# OPTIMALISASI JUMLAH *CLUSTER* DATA SEKOLAH DASAR (SD) MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING*

**Siti Maryam [1], Rini Astuti [2], Fadhil M. Basysyar [3]**
[1] Teknik Informatika, STMIK IKMI Cirebon
[2,3] Sistem Informasi, STMIK LIKMI Bandung
Jl. Perjuangan No. 10B, Majasem, Cirebon
*sitimaryam01823@gmail.com*

**ABSTRAK**

Ketersediaan fasilitas belajar dan distribusi tenaga pendidik yang merata di setiap sekolah berperan penting dalam menjamin penyelenggaraan pendidikan yang berkualitas. Penelitian ini bertujuan untuk mengelompokkan Sekolah Dasar (SD) di Kabupaten Ciamis berdasarkan kesamaan karakteristik menggunakan *K-Means Clustering*. Saat ini data SD belum dikelompokkan sehingga perencanaan peningkatan mutu pendidikan belum maksimal. Oleh karena itu dibutuhkan klasterisasi agar pemahaman mengenai pola dan kesamaan karakteristik antar SD lebih mendalam. Metode yang digunakan adalah klasterisasi dengan Algoritma *K-Means Clustering*, menggunakan *tools* RapidMiner. Parameter yang digunakan meliputi *Davies Bouldin Index* (DBI), *Measure Types Numerical Measures* jenis *Euclidean Distance* dan *Manhattan Distance*, serta melakukan iterasi menggunakan parameter *Max Optimization* dimulai dari 1 hingga 20 kali. Penentuan *cluster* dimulai dari K= 2 hingga K= 10 untuk mendapatkan hasil *cluster optimal* berdasarkan nilai DBI terendah mendekati 0. Dari hasil analisis diketahui bahwa *cluster optimal* berada pada K= 10 dengan nilai DBI 0.450 menggunakan parameter *Measure Types Numerical Measures* jenis *Manhattan Distance* dan proses iterasi berhenti pada iterasi ke 10, karena pada iterasi tersebut nilai *centroid* dan keanggotaan pada setiap *cluster* sudah stabil. Hasil dari penelitian ini akan memberikan pemahaman yang lebih mendalam mengenai sekolah-sekolah dengan karakteristik yang serupa.

***Kata kunci:*** *K-Means Clustering, Davies Bouldin Index (DBI),* Iterasi, *Numerical Measures, Euclidean Distance, Manhattan Distance*

## 1. PENDAHULUAN

Setiap warga negara berhak memperoleh pendidikan yang layak [1] .Ketersediaan fasilitas belajar dan penyebaran tenaga pendidik yang merata di setiap sekolah memiliki peran penting dalam memastikan penyelenggaraan pendidikan [2]. Oleh karena itu, memastikan kualitas pendidikan pada jenjang Sekolah Dasar (SD) menjadi prioritas utama. Meskipun terdapat banyak data SD di Kabupaten Ciamis, belum ada pendekatan analisis yang menyeluruh untuk mengelompokan sekolah-sekolah berdasarkan ciri-ciri dan performa. Kekurangan dalam mengidentifikasi pola-pola khusus dalam data dapat menjadi kendala dalam meningkatkan kualitas pendidikan diwilayah ini. Untuk meningkatkan mutu pendidikan pada jenjang SD, pemahaman yang mendalam tentang karakteristik sekolah, penyebaran tenaga pendidik dan ketersediaan fasilitas sangat diperlukan guna mendukung berjalannya kegiatan belajar mengajar [3].

Saat ini data Sekolah Dasar (SD) di Kabupaten Ciamis belum dikelompokan berdasarkan kesamaan karakteristiknya, sehingga perencanaan program pendidikan dasar dalam upaya peningkatan mutu pendidikan belum dapat disesuaikan secara tepat dengan kondisi setiap *cluster*. Oleh karena itu, diperlukan klasterisasi data SD di Kabupaten Ciamis dengan menggunakan algoritma *K-Means Clustering* untuk pengelompokan berdasarkan karakteristik yang serupa.

Penelitian terdahulu di bidang pendidikan dan analisis data telah menunjukan bahwa metode klasterisasi dengan algoritma *K-Means Clustering* dapat digunakan untuk mengelompokan berbagai data berdasarkan kesamaan karakteristiknya [4]. Metode ini mampu memaksimalkan kesamaan karakteristik dalam satu *cluster* dan memaksimalkan perbedaan antar *cluster*, sehingga dihasilkan pengelompokan data yang diinginkan [5]. Namun, pada penelitian ini akan lebih difokuskan pada penerapan metode klasterisasi pada data Sekolah Dasar (SD) di Kabupaten Ciamis.

Tujuan utama dari penelitian ini adalah pengelompokan Sekolah Dasar (SD) di Kabupaten Ciamis kedalam *cluster-cluster* yang memiliki kesamaan karakteristik dengan menggunakan metode algoritma *K-Means Clustering*. Hasil dari penelitian ini dapat dijadikan dasar pengambilan keputusan terkait pendampingan dan pemberian kebutuhan sekolah dalam perencanaan dan pengembangan pendidikan di Kabupaten Ciamis.

Metode penelitian ini menggunakan pendekatan analisis data klasterisasi menggunakan algoritma *K-Means Clustering* berdasarkan nilai *Davies Bouldin Index* (DBI), dan *Measure types Numerical Measures* dengan jenis *Euclidean Distance*, dan *Manhattan Distance*, kemudian iterasi juga digunakan dalam proses klasterisasi. Algoritma *K-Means Clustering* dipilih karena mampu mengelompokkan objek berdasarkan kesamaan karakteristik dalam satu *cluster* [6]. Selanjutnya, dilakukan pengujian akurasi dengan metode *Davies Bouldin Index* (DBI) untuk

mengevaluasi performa terbaik klasterisasi dalam menghasilkan *cluster-cluster* yang optimal [7].

Hasil dari penelitian ini akan memberikan pemahaman yang lebih mendalam mengenai sekolah-sekolah dengan karakteristik yang serupa. Informasi ini diharapkan dapat memberikan dampak positif bagi pemerintah daerah, lembaga pendidikan, serta membantu dalam upaya peningkatan kualitas pendidikan. Selain itu, hasil penelitian ini juga memiliki potensi sebagai dasar bagi penelitian lanjutan dalam analisis data dengan menggunakan metode klasterisasi.

## 2. TINJAUAN PUSTAKA

Penelitian terdahulu di bidang pendidikan dan analisis data telah menunjukan bahwa metode klasterisasi dengan algoritma *K-Means Clustering* dapat digunakan untuk mengelompokan berbagai data berdasarkan kesamaan karakteristiknya. Metode ini mampu memaksimalkan kesamaan karakteristik dalam satu *cluster* dan memaksimalkan perbedaan antar *cluster*, sehingga dihasilkan pengelompokan data yang diinginkan [5]. Namun, pada penelitian ini akan lebih difokuskan pada penerapan metode klasterisasi pada data Sekolah Dasar (SD) di Kabupaten Ciamis.

Dalam penelitian yang dilakukan oleh [1], klasterisasi sekolah dilakukan dengan menggunakan algoritma *K-Means Clustering* pada tingkat SD dan SMP. Atribut yang dianalisis meliputi fasilitas (bentuk pendidik, status sekolah, jumlah ruang perpus, jumlah WC guru perempuan, jumlah WC siswa laki-laki, jumlah ruang kelas), guru dan tenaga pendidik. Hasil penelitian ini menghasilkan 3 kelompok (*cluster)*, dengan *Davies Bouldin Index* (DBI) senilai -0,695.

Penelitian lain yang dilakukan oleh [8], melakukan klasterisasi SMP berdasarkan minat siswa menggunakan algoritma *K-Means Clustering* berdasarkan jumlah rombel (rombongan belajar), jumlah peserta didik, jumlah data guru dan tenaga pendidik. Hasil penelitian ini mengidentifikasi 3 *cluster* yaitu *Cluster* 0 yang kurang diminati, *Cluster* 1 yang cukup diminati, dan *Cluster* 2 yang sangat diminati siswa dalam memilih sekolah.

Penelitian yang dilakukan oleh [4], melakukan klasterisasi dalam menentukan kriteria pemilihan Sekolah Dasar menggunakan algoritma *K-Means Clustering* berdasarkan kriteria lokasi sekolah, sejarah keluarga, dan pendidikan orangtua dan lokasi rumah.

Penelitian yang dilakukan [9] membahas tentang klasterisasi Sekolah Negeri di Kabupaten Malang berdasarkan fasilitas dengan metode RCE (*Rapid Centroid Estimation*) dan K-Means. Tujuan dari penelitian ini adalah untuk mengelompokan sekolah-sekolah di Kabupaten Malang sehingga distribusi sarana dan prasarana yang mendukung proses pembelajaran menjadi lebih merata. Hasil dari penelitian ini mengungkapkan rata-rata nilai *Silhouette Coefficient* adalah 0.351 untuk pengelompokan SD di Kecamatan Bululawang, 0.399 untuk seluruh data SMP, 0.488 untuk seluruh data SMA, dan 0.468 untuk seluruh data SMK.

Penelitian yang dilakukan [10] membahas tentang pengelompokan Sekolah Dasar di Kota Samarinda berdasarkan indikator standar pelayanan minimal pendidikan dengan tujuan untuk mengidentifikasi Sekolah Dasar kedalam tiga kategori yaitu capaian yang baik, sedang dan kurang. Hasil dari penelitian ini mengungkap bahwa tiga *cluster*, di mana 194 sekolah yang masuk dalam kategori baik, 17 sekolah kategori sedang, dan 11 sekolah yang masuk dalam kategori kurang. Hasil ini diperoleh dengan nilai *squared error* sebesar 32.2505.

## 3. METODE PENELITIAN

Metode analisis data yang digunakan dalam penelitian ini adalah klasterisasi K-Means. Algoritma K-Means merupakan salah satu algoritma klastering paling populer karena kesederhanaan dan efisiensinya.

Metode K-Means digunakan untuk menentukan jumlah klaster. Selain itu, metode ini juga digunakan untuk menentukan pusat (*centroid*) dari masing-masing *cluster*. Sebagai contoh, ketika melakukan pengelompokan data sekolah dasar berdasarkan Status, PD, Rombel, Guru, Tendik, R.Kelas, R.Lab dan R.Perpus, maka terbentuk *cluster-cluster* yang berisi sekolah dengan karakteristik atribut yang serupa.

Penelitian ini juga melibatkan perhitungan jarak antara titik data dengan *centroid* menggunakan rumus tertentu. Secara keseluruhan, metode *K-Means clustering* diaplikasikan dalam penelitian ini untuk pengelompokan data sekolah dasar secara sederhana dan efisien berdasarkan kesamaan karakteristik atributnya.

Berikut Rumus yang digunakan [11] dalam penelitian ini:

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^{p} (x_{2j} - x_{1j})^2}$$

Gambar 1. Rumus Perhitungan *Centroid Euclidean Distance*

$$D_{L_1}(x_2, x_1) = \|x_2 - x_1\|_1 = \sum_{j=1}^{p} |x_{2j} - x_{1j}|$$

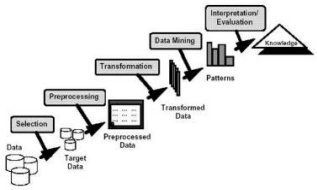
Gambar 2. Rumus Perhitungan *Centroid Manhattan Distance*

Dimana :

DL2 = jarak kuadrat Eucliden antar objek ke x2 dengan x1.
P = jumlah variabel cluster.
X2j = nilai atau data dari objek ke-2 pada variabel ke-j.
X1j = nilai atau data dari objek ke-1 pada variabel ke-j

Implementasi *Knowledge Discovery In Database* (KDD) dalam penelitian ini dilakukan dengan tujuan mendapatkan pemahaman yang mendukung dalam proses pengambilan keputusan. Seperti yang telah diuraikan oleh Yuli Mardi, KDD adalah pendekatan

yang digunakan untuk mengidentifikasi dan memperoleh wawasan dari data dalam basis data untuk mendukung pengambilan keputusan [1]. Berikut adalah langkah-langkah KDD seperti pada Gambar 3, sebagai berikut:




Gambar 3. KDD

### 3.1. *Data Selection*

Pada tahap seleksi data (*selection)*, dilakukan identifikasi atribut yang akan digunakan dalam proses *data mining*. Data awal memiliki 11 atribut, namun dipilih menjadi 9 atribut untuk proses selanjutnya, yaitu: Nama Sekolah, Status, PD, Rombel, Guru, Tendik, R.Kelas, R.Lab dan R.Perpus. Atribut Nama Sekolah dijadikan *type* ID agar tidak diproses dalam perhitungan, karena merupakan atribut identitas objek penelitian untuk mengetahui hasil klasterisasi.

### 3.2. *Data Preprocessing*

Pada tahap *data preprocessing* akan dilakukan pembersihan data dan mempersiapkan data untuk analisis, termasuk mengatasi nilai-nilai yang hilang atau tidak valid.

### 3.3. *Data Transformation*

Transformasi data mengacu pada proses mengubah tipe data sehingga sesuai dengan kebutuhan analisis yang akan dilakukan. Dalam penelitian ini transformasi data dilakukan dengan mengubah data *non numeric* menjadi *numerical* pada atribut status.

### 3.4. *Data Mining*

Tahap *data mining* adalah inti dari proses KDD. Pada penelitian ini proses *data mining* menggunakan *Algoritma K-Means Clustering.* Selain itu, akan dilakukan evaluasi performa menggunakan *Davies Bouldin Index* (DBI), dengan tujuan untuk mengetahui jumlah *cluster* optimal. Pada penelitian ini juga menggunakan parameter *Numerical Measures* jenis *Euclidean Distance*, dan Manhattan *Distance.* Kemudian dilakukan proses iterasi untuk mendapatkan hasil nilai *centroid* dan anggota tiap *cluster* tidak berubah.

### 3.5. *Interpretation/ Evaluation*

Pada tahap *Interpretation/Evaluation*, penulis akan menganalisis hasil klasterisasi guna memahami dan relevansi temuan-temuan yang diperoleh.

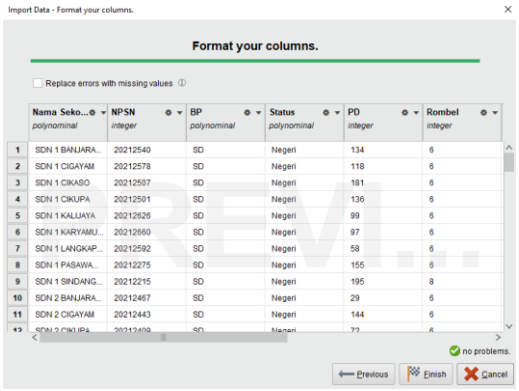
## 4. HASIL DAN PEMBAHASAN

Analisis data menggunakan teknik *Knowledge Discovery In Database* (KDD).

### 4.1. *Data Selection*

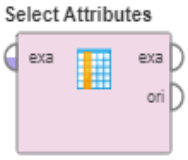
Pada tahap ini menggunakan operator *Read Excel* untuk meng *import* data berformat excel yang tersimpan di komputer pengguna ke dalam proses pengolahan data yang sedang berlangsung di perangkat lunak RapidMiner. Untuk parameter yang digunakan adalah parameter *default.* Pada Gambar 4 merupakan operator *Read Excel*.


Gambar 4. *Read Excel*

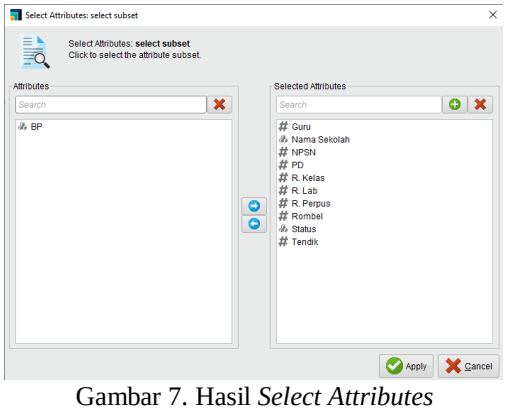



Gambar 5. Hasil Read Excel

Kemudian operator *Select Attributes* digunakan untuk menentukan atribut yang dipilih dalam pengolahan data di perangkat lunak RapidMiner. Pada Gambar 6 menunjukkan operator *Select Attributes*.

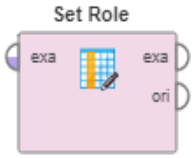



Gambar 6 *Select Attributes*




Gambar 7. Hasil *Select Attributes*

Kemudian, untuk menentukan ID pada dataset, digunakan operator *Set Role* seperti pada Gambar 8.




Gambar 8. *Set Role*

Parameter pada operator *Set Role* yang digunakan, seperti pada Tabel 1, dimana atribut Nama Sekolah dijadikan sebagai ID.

Tabel 1. Operator *Set Rol*

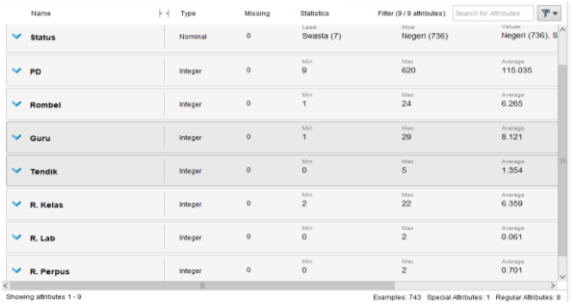
| No | Parameter | Isi |
|---|---|---|
| 1 | *Attribute Name* | Nama Sekolah |
| 2 | *Target Role* | ID |

| Row No. | Nama Sekol... | Status |
|---|---|---|
| The position of the example in the (filtered) view on the | | |
| 2 | SDN 1 CIGAY... | Negeri |
| 3 | SDN 1 CIKASO | Negeri |
| 4 | SDN 1 CIKUPA | Negeri |
| 5 | SDN 1 KALIJ... | Negeri |

Gambar 9. Hasil *Set Role*

### 4.2. *Data Preprocessing*

Karena tidak ada data yang hilang atau bernilai kosong pada dataset seperti pada Gambar 10, maka tidak perlu dilakukan tahap *preprocessing*.

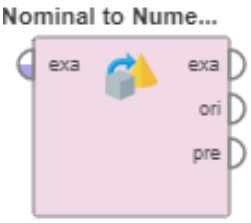


Gambar 10. Statistik Dataset

Berdasarkan hasil yang ditampilkan, tidak ditemukan adanya *missing value* pada dataset. Dengan demikian, proses analisis dapat dilanjutkan ke tahap selanjutnya.

### 4.3. *Data Transformation*

Operator *Nominal to Numeric* seperti pada Gambar 11, ini digunakan untuk mengubah tipe data atribut non-numerik menjadi numerik.




Gambar 11. *Nominal to Numerical*

Parameter pada operator *Nominal to Numerical* yang digunakan, seperti pada Tabel 3, dimana atribut "Status" akan diubah menjadi tipe numerik.

Tabel 2 Operator *Nominal to Numerical*

| No | Parameter | Isi |
|---|---|---|
| 1 | *Attribute Filter Type* | *Subset* |
| 2 | *Select Attributes* | Status |
| 3 | *Coding Type* | *Unique Integers* |

| Row No. | Nama Sekol... | Status |
|---|---|---|
| 1 | SDN 1 BANJ... | 0 |
| 2 | SDN 1 CIGAY... | 0 |
| 3 | SDN 1 CIKASO | 0 |
| 4 | SDN 1 CIKUPA | 0 |
| 5 | SDN 1 KALIJ... | 0 |

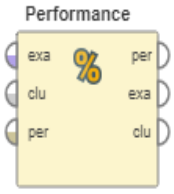
Gambar 12. Hasil Operator *Nominal to Numerical*

### 4.4. *Data Mining*

Penentuan jumlah *cluster* merupakan langkah yang harus dilakukan. Nilai *cluster* atau k yang digunakan ditentukan berdasarkan jumlah *cluster* optimal dan jumlah anggota optimal dalam tiap *cluster*. Dalam proses ini, Gambar 13 merupakan operator utama yang digunakan dalam pemodelan proses klasterisasi.

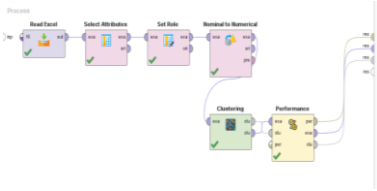

Gambar 13. *K-Means Clustering*

Selanjutnya, operator *Performance (Cluster Distance Performance)* seperti pada Gambar 14, ditambahkan untuk mengukur performa dari operator *K-Means*, dengan menggunakan metode *Davies Bouldin Index* (DBI). Hal ini dilakukan dengan tujuan untuk mendapatkan informasi mengenai nilai DBI hasil dari proses *clustering*.




Gambar 14 *Performance*

Proses pemodelan pada tahap *K-Means Clustering*, dapat dilihat pada Gambar 15.



Gambar 15 Pemodelan *K-Means Clustering*

Penelitian ini didasarkan pada perbandingan nilai Davies Bouldin Index (DBI) dengan metode K-Means Clustering, menggunakan Measure Types Numerical Measures jenis Euclidean Distance, dan Manhattan Distance serta dilakukan iterasi untuk mendapatkan nilai performa terbaik dalam pengolahan data. Pengujian diawali dengan nilai K=2 hingga K=10.

Hasil analisis menunjukan *cluster* dengan nilai DBI yang paling mendekati 0 diperoleh pada K=10 yaitu sebesar 0.450. Nilai ini mengindikasikan performa klasterisasi yang cukup baik dalam menghasilkan *cluster-cluster* yang optimal.

Selanjutnya, parameter *Measure Types Numerical Measures* jenis *Manhattan Distance* menghasilkan pengelompokan terbaik dengan nilai DBI mendekati 0 yaitu 0.450. Sementara pada *Measure Types Numerical Measures* jenis *Euclidean Distance* menghasilkan nilai DBI yaitu 0.494.

Hasil analisis juga menunjukan bahwa pada iterasi ke 10, nilai *centroid* dan jumlah anggota *cluster* sudah stabil tanpa ada perubahan lebih lanjut. Hal ini mengindikasikan hasil klasterisasi yang sudah optimal.. Dengan demikian dapat disimpulkan bahwa pembagian data menjadi 10 *cluster* memiliki akurasi yang lebih baik.

Pada Gambar 16 menunjukkan nilai *Davies Bouldin Index* (DBI) sebesar 0,450.


Gambar 16. DBI

Pada iterasi ke-17, nilai *centroid* dan anggota setiap *cluster* tidak mengalami perubahan atau sudah stabil. Pada Gambar 15 menunjukkan nilai *centroid* pada iterasi ke-10 untuk setiap *cluster* berdasarkan atribut-atribut yang digunakan dalam analisis.

Gambar 17. Nilai *Centroid*

Pada Tabel 5 menunjukan hasil *clustering* beserta jumlah anggota pada masing-masing *cluster* nya.

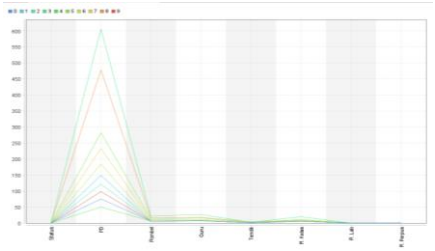
Tabel 3. Hasil *Clustering*

| ***Cluster*** | **Jumlah Anggota** |
|---|---|
| *Cluster* 0 | 137 *items* |
| *Cluster* 1 | 108 *items* |
| *Cluster* 2 | 143 *items* |
| *Cluster* 3 | 2 *items* |
| *Cluster* 4 | 98 *items* |
| *Cluster* 5 | 9 *items* |
| *Cluster* 6 | 70 *items* |
| *Cluster* 7 | 18 *items* |
| *Cluster* 8 | 1 *items* |
| *Cluster* 9 | 157 *items* |
| *Total number of items* | 743 |

### 4.5. *Interpretation/ Evaluation*

Penelitian ini didasarkan pada perbandingan nilai Davies Bouldin Index (DBI) dengan metode K-Means Clustering, menggunakan Measure Types Numerical Measures jenis Euclidean Distance, dan Manhattan Distance serta dilakukan iterasi untuk mendapatkan nilai performa terbaik dalam pengolahan data. Pengujian diawali dengan nilai K=2 hingga K=10. Ditemukan bahwa *cluster* dengan nilai yang paling mendekati 0 berada K=10 dengan nilai DBI 0.450 pada *Measure Types Numerical Measures* jenis *Manhattan Distance.*

Analisis klasterisasi telah dilakukan terhadap data sekolah untuk mengelompokkan sekolah-sekolah berdasarkan kesamaan karakteristik. Hasil klasterisasi menunjukkan pembentukan beberapa kelompok (*cluster*) sekolah dengan pola dan rentang nilai atribut yang bervariasi antar cluster. Pada Gambar 18, disajikan visualisasi hasil klasterisasi berupa *scatter plot* berdasarkan atribut yang ada pada dataset Sekolah Dasar (SD).



Gambar 18 Visualisasi

Berdasarkan visualisasi diatas, *cluster* 3 memiliki kecenderungan yang berbeda dibandingkan *cluster* lainnya. *Cluster* 3 mendominasi, sementara *cluster*

yang lainnya memiliki kemiripan atribut yang lebih tinggi. Terdapat kesamaan pola pada atribut Status, Rombel, Guru, Tendik, R. Kelas, R. Lab, dan R. Perpus.

a. *Cluster* 0 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 1 hingga 17 untuk atribut Rombel dan atribut Guru, 0 hingga 4 untuk atribut Tendik, 2 hingga 13 untuk atribut R. Kelas, 0 hingga 2 untuk atribut R. Lab dan R. Perpus. Namun pada atribut PD terjadi peningkatan signifikan dengan nilai antara 63 hingga 86.
b. *Cluster* 1 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 1 hingga 17 untuk atribut Rombel dan atribut Guru, 0 hingga 4 untuk atribut Tendik, 2 hingga 13 untuk atribut R. Kelas, 0 hingga 2 untuk atribut R. Lab dan R. Perpus. Namun pada atribut PD terjadi peningkatan signifikan dengan nilai antara 138 hingga 167.
c. *Cluster* 2 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 1 hingga 17 untuk atribut Rombel dan atribut Guru, 0 hingga 4 untuk atribut Tendik, 2 hingga 13 untuk atribut R. Kelas, 0 hingga 2 untuk atribut R. Lab dan R. Perpus. Namun pada atribut PD terjadi peningkatan signifikan dengan nilai antara 110 hingga 135.
d. *Cluster* 3 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 0 hingga 4 untuk atribut Tendik, 0 hingga 2 untuk atribut R. Lab dan R. Perpus. Tetapi memiliki pola berbeda pada beberapa atribut yang lain yaitu pada atribut PD memiliki rentang nilai 256 hingga 314, atribut Rombel 19 hingga 24, atribut Guru 25 hingga 29, atribut R. Kelas 18 hingga 22.
e. *Cluster* 4 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 9 hingga 63 untuk atribut PD, 1 hingga 17 untuk atribut Rombel dan atribut Guru, 0 hingga 4 untuk atribut Tendik, 2 hingga 13 untuk atribut R. Kelas, 0 hingga 2 untuk atribut R. Lab dan R. Perpus.
f. *Cluster* 5 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 1 hingga 17 untuk atribut Rombel dan atribut Guru, 0 hingga 4 untuk atribut Tendik, 2 hingga 13 untuk atribut R. Kelas, 0 hingga 2 untuk atribut R. Lab dan R. Perpus. Namun pada atribut PD terjadi peningkatan signifikan dengan nilai antara 256 hingga 314.
g. *Cluster* 6 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 1 hingga 17 untuk atribut Rombel dan atribut Guru, 0 hingga 4 untuk atribut Tendik, 2 hingga 13 untuk atribut R. Kelas, 0 hingga 2 untuk atribut R. Lab dan R. Perpus. Namun pada atribut PD terjadi peningkatan signifikan dengan nilai antara 168 hingga 209.
h. *Cluster* 7 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 1 hingga 17 untuk atribut Rombel dan atribut Guru, 0 hingga 4 untuk atribut Tendik, 2 hingga 13 untuk atribut R. Kelas, 0 hingga 2 untuk atribut R. Lab dan R. Perpus. Namun pada atribut PD terjadi peningkatan signifikan dengan nilai antara 205 hingga 254.
i. *Cluster* 8 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 1 hingga 17 untuk atribut Rombel dan atribut Guru, 0 hingga 4 untuk atribut Tendik, 2 hingga 13 untuk atribut R. Kelas, 0 hingga 2 untuk atribut R. Lab dan R. Perpus. Namun pada atribut PD terjadi peningkatan signifikan dengan nilai 478.
j. *Cluster* 9 memiliki rentang nilai 0 hingga 1 untuk atribut Status, 1 hingga 17 untuk atribut Rombel dan atribut Guru, 0 hingga 4 untuk atribut Tendik, 2 hingga 13 untuk atribut R. Kelas, 0 hingga 2 untuk atribut R. Lab dan R. Perpus. Namun pada atribut PD terjadi peningkatan signifikan dengan nilai antara 87 hingga 109.

Berdasarkan data hasil klasterisasi, dapat diamati bahwa setiap atribut (Status, PD, Rombel, Guru, Tendik, R. Kelas, R. Lab dan R. Perpus) memiliki pola dan karakteristik yang bervariasi. Dengan kata lain, setiap atribut memberikan kontribusi yang berbeda-beda dalam membentuk *cluster-cluster* hasil pengelompokan data. Perbedaan pola dan karakteristik inilah yang mengakibatkan terbentuknya *cluster* dengan ciri khas masing-masing. Kesimpulannya, setiap atribut memiliki karakteristik tertentu dalam proses pembentukan cluster pada data yang dianalisis.

## 5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa jumlah cluster yang paling optimal untuk pengelompokan data SD adalah 10 cluster, berdasarkan nilai *Davies Bouldin Index* (DBI) terendah yaitu 0,450. Jenis parameter *Measure Types* yang menghasilkan performa klasterisasi paling optimal adalah *Numerical Measure* dengan jenis *Manhattan Distance*. Iterasi optimal pada proses *K-Means Clustering* juga diperoleh pada iterasi ke-10, ditandai dengan nilai centroid serta keanggotaan *cluster* yang sudah stabil sehingga tidak ada perpindahan data antar *cluster*.

## DAFTAR PUSTAKA

[1] N. Nurahman, A. Purwanto, and S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, Mar. 2022, doi: 10.30812/matrik.v21i2.1411.

[2] H. S. Sutanto and I. Susilawaty, "ANALISIS KLASTERISASI DATA SEKOLAH SMP DI KABUPATEN KOTAWARINGIN TIMUR MENGGUNAKAN METODE K-MEANS," *JURSISTEKNI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 298–309, 2023, [Online]. Available: https://jursistekni.nusaputra.ac.id/article/view/269

[3] M. R. Sholihin and Rudiman, "Pemetaan Sekolah Muhammadiyah di Kabupaten PPU Berdasarkan

Fasilitas , Pendidik dan Tenaga Pendidik Menggunakan Metode K-Means Clustering," *J. Keilmuan dan Apl. Tek. Inform.*, vol. 5, no. 36, pp. 45–51, 2022, [Online]. Available: https://mail.jurnal.yudharta.ac.id/v2/index.php/EXPLORE-IT/article/view/3425

[4] D. W. Sari and K. G. Ayu, "Penentuan Kriteria Dalam Memilih Sekolah Dasar Dengan Menerapkan K-Means Clustering (Studi Kasus : Wilayah Kecamatan Mampang)," *Sainstech J. Penelit. dan Pengkaj. Sains dan Teknol.*, vol. 29, no. 2, pp. 24–28, Aug. 2019, doi: 10.37277/stch.v29i2.334.

[5] A. Wahyu and Rushenda, "Klasterisasi Dampak Bencana Gempa Bumi," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 8, no. 1, pp. 175–179, 2022, doi: 10.26418/jp.v8i1.

[6] Abdussalam Amrullah, I. Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, "ANALISIS CLUSTER FAKTOR PENUNJANG PENDIDIKAN MENGGUNAKAN ALGORITMA K-MEANS (STUDI KASUS: KABUPATEN KARAWANG)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, Nov. 2022, doi: 10.36595/jire.v5i2.701.

[7] Nurahman and D. D. Aulia, "Klasterisasi Pendidikan Masyarakat untuk mengetahui Daerah dengan Pendidikan Terendah menggunakan Algoritma K-Means," vol. 5, no. 1, pp. 38–44, 2023.

[8] S. Oktarian, S. Defit, and Sumijan, "Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering," *J. Inf. dan Teknol.*, vol. 2, no. 3, pp. 68–75, Sep. 2020, doi: 10.37034/jidt.v2i3.65.

[9] D. Kusbianto, I. A. Mashudi, and F. D. Prianggoro, "IMPLEMENTASI METODE RAPID CENTROID ESTIMATION DAN K-MEANS UNTUK KLASTERISASI SEKOLAH BERDASARKAN FASILITAS (STUDI KASUS SEKOLAH NEGERI KABUPATEN MALANG)," *Semin. Inform. Apl. Polinema*, pp. 254–259, 2020, [Online]. Available: http://jurnalti.polinema.ac.id/index.php/SIAP/article/view/780

[10] T. R. Tulili, K. B. Utomo, and D. Lestari, "PENGELOMPOKAN SEKOLAH DASAR DI KOTA SAMARINDA BERDASARKAN INDIKATOR STANDAR PELAYANAN MINIMAL PENDIDIKAN MENGGUNAKAN METODE K-MEANS," *Just TI (Jurnal Sains Terap. Teknol. Informasi)*, vol. 11, no. 2, pp. 24–28, Jul. 2019, doi: 10.46964/justti.v11i2.147.

[11] R. I. Fajriah, H. Sutisna, and B. K. Simpony, "Perbandingan Distance Space Manhattan Dengan Euclidean Pada K-Means Clustering Dalam Menentukan Promosi," *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 4, no. 1, pp. 36–49, 2019.