 36 | Man 3                   |
| 7  | SMAN 7       | 17 | SMAN 17         | 27 | SMA Bukit Barisan | 37 | Man PGAI                |
| 8  | SMAN 8       | 18 | SMA Pertiwi 1   | 28 | SMA SIMA          | 38 | Man Limau Manis         |
| 9  | SMAN 9       | 19 | SMA Kartika 1-5 | 29 | SMA pembangunan   |    |                         |
| 10 | SMAN 10      | 20 | SMA Don Bosco   | 30 | MAS MTI           |    |                         |

Peubah yang digunakan adalah peubah standar mutu pendidikan, dari masing-masing peubah terdapat perbedaan bobot penilaian sehingga data yang digunakan adalah hasil kali antara skor setiap item dengan bobotnya. Berikut akan ditampilkan daftar peubah dan bobot masing-masing peubah pada Tabel 2.4.6.

**Tabel 2.4.6** Skor Penilaian Sekolah

| No | Standar Mutu                 | Komponen Bobot | Bobot Butir | Kode |
|----|------------------------------|----------------|-------------|------|
| 1  | Standar Isi                  | 15             | 1,00        | X1   |
| 2  | Standar Proses               | 10             | 1,00        | X2   |
| 3  | Standar Kompetensi lulusan   | 10             | 0,40        | X3   |
| 4  | Standar Pendidik dan Tendik  | 15             | 0,75        | X4   |
| 5  | Standar Sarana dan Prasarana | 15             | 0,50        | X5   |
| 6  | Standar Pengelolaan          | 10             | 0,50        | X6   |
| 7  | Standar Pembiayaan           | 15             | 0,60        | X7   |
| 8  | Standar Penilaian            | 10             | 0,50        | X8   |

Anggota *ensemble* diperoleh dengan menggerombolkan objek-objek menggunakan metode penggerombolan berhirarki (*Complete Linkage* dan *Average Linkage*) dan metode *k-means* (*Distance Based Starting Point*, *Density Based Starting Point*, dan *Hierarchical Starting Point*). Setiap metode menghasilkan solusi 2, 3 dan 4 gerombol. Setelah membangun anggota *ensemble* maka tahap selanjutnya adalah membentuk solusi *consensus* dengan mengkombinasikan seluruh anggota *ensemble* dengan menggunakan algoritma *meta-clustering*. Solusi *consensus* yang terbentuk adalah sebagai berikut.

- Solusi 2 gerombol  
Solusi 2 gerombol yang terbentuk adalah 13 sekolah pada gerombol 1 dan 24 sekolah pada gerombol 2. Pada solusi 2 gerombol ini, variabel standar penilaian pendidikan merupakan variabel yang paling membedakan kedua gerombol ini. Sedangkan, variabel standar pendidik dan tenaga kependidikan adalah yang paling tidak membedakan kedua gerombol. Nilai *reproducibility* pada solusi dua gerombol adalah 0,8290 dan nilai *reproducibility adjusted* pada solusi dua gerombol adalah 0,658.
- Solusi 3 gerombol  
Solusi 3 gerombol yang terbentuk adalah 9 sekolah pada gerombol 1, 24 sekolah pada gerombol 2 dan 3 sekolah pada gerombol 3. Variabel standar sarana dan prasarana merupakan variabel yang paling membedakan ketiga gerombol ini. Sedangkan, variabel standar pendidik dan tenaga kependidikan adalah variabel yang paling tidak membedakan ketiga gerombol. Nilai *reproducibility* pada solusi 3 gerombol adalah 0,768 dan nilai *reproducibility adjusted* pada solusi 3 gerombol adalah sebesar 0,652.
- Solusi 4 gerombol  
Solusi 4 gerombol yang terbentuk adalah Gerombol 1 terdiri dari 9 sekolah, gerombol 2 terdiri dari 16 sekolah, gerombol 3 terdiri dari 4 sekolah dan gerombol 4 terdiri dari 8 sekolah. Variabel standar sarana dan prasarana merupakan variabel yang paling membedakan keempat gerombol ini. Sedangkan, variabel standar isi adalah variabel yang paling tidak membedakan keempat gerombol. Nilai *reproducibility* pada solusi 4 gerombol adalah 0,745 dan nilai *reproducibility adjusted* solusi 4 gerombol adalah 0,66.

Dari ketiga solusi gerombol diperoleh nilai *reproducibility adjusted* sebesar 0,658 untuk solusi 2 gerombol, 0,652 untuk solusi 3 gerombol dan 0,66 untuk solusi dengan 4 gerombol, solusi yang memiliki nilai *reproducibility adjusted* tertinggi adalah solusi terbaik yang dapat dipilih. Dari nilai *reproducibility adjusted* yang diperoleh dapat disimpulkan bahwa solusi penggerombolan terbaik adalah solusi dengan 4 gerombol, karena memiliki nilai *reproducibility adjusted* yang paling tinggi. Berikut akan ditampilkan nilai rata-rata dari gerombol terbaik yang terbentuk.

**Tabel 2.4.8** Rata-rata Standar Mutu Setiap Gerombol

| Standar              | Gerombol   |            |            |            |
|----------------------|------------|------------|------------|------------|
|                      | Gerombol 1 | Gerombol 2 | Gerombol 3 | Gerombol 4 |
| Isi                  | 72.93      | 87.60      | 75.41      | 92.59      |
| Proses               | 69.22      | 82.64      | 61.13      | 92.5       |
| Kompetensi Lulus     | 59.33      | 81.39      | 67         | 90.81      |
| Tendik dan Pendidik  | 69.78      | 81.48      | 80.94      | 94.42      |
| Sarana dan prasarana | 50.26      | 77.96      | 79.59      | 90.52      |
| Pengelolaan          | 65.17      | 85.09      | 69.13      | 93.25      |
| Pembiayaan           | 70.11      | 87.97      | 77         | 93.83      |
| Penilaian Pendidikan | 76.56      | 87.77      | 77.44      | 95.47      |

Dari tabel dapat dilihat bahwa, gerombol 1 memiliki nilai rata-rata standar mutu yang lebih kecil dari nilai rata-rata gerombol lainnya, anggota gerombol 1 adalah SMA/MA yang berakreditasi C. Sedangkan gerombol 2 memiliki nilai rata-rata standar mutu yang lebih besar dari gerombol 1 dan gerombol 3, anggota gerombol 2 merupakan SMA/MA yang berada pada selang batas atas B dan berada pada selang batas bawah A, ini berarti anggota dari gerombol 2 merupakan SMA/MA berakreditasi A dan B. Selanjutnya, anggota dari gerombol 3 merupakan SMA/MA yang berakreditasi B, terlihat dari nilai rata-rata standar mutu pendidikan yang lebih besar dari gerombol 1 dan lebih kecil dari gerombol 2. Nilai rata-rata tertinggi dapat di lihat pada gerombol 4 dengan anggota gerombol merupakan SMA/MA yang berakreditasi A.

## 5. Kesimpulan

Pada metode ini, penentuan solusi terbaik dilakukan dengan suatu ukuran yang dinamakan sebagai nilai *reproducibility* dan nilai *reproducibility adjusted*. Metode ini menghasilkan solusi 2,3 dan 4 gerombol dengan nilai *reproducibility adjusted* masing-masing sebesar 0,658 ; 0,652 dan 0,66. Dapat disimpulkan bahwa solusi yang paling baik pada penelitian ini adalah solusi 4 gerombol, dimana anggota dari gerombol 1 merupakan SMA/MA yang berakreditasi C, gerombol 2 SMA/MA yang berakreditasi A dan B, gerombol 3 SMA/MA yang berakreditasi B dan gerombol 4 SMA/MA yang berakreditasi A.

## 6. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Bapak Dr. Dodi Devianto, Bapak Dr. Admi Nazra, dan Ibu Dr. Maiyastri, yang telah memberikan masukan dan saran sehingga paper ini dapat diselesaikan dengan baik.

## Daftar Pustaka

- [1] Johnsons, R. A. dan Winchern, D. W. 2002. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall
- [2] Strehl, A. dan Gosh, J. 2002. Cluster Ensembles-A Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research* **3**: 583 – 617
- [3] Orme, B. dan Johnsons, R. 2008. Improving K-Means Cluster Analysis: Ensemble Analysis Instead of Highest Reproducibility Replicates. Sawtooth software
- [4] Jolliefe, I.T. 2002. *Principal Component Analysis*. New York: Spinger-Verla

- [5] Megawati. 2010. Penggerombolan Kabupaten di Jawa Barat berdasarkan Indikator Ketahanan dan Kerentanan Pangan menggunakan Cluster Ensemble. Bogor: Institut Pertanian Bogor
- [6] Anggiani, D. 2009. Penerapan Cluster Ensemble dalam Penggerombolan Kecamatan di Kabupaten Bogor. Bogor: Institut Pertanian Bogor
- [7] Retzer, J. dan Shan, M. 2007. Cluster Ensemble Analysis and Graphical Depiction of Cluster Partitions. *Proceedings of the 2007 Sawtooth Software Conference: California*, 17 – 19 October 2007. Sequim WA : 2008. 239 — 250
- [8] Hu, X dan Yoo, I. 2004. *Cluster Ensemble and its Applications in Gene Expression Analysis*. Philadelphia: Drexel University
- [9] Heryanti, Vivi. 2010. Cluster Ensemble dalam Menggerombolkan Varietas Padi. Bogor: Institut Pertanian Bogor