

# IMPLEMENTASI METODE RAPID CENTROID ESTIMATION DAN K-MEANS UNTUK KLASTERISASI SEKOLAH BERDASARKAN FASILITAS (STUDI KASUS SEKOLAH NEGERI KABUPATEN MALANG)

Deddy Kusbianto<sup>1</sup>, Irsyad Arif Mashudi<sup>2</sup>, Fransdhinta Dewana Prianggoro<sup>3</sup>

<sup>1,2,3</sup> Teknik Informatika, Teknologi Informasi, Politeknik Negeri Malang

<sup>1</sup>[deddy\\_kusbianto@polinema.ac.id](mailto:deddy_kusbianto@polinema.ac.id), <sup>2</sup>[irsyad.arif@polinema.ac.id](mailto:irsyad.arif@polinema.ac.id), <sup>3</sup>[dhintaF@gmail.com](mailto:dhintaF@gmail.com)

**Abstrak**— Sebagai fasilitas pembentuk karakter bangsa sekolah diharuskan memenuhi kebutuhan sarana dan prasarana penunjang proses belajar siswa. Faktanya banyak sekolah Negeri di Kabupaten Malang yang belum bisa memenuhi kebutuhan tersebut, seperti tidak adanya perpustakaan dan laboratorium yang seharusnya dapat dimanfaatkan siswa untuk menambah pengetahuan dan mengembangkan potensi diri mereka di sekolah, hal ini disebabkan kurang meratanya distribusi fasilitas pada setiap sekolah yang dapat mengakibatkan proses belajar siswa tidak maksimal. Permasalahan tersebut dapat diatasi dengan mengetahui sekolah mana saja yang perlu dilakukan pemerataan, untuk itu perlu dilakukan pengelompokan sekolah-sekolah di Kabupaten Malang berdasarkan fasilitas yang dimiliki sekolah tersebut. Pengelompokan dapat dilakukan dengan cara klasterisasi. Pada penelitian ini akan dilakukan klasterisasi sekolah dengan menggunakan kombinasi metode RCE (Rapid Centroid Estimation) dan K-Means. Untuk menjalankan metode yang digunakan pada penelitian ini dibuat sebuah sistem klasterisasi Sekolah Negeri di Kabupaten Malang. Pada penelitian ini didapatkan rata-rata nilai pengujian Silhouette Coefficient sebesar 0,351 pada hasil klasterisasi sekolah dasar di kecamatan Bululawang, nilai Silhouette Coefficient 0,399 untuk seluruh data SMP, nilai Silhouette Coefficient 0,488 untuk seluruh data SMA dan nilai Silhouette Coefficient 0,468 untuk seluruh data SMK, yang menunjukkan hasil klasterisasi yang baik, karena jika nilai Silhouette Coefficient lebih mendekati nilai 1 dan menjauhi nilai -1 menunjukkan data tepat di cluster tersebut.

**Kata kunci**— RCE, Rapid Centroid Estimation, K-Means, Silhouette Coefficient, Klasterisasi.

## I. PENDAHULUAN

Sebagai salah satu fasilitas membentuk karakter bangsa, sekolah menjadi tempat penting dimana para pelajar Indonesia diharapkan dapat berjuang membawa negara bersaing di kancah global. Dunia pendidikan di Indonesia memiliki beberapa kendala diantaranya adalah keterbatasan akses pada pendidikan dan fasilitas sekolah yang belum merata. Contoh untuk fasilitas pada sekolah yang berlokasi

di luar perkotaan, kelengkapan fasilitas dapat disebut kurang sehingga berpotensi menyebabkan proses belajar siswa terganggu, yang dimaksud fasilitas dalam hal ini adalah guru, ruang kelas, ruang laboratorium dan ruang perpustakaan untuk pemerataan guru juga belum merata. Terbatasnya akses pendidikan di Indonesia, terlebih lagi di daerah berujung pada meningkatnya arus perpindahan untuk mendapat akses ilmu yang lebih baik di perkotaan, yang dapat mengakibatkan penumpukan siswa sehingga akan banyak sekolah yang kekurangan siswa. Hal ini menunjukkan bahwa permasalahan ada pada pemerataan fasilitas dan jumlah guru pada tiap sekolah.

Permasalahan diatas dapat diatasi dengan cara mengetahui sekolah mana saja yang perlu dilakukan pemerataan, untuk itu pemerintah perlu mengelompokkan sekolah-sekolah yang ada di Kabupaten Malang. Pengelompokan sekolah dapat dilakukan dengan cara klasterisasi, dimana nantinya sekolah dengan karakteristik yang sama dikelompokkan menjadi satu cluster, sehingga jika terdapat sekolah yang fasilitasnya kurang maka sekolah dengan fasilitas serupa akan terkelompok dalam cluster yang sama. Pada penelitian ini peneliti melakukan klasterisasi dengan menggunakan kombinasi metode RCE dan K-Means karena data yang akan digunakan adalah data numeric yang tidak memiliki label, dan selanjutnya data cluster akan diurutkan menggunakan metode MOORA untuk mengurutkan peringkat cluster, sehingga cluster dengan rata-rata nilai MOORA yang tinggi menjadi cluster sekolah dengan fasilitas yang lengkap dan cluster dengan rata-rata nilai MOORA yang rendah menjadi cluster sekolah dengan fasilitas kurang.

Metode K-Means merupakan metode clustering yang paling sederhana dan umum, (Rahma, Arhandi, & Firdausi, 2020). K-Means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien, (Rahma, Arhandi, & Firdausi, 2020). Tetapi, K-Means mempunyai kelemahan

yang diakibatkan oleh penentuan pusat awal cluster, Hasil cluster yang terbentuk dari metode K-Means ini sangat tergantung pada inisiasi nilai pusat awal cluster yang diberikan, (Rahma, Arhandi, & Firdausi, 2020). Untuk meningkatkan efisiensi penentuan pusat awal cluster K-Means, peneliti menggunakan metode Rapid Centroid Estimation. Hasil evaluasi clustering dengan kombinasi K-Means dan RCE, mampu memperbaiki dalam hal peningkatan hasil iterasi klasterisasi data. Penentuan titik pusat clustering menyebabkan terjadinya penurunan iterasi dalam cluster data dibandingkan dengan titik pusat cluster secara random (Selvida, 2019).

Metode Multi-Objective Optimization by Ratio Analysis atau biasa disingkat dengan metode MOORA adalah metode yang memiliki perhitungan dengan kalkulasi yang minimum dan sangat sederhana. Metode ini memiliki tingkat selektifitas yang baik karena dapat menentukan tujuan dari kriteria yang bertentangan, dimana kriteria dapat bernilai menguntungkan (benefit) atau yang tidak menguntungkan (cost), (Rokhman, Rozi, & Asmara, 2017).

Untuk membantu melakukan klasterisasi sekolah berdasarkan fasilitas pada sekolah negeri Kabupaten Malang, diperlukan sebuah sistem untuk menjalankan metode klasterisasi. Pada penelitian ini akan dibuat sistem klasterisasi sekolah berdasarkan fasilitas pada sekolah negeri Kabupaten Malang dengan mengimplementasikan metode Rapid Centroid Estimation dan K-Means

## II. DATA DAN METODE PENELITIAN

### A. Data

Metode pengambilan data digunakan untuk mengumpulkan data pendukung yang dibutuhkan dalam pembuatan aplikasi. pengumpulan data yang diperoleh dengan cara Mengakses website [dapo.dikdasmen.kemdikbud.go.id](http://dapo.dikdasmen.kemdikbud.go.id) kemudian memilih data sekolah Kabupaten malang, lalu mengunduh file csv yang berisi data sekolah masing-masing kecamatan.

### B. Metode Penelitian

#### 1) Rapid Centroid Estimation (RCE)

Rapid Centroid Estimation berbasis pada algoritma Particle Swarm Clustering (PSC), namun dikonfigurasi ulang untuk mengurangi kerumitan komputasi (Selvida, 2019). RCE mampu mencapai kinerja dengan stabilitas yang lebih tinggi dan kecepatan optimasi yang lebih cepat dari pada PSC (Selvida, 2019). Adapun pseudocode algoritma RCE adalah sebagai berikut (Selvida, 2019):

- a) Algorithm  $S = RCE(\text{dataset}, \text{jumlah\_cluster})$
- b) Inisialisasikan  $n_c$  partikel, random  $x$
- c) Hitung jarak  $p, g$  untuk setiap partikel kesetiap datum.
- d) While  $t < \text{max\_iterasi}$
- e) Update matrik jarak

$$D(t) = d(y_j, x_i): \forall i, j \quad (1)$$

- f) Temukan titik data terdekat dari setiap partikel

$$[Dx^{min}(t) \quad lx^{min}] = \min(D, i) \quad (2)$$

- g) Temukan setiap partikel terdekat untuk setiap data

$$[Dy^{min}(t) \quad ly^{min}] = \min(D, y) \quad (3)$$

- h) Tentukan partikel pemenang (Partikel yang paling dekat terhadap pola inputan dengan menggunakan rumus Euclidean distance)

$$x_{most\ win}(t) = x(t) \in \min(d(p_i(t) - x_i(t))): \forall i \quad (4)$$

- i) **For** untuk setiap partikel  $x$

- j) Tentukan elemen dari setiap anggota partikel  $i$  (centroid)

$$y_i^{cluster} = \forall y \in x_i(t) \quad (5)$$

$$N_i = size(y_i^{cluster}) \quad (6)$$

- k) Update posisi dengan menggunakan langkah (g) jika  $N_i$  lebih besar dari nol, jika tidak langsung menuju koordinat partikel pemenang.

$$x_i(t+1) = \begin{cases} x_i + \Delta x_i(t+1) & \text{jika } N_i > 0 \\ x_i(t) + \phi_s \otimes (x_{most\ win}(t) - x_i(t)) & \text{jika tidak} \end{cases} \quad (7)$$

- l) **End For**

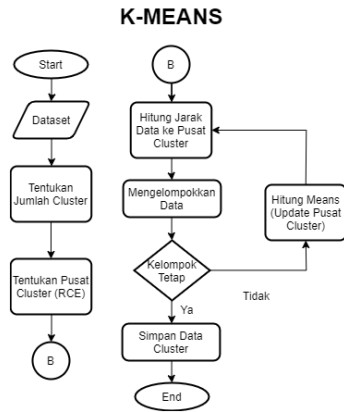
$$m) t = t + 1 \quad (8)$$

- n) **End While.**

#### 2) K-Means

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok.

Pengelompokan K-Means merupakan metode analisis kelompok yang mengarah pada pemartisian  $N$  objek pengamatan ke dalam  $K$  kelompok (cluster), dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan rata-rata terdekat. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dikelompokkan ke dalam satu kelompok yang sama data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Tujuan pengelompokan ini adalah untuk meminimalkan fungsi objek yang diset dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antarkelompok.



Gambar 1 Flowchart K-Means

### 3) MOORA

Metode Multi-Objective Optimization by Ratio Analysis (MOORA) adalah metode yang diperkenalkan oleh (Rokhman, Rozi, & Asmara, 2017). Metode yang relative baru ini pertama kali digunakan oleh Brauers dalam suatu pengambilan keputusan dengan multi-kriteria. Metode MOORA memiliki tingkat fleksibilitas dan kemudahan untuk dipahami dalam memisahkan bagian subjektif dari suatu proses evaluasi kedalam kriteria bobot keputusan. Metode ini memiliki tingkat selektifitas yang baik karena dapat menentukan tujuan dari kriteria yang bertentangan. Dimana kriteria dapat bernilai menguntungkan (benefit) atau yang tidak menguntungkan (cost).

Metode MOORA terdiri dari lima langkah utama sebagai berikut:

- Menentukan tujuan dan mengidentifikasi atribut evaluasi yang bersangkutan.
- Menampilkan semua informasi yang tersedia untuk atribut dalam bentuk matriks keputusan.  $x$  adalah nilai kriteria masing-masing kriteria yang dipresentasikan sebagai matriks.

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdot & x_{1n} \\ x_{21} & x_{22} & \cdot & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{m1} & x_{m2} & \cdot & x_{mn} \end{bmatrix}$$

- Menentukan skor ternormalisasi dari masing-masing alternative untuk tiap kriteria ( $x^*_{ij}$ ), dengan persamaan:

$$x^*_{ij} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^m x_{ij}^2}} \quad (9)$$

- Menghitung nilai optimasi multiobjektif yaitu: kriteria pada masing-masing alternative diberikan nilai bobot kepentingan, maka pemberian nilai bobot pada kriteria, menentukan skor terbobot ( $y_{ij}$ ), dengan cara

mengalikan skor ternormalisasi dengan bobot dari kriteria ( $w_j$ ), menggunakan persamaan:

$$y_{ij} = w_j \times x^*_{ij} \quad (10)$$

Dengan ketentuan nilai bobot jenis kriteria maximum lebih besar dari nilai bobot jenis kriteria minimum. Perkalian bobot kriteria terhadap nilai atribut maximum dikurangi perkalian bobot kriteria terhadap nilai atribut minimum, jika dirumuskan maka:

$$y_j^* = \sum_{i=1}^{i=g} w_j x^*_{ij} - \sum_{i=g+1}^{i=n} w_j x^*_{ij}, \quad (11)$$

Keterangan :  $i = 1, 2, \dots, g$ -kriteria/atribut dengan status maximized;

$i = g+ 1, g+ 2, \dots, n$ - kriteria/ atribut dengan status minimized;

$W_j$  = bobot terhadap  $j$

$y_i$  = Nilai penilaian yang telah

dinormalisasi dari alternatif 1 terhadap semua atribut.

- Menentukan nilai rangking dari hasil perhitungan ( $Y_i$ ).

### 4) Silhouette Coefficient

Silhouette Coefficient merupakan metode evaluasi untuk menguji optimal atau ketepatan sebuah cluster yang telah terbentuk dari proses clustering. Silhouette Coefficient memberikan hasil kualitas visual objek dalam tiap cluster, memberikan informasi sesuai dengan jumlah cluster pada dataset. Metode ini gabungan dari metode separation dan cohesion. Tahapan perhitungan Silhouette Coefficient adalah sebagai berikut :

- Hitung rata-rata jarak dari suatu data, maka didapatkan rata-rata dengan cara memisalkan  $i$  terhadap semua data lain yang berada dalam satu cluster sebagai berikut :

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (12)$$

dimana :

$a(i)$  = Perbedaan rata-rata objek ( $i$ ) ke semua objek lain pada A.

A = Cluster.

$d(i,j)$  = jarak antara data I dengan j.

- Hitung rata-rata jarak data I tersebut dengan semua data di Cluster lain, dan diambil nilai terkecilnya.

$$d(i, c) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (13)$$

dimana :

$d(i,C)$  = Perbedaan rata-rata objek ( $i$ ) ke semua objek lain pada C

C = cluster lain selain cluster A atau cluster C tidak sama dengan cluster A.

- Setelah menghitung  $d(i,C)$  untuk semua C, maka diambil nilai terkecil dengan menggunakan persamaan sebagai berikut:

$$b(i) = \min_{C \neq A} d(i, C). \quad (14)$$

d) Nilai Silhouette Coefficient didefinisikan seperti persamaan berikut:

$$s(i) = \frac{(b(i) - a(i))}{\max(b(i), a(i))} \quad (15)$$

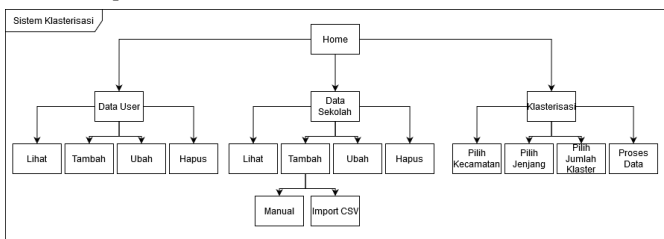
s(i) selalu terletak antara nilai -1 sampai dengan 1. Nilai s(i) dapat dijelaskan sebagai berikut :

Tabel 1 Nilai Silhouette Coefficient

Nilai s(i)	Keterangan
Mendekati -1	Klasterisasi objek (i) belum tepat
0	objek I terletak diantara cluster
Mendekati 1	Klasterisasi objek (i) sudah tepat

### III. PERANCANGAN

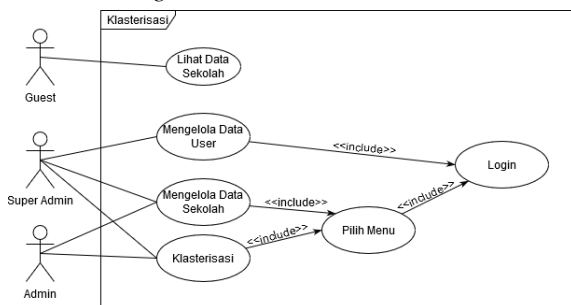
#### A. Deskripsi Sistem



Gambar 2. Deskripsi Sistem

Pada sistem klasterisasi ini terdapat dua menu utama dalam sistem yaitu Data Sekolah dan Klasterisasi yang dapat diakses oleh Admin dan Super Admin. Menu Data User hanya dapat diakses oleh Super Admin. Dalam menu Data Sekolah Admin dapat mengelola Data Sekolah yang nantinya data dimanfaatkan untuk proses klasterisasi. Dalam menu Klasterisasi Admin dan Super Admin dapat melakukan Klasterisasi dengan memilih data mana yang akan diproses dan setelah itu hasilnya ditampilkan dalam bentuk tabel.

#### B. Use Case Diagram



Gambar 3. Deskripsi Sistem

Dari use case diatas dapat dilihat ada tiga macam pengguna, yaitu guest, admin dan super admin, untuk guest hanya dapat melihat data sekolah tanpa login kedalam sistem. Admin dapat mengolah data yaitu lihat, tambah, ubah dan hapus. Selain itu Admin juga dapat melakukan Klasterisasi data. Super Admin sama dengan Admin dengan tambahan fitur kelola data user.

#### C. Sampel Data

Sampel data yang digunakan dalam penelitian adalah data Sekolah Negeri Kabupaten Malang seperti di bawah ini:

No.	Nama Sekolah	Rasio PD:RB (X1)	Rasio G:RB (X2)	Rasio R,Kelas:RB (X3)	Rasio Lab:PD (X4)	Rasio Perpus:PD (X5)
1	SD NEGERI 1 BULULAWANG	24.143	1.286	1.143	0,006	0,006
2	SD NEGERI 1 KREBET	30.500	1.500	1.500	0,008	0,004
3	SD NEGERI 1 KREBET SENGRONG	15.833	0,833	1,000	0	0
4	SD NEGERI 1 SUDIMORO	19.333	1.333	1,000	0	0
5	SD NEGERI 1 WANDANPURO	24.333	1.667	1.667	0	0
6	SD NEGERI 2 BAKALAN	23.333	1.333	1,000	0,007	0
7	SD NEGERI 2 BULULAWANG	13.000	1,000	1,000	0,013	0,013
8	SD NEGERI 2 KASEMBON	15.333	0,833	1,000	0	0,011
9	SD NEGERI 2 KASRI	18.000	1.167	1.167	0	0,009
10	SD NEGERI 2 KREBET SENGRONG	35.000	1.667	1,000	0	0,005
11	SD NEGERI 2 LUMBANGSARI	27.167	1.167	1.333	0	0
12	SD NEGERI 2 SUDIMORO	17.333	1,000	1,000	0	0
13	SD NEGERI 2 WANDANPURO	29.400	1.400	1,000	0	0,002
14	SD NEGERI 3 KREBET	8.000	1.333	1,000	0	0
15	SD NEGERI 3 LUMBANGSARI	24.167	1.333	1,000	0	0
16	SD NEGERI 3 SUDIMORO	21.417	1.167	1.167	0	0
17	SD NEGERI 4 WANDANPURO	34.333	1.167	1,000	0,005	0,005
18	SD NEGERI GADING	17.833	1.333	1,000	0	0
19	SD NEGERI KUWOLU	8.500	0,5	1,000	0	0,002
20	SD NEGERI PRINGO	21.500	1,000	1,000	0,016	0,016

Gambar 4. Sampel Data

Data pada Gambar 4 adalah data mentah sebelum dilakukan preprocessing, dari data diatas perlu dilakukan preprocessing agar tidak terdapat kolom yang mendominasi, preprocessing dilakukan dengan cara normalisasi data sehingga hasilnya seperti berikut ini :

No.	Nama Sekolah	Rasio PD:RB (X1)	Rasio G:RB (X2)	Rasio R,Kelas:RB (X3)	Rasio Lab:PD (X4)	Rasio Perpus:PD (X5)
1	SD NEGERI 1 BULULAWANG	0,598	0,673	0,214	0,382	0,302
2	SD NEGERI 1 KREBET	0,833	0,857	0,750	0,529	0,209
3	SD NEGERI 1 KREBET SENGRONG	0,290	0,286	0,000	0,000	0,000
4	SD NEGERI 1 SUDIMORO	0,420	0,714	0,000	0,000	0,000
5	SD NEGERI 1 WANDANPURO	0,605	1,000	1,000	0,000	0,000
6	SD NEGERI 2 BAKALAN	0,568	0,714	0,000	0,461	0,000
7	SD NEGERI 2 BULULAWANG	0,185	0,429	0,000	0,827	0,654
8	SD NEGERI 2 KASEMBON	0,272	0,286	0,000	0,000	0,554
9	SD NEGERI 2 KASRI	0,370	0,571	0,250	0,000	0,472
10	SD NEGERI 2 KREBET SENGRONG	1,000	1,000	0,000	0,000	0,243
11	SD NEGERI 2 LUMBANGSARI	0,710	0,571	0,500	0,000	0,000
12	SD NEGERI 2 SUDIMORO	0,346	0,429	0,000	0,000	0,000
13	SD NEGERI 2 WANDANPURO	0,793	0,771	0,000	0,000	0,116
14	SD NEGERI 3 KREBET	0,000	0,714	0,000	0,000	0,000
15	SD NEGERI 3 LUMBANGSARI	0,599	0,714	0,000	0,000	0,000
16	SD NEGERI 3 SUDIMORO	0,497	0,571	0,250	0,000	0,000
17	SD NEGERI 4 WANDANPURO	0,975	0,571	0,000	0,313	0,248
18	SD NEGERI GADING	0,364	0,714	0,000	0,000	0,000
19	SD NEGERI KUWOLU	0,019	0,000	0,000	0,000	1,000
20	SD NEGERI PRINGO	0,500	0,429	0,000	1,000	0,791

Gambar 5. Sampel Data Preprocessing

### IV. UJI COBA DAN PEMBAHASAN

#### A. Uji Coba Perhitungan Metode

Berikut hasil perhitungan metode RCE dan K-Means dengan jumlah klaster 3 pada sampel data yang digunakan :

Klaster 1						
Rata - rata Silhouette Coefficient : 0.3836339830508865						
Nama Sekolah	Silhouette Coefficient	Rasio PD:RB	Rasio G:RB	Rasio R,Kelas:RB(X3)	Rasio Lab:PD(X4)	Rasio Perpus:PD(X5)
SD NEGERI 2 KASEMBON	0.15188478221319868	15.333	0.833	1.000	0.000	0.011
SD NEGERI KUWOLU	0.4494881302731176	8.500	0.500	1.000	0.000	0.020
SD NEGERI PRINGO	0.4553921356344493	21.500	1.000	1.000	0.016	0.016
SD NEGERI 2 BULULAWANG	0.47777088408278023	13.000	1.000	1.000	0.013	0.013

Gambar 6. Hasil Klaster 1

Klaster 2						
Rata - rata Silhouette Coefficient : 0.4225173459936577						
Nama Sekolah	Silhouette Coefficient	Rasio PD:RB	Rasio G:RB	Rasio R,Kelas:RB(X3)	Rasio Lab:PD(X4)	Rasio Perpus:PD(X5)
SD NEGERI 1 KREBET SENGRONG	0.4515863632480745	15.833	0.833	1.000	0.000	0.000
SD NEGERI 2 SUDIMORO	0.5395806096119905	17.333	1.000	1.000	0.000	0.000
SD NEGERI 3 LUMBANGSARI	0.5325231737781396	24.167	1.333	1.000	0.000	0.000
SD NEGERI 2 WANDANPURO	0.44700034954205453	29.400	1.400	1.000	0.000	0.002
SD NEGERI 1 SUDIMORO	0.5640664248211072	19.333	1.333	1.000	0.000	0.000
SD NEGERI GADING	0.55902853503902	17.833	1.333	1.000	0.000	0.000
SD NEGERI 2 KREBET SENGRONG	0.27250287473550816	35.000	1.667	1.000	0.000	0.005
SD NEGERI 3 SUDIMORO	0.36982868330189334	21.417	1.167	1.167	0.000	0.000
SD NEGERI 4 WANDANPURO	0.3256652839882907	34.333	1.167	1.000	0.005	0.005

SD NEGERI 3 KREBET	0.44479222161729787	8.000	1.333	1.000	0.000	0.000
SD NEGERI 2 BAKALAN	0.4058249683262882	23.333	1.333	1.000	0.007	0.000
SD NEGERI 2 KASRI	0.3012940936820063	18.000	1.167	1.167	0.000	0.009
SD NEGERI 1 BULULAWANG	0.2790319162258788	24.143	1.286	1.143	0.006	0.006

Gambar 7. Hasil Kluster 2

Kluster 3						
Rata - rata Silhouette Coefficient : 0.4689231275447095						
Nama Sekolah	Silhouette Coefficient	Rasio PD:RB	Rasio G:RB	Rasio R.Kelas:RB(X3)	Rasio Lab:PD(X4)	Rasio Perpus:PD(X5)
SD NEGERI 2 LUMBANGSARI	0.28424833479177825	27.167	1.167	1.333	0.000	0.000
SD NEGERI 1 WANDANPURO	0.5894907025444375	24.333	1.667	1.667	0.000	0.000
SD NEGERI 1 KREBET	0.5330303452979128	30.500	1.500	1.500	0.008	0.004

Gambar 8. Hasil Kluster 3

Gambar 6, 7 dan 8 adalah hasil perhitungan metode RCE dan K-Means yang telah diimplementasikan pada sistem menggunakan Bahasa pemrograman python.

### B. Pembahasan

Pembahasan dilakukan untuk menguji akurasi hasil klusterisasi apakah data sudah tepat berada dalam *cluster* tersebut dengan menggunakan metode *Silhouette Coefficient*.

Kluster 1	Rata - rata Silhouette Coefficient : 0.3836339830508865
Kluster 2	Rata - rata Silhouette Coefficient : 0.4225173459936577
Kluster 3	Rata - rata Silhouette Coefficient : 0.4689231275447095

Gambar 9. Rata – rata Silhouette Coefficient

Semakin nilai *Silhouette Coefficient* mendekati nilai 1 berarti akurasi pengelompokan kluster baik, jika nilai *Silhouette Coefficient* mendekati nilai -1 berarti pengelompokan kluster kurang baik.

## V. PENUTUP

### A. Kesimpulan

Berdasarkan penelitian dan uji coba yang telah dilaksanakan oleh penulis pada Klusterisasi Sekolah Negeri Kabupaten Malang menggunakan metode Rapid Centroid Estimation dan K-means maka dapat ditarik kesimpulan bahwa :

- 1) Metode rapid Centroid Estimation dan K-Means dapat digunakan untuk melakukan klusterisasi sekolah Negeri Kabupaten Malang.
- 2) Metode RCE dan K-Means dapat melakukan klusterisasi Sekolah Negeri kabupaten Malang dengan hasil klusterisasi yang baik, rata-rata nilai *Silhouette Coefficient* 0,351 untuk klusterisasi data sekolah dasar Kecamatan Bululawang, nilai *Silhouette Coefficient* 0,399 untuk seluruh data SMP, nilai *Silhouette Coefficient* 0,488 untuk seluruh data SMA dan nilai *Silhouette Coefficient* 0,468 untuk seluruh data SMK, di mana nilai *Silhouette Coefficient* lebih mendekati nilai 1 dan lebih menjauhi nilai -1, artinya rata-rata pengelompokan data adalah benar.

- 3) Sistem Klusterisasi dapat menjalankan metode RCE dan K-Means serta hasil Klusterisasi dapat menjadi pertimbangan dalam pemerataan fasilitas Sekolah Negeri kabupaten Malang.

### B. Saran

Berdasarkan penelitian dan uji coba aplikasi ini, ada beberapa hal yang disarankan, yaitu :

- 1) Data sekolah dapat ditambahkan atribut dari sarana dan prasarana lain supaya pengelompokan dapat menjadi lebih detail.
- 2) Sistem dapat dikembangkan menggunakan metode klusterisasi lain yang diharapkan mendapat hasil klusterisasi yang lebih baik.
- 3) Sistem dapat dikembangkan menggunakan teknologi dan bahasa pemrograman lain sehingga proses dapat berjalan lebih baik.

### DAFTAR PUSTAKA

- [1] I. F. Rahma, P. P. Arhandi dan A. T. Firdausi, "Penerapan Metode Hierarchical Clustering Dan K-Means Clustering Untuk Mengelompokkan Potensi Lokasi Penjualan Linkaja," Jurnal Informatika Polinema, 2020.
- [2] D. Selvida, "Analisis Klasifikasi Data dengan Kombinasi Metode K-Means dan Rapid Centroid Estimation," Repositori Institusi USU, 2019.
- [3] S. Rokhman, I. F. Rozi dan R. A. Asmara, "Pengembangan Sistem Penunjang Keputusan Penentuan UKT Mahasiswa Dengan Menggunakan Metode MOORA Studi Kasus Politeknik Negeri Malang," Jurnal Informatika Polinema, 2017.
- [4] Athifaturrofifah, R. Goejantoro dan D. Yuniarti, "Perbandingan Pengelompokan K-Means dan K-Medoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas," EKSPONENSIAL Volume 10, 2019.
- [5] R. L. Rohmah, "Zonasi Daerah Terdampak Bencana Angin Puting Beliung Menggunakan K-Means Clustering Dengan Analisis Silhouette Coefficient, Davies Bouldin Index dan Purity," 2019.
- [6] Putra, Angger Kartyasa Pribadi; Purwanto, Yudha; Novianty, Astri, "Analisis Sistem Deteksi Anomali Trafik Menggunakan Algoritma CURE (Clustering Using Representatives) dengan Koefisien Silhouette dalam validasi Clustering," e-Proceeding of Engineering, 2015.



- [7] Puspitasari, Novianti; Havaluddin, "Penerapan Metode K-Means dalam Pengelompokan Curah hujan Di Kalimantan Timur".
- [8] E. Muningsih, "Prosiding Seminar Nasional Elinvo," Optimasi Jumlah Cluster K-Means dengan Metode Elbow Untuk Pemetaan Pelanggan, 2017.
- [9] S. Rustam, H. A. Santoso dan S. Catur, "Optimasi K-Means Clustering untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization Di Kota Semarang," ILKOM, 2018.
- [10] D. N. batubara, A. P. Windarto, D. Hartama dan H. Satria, "Analisis Metode K-Means Pada Pengelompokan Keberadaan Area resapan Air Menurut Provinsi," SENSASI, 2019.
- [11] M. S. Dewi, A. P. Windarto, I. S. Demanik dan H. Satria, "Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah," SENSASI, 2019.
- [12] Ramadiani, F. P. Rani, D. M. Khairina dan H. R. Hatta, "Sistem Pendukung Keputusan Pemilihan Pramuka Pandega Berprestasi Menggunakan Metode Multi Objective Optimization On The Basis of Ratio Analysis," JTIK, 2018.
- [13] M. A. Amri, A. P. Windarto, A. Wanto dan I. S. Damantik, "Analisis Metode K-Means Pada Pengelompokan Perguruan Tinggi menurut Provinsi Berdasarkan Fasilitas yang dimiliki Desa," KOMIK, 2019.
- [14] S. Wardani dan R. A. Parlina Iin, "Analisis Perhitungan Metode MOORA dalam Pemilihan Supplier Bahan Bangunan Di Toko Megah Gracindo Jaya," Jurnal Nasional Informatika dan Teknologi Jaringan, 2018.
- [15] E. Muningsih dan S. Kiswati, "Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok barang," Jurnal Bianglala Informatika, 2015.
- [16] K. U. S. b. Lukman, M. Zulfahmi dan A. A. Nababan, "Perbandingan RapidCentroid Estimation (RCE) - K Nearest Neighbor (K-NN) dengan K Means - K Nearest Neighbor (K-NN)," Jurnal Nasional Informatika dan Jaringan, 2017.
- [17] Sapriadi, "Peningkatan Kinerja K-Means Menggunakan Kombinasi Principal Component Analysis dan Rapid Centroid Estimation," Repositori USU, 2019.