

Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik

School Clusterization Using K-Means Algorithm Based on Facilities, Educators, and Educations

Nurahman¹, Agung Purwanto², Sigit Mulyanto³
Universitas Darwan Ali, Indonesia

Informasi Artikel

Genesis Artikel:

Diterima, 5 November 2021

Direvisi, 22 Desember 2021

Disetujui, 5 Januari 2022

Kata Kunci:

Data Mining

DBI

K-Means

Kelasterisasi

Keywords:

Clusterisasi

Data Mining

DBI

K-Means

ABSTRAK

Fasilitas pendidikan merupakan salah satu indikator untuk tercapainya capaian pembelajaran di sekolah. Keberadaan fasilitas, guru, dan tenaga pendidik sangat dibutuhkan pada lingkungan sekolah. Untuk memudahkan pemerintah setempat dalam penanganan pemerataan kebutuhan sekolah, maka diperlukan pengklasteran atau pengelompokan sekolah. Tujuan dari penelitian ini adalah membantu pihak pemerintah dalam mengklaster sekolah tertentu menjadi beberapa *cluster*, sehingga memudahkan untuk melakukan pendampingan maupun pengadaan kebutuhan sekolah dilingkungan pemerintah Kabupaten Seruyan. Klasterisasi dilakukan dengan menggunakan algoritma *K-Means*. Penerapan Algoritma *K-Means* dengan menentukan nilai *Cluster* yaitu 3. Hasil penelitian menunjukkan bahwa setiap *Cluster* memiliki jumlah keanggotaannya masing-masing. Pada *Cluster 0* terdiri dari 178 sekolah, *Cluster 1* terdiri dari 3 sekolah, dan *Cluster 2* terdiri dari 43 sekolah. Kemudian hasil pemeringkatan *Cluster* secara berurutan diperoleh dari *Cluster 1*, *Cluster 0*, dan terakhir adalah *Cluster 2*. Selanjutnya dalam pengujian *performance* algoritma *K-Means* dengan membagi menjadi 3 *Cluster* diperoleh nilai *Davies Bouldin Index* senilai -0,695. Dari hasil pengolahan dan analisis data terdapat 43 sekolah dalam *cluster* rendah, sehingga penelitian ini merekomendasikan perlu dilakukan pendampingan dan pengadaan kebutuhan sekolah terhadap sekolah *cluster* rendah oleh dinas pendidikan pada pemerintah kabupaten Seruyan.

ABSTRACT

Educational facilities are one of the indicators for achieving learning outcomes in schools. The existence of facilities, teachers, and educators is needed in the school environment. To facilitate the local government in handling the distribution of school needs, it is necessary to cluster or group schools. The purpose of this study is to assist the government in clustering certain schools into several clusters, making it easier to provide assistance and procurement of school needs within the Seruyan Regency government. Clustering is done by using the *K-Means* algorithm. The application of the *K-Means* Algorithm by determining the *Cluster* value is 3. The results show that each *Cluster* has its own number of memberships. *Cluster0* consists of 178 schools, *Cluster1* consists of 3 schools, and *Cluster2* consists of 43 schools. Then the results of the *Cluster* ranking are sequentially obtained from *Cluster1*, *Cluster0*, and the last is *Cluster2*. Furthermore, in testing the performance of the *K-Means* algorithm by dividing it into 3 clusters, the *Davies Bouldin Index* value is -0.695. From the results of data processing and analysis, there are 43 schools in the low cluster, so this study recommends the need for assistance and provision of school needs for low cluster schools by the education office at the Seruyan district government.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Korespondensi:

Nurahman,

Program Studi Sistem Informasi,

Universitas Darwan Ali,

Email: nurrahman.ikhtiar@gmail.com

1. PENDAHULUAN

Indonesia memiliki jenjang pendidikan yaitu kelompok bermain, taman kanak-kanak, sekolah dasar, sekolah menengah pertama, dan sekolah menengah atas, serta perguruan tinggi. Sebagai warga negara memiliki hak mendapatkan pendidikan yang layak. Kebutuhan fasilitas pada sekolah sangatlah penting dan perlu menjadi perhatian untuk dapat menerapkan pendidikan yang layak. Semua jenjang pendidikan di Indonesia sudah tersebar di berbagai wilayah kabupaten demi mensukseskan pemerataan pendidikan di berbagai wilayah. Pemerataan pendidikan diberbagai wilayah kabupaten dengan dukungan pemerintah kabupaten dilakukan dengan memperhatikan perkembangan jumlah masyarakat disekitar. Diketahui bahwa pada bulan September tahun 2020 hasil sensus penduduk tercatat Indonesia mempunyai jumlah penduduk sebesar 270,20 juta jiwa [1]. Pada Kalimantan Tengah mencapai 2.669.969 jiwa [1]. Tahun 2019 pada Kabupaten Seruyan tercatat jumlah penduduk sebesar 147.545 jiwa.

Kabupaten Seruyan merupakan kabupaten yang berada di Provinsi Kalimantan Tengah. Kabupaten Seruyan pernah berada dalam posisi daerah tertinggal. Untuk itu, perlu menjadi perhatian pihak pemerintah agar pendidikan di wilayah tersebut dapat terus berkembang sesuai dengan standar pendidikan Nasional. Fasilitas pendidikan merupakan salah satu indikator untuk tercapainya capaian pembelajaran di sekolah. Pada prinsipnya fasilitas perlu dimanfaatkan secara efisiensi terhadap pemakaian fasilitas pembelajaran agar tujuan pembelajaran tercapai sehingga meningkatkan mutu sekolah. Selain itu, peran pendidik juga sangat penting dalam dunia pendidikan. Peranan pendidik dalam dunia pendidikan tidak hanya semata-mata memberikan informasi saja. Melainkan lebih dari itu, yaitu mengarahkan dan memberi fasilitas belajar (*directing and facilitating the learning*) agar proses belajar lebih memadai [2]. Dengan demikian, Keberadaan fasilitas dan guru sangat dibutuhkan pada lingkungan sekolah. Bahkan keberadaan tenaga pendidik juga dibutuhkan dalam proses berjalannya pendidikan dilingkungan sekolah.

Dinas pendidikan kabupaten memiliki salah satu tanggungjawab berkaitan dengan Pelayanan Pendidikan Dasar. Hal tersebut sebagai mana tercantum pada UU Nomor 23 Tahun 2014 pasal 12. Merujuk pada undang undang tersebut pada penelitian ini berfokus pada data pendidikan dasar yaitu data sekolah dasar (SD) dan data sekolah menengah pertama (SMP). Data sekolah yang menjadi tanggungjawab dinas pendidikan kabupaten seruyan sebanyak 232 sekolah menurut perekaman data Dapodik baik SD dan SMP. Data SMP terekam sebanyak 65 sekolah dan data SD terekam sebanyak 167. Diketahui, melalui perekaman data dapodik bahwa pengadaan fasilitas sekolah belum terlaksana secara merata diberbagai sekolah yang menjadi tanggungjawab dinas pendidikan kabupaten seruyan. Masih terdapat sekolah-sekolah yang belum memiliki fasilitas lengkap. Bahkan terdapat sekolah-sekolah yang tidak memiliki ruang perpustakaan yang dapat dijadikan sebagai prasarana pendukung dalam pembelajaran. Selain itu, menurut perekaman data dapodik bahkan pengadaan guru dan tendik masih tidak seimbang. Dari beberapa sekolah terekam hanya memiliki guru 3, 2 bahkan 1 guru pada sekolah tersebut. Hal ini tentunya perlu menjadi perhatian pihak pemerintah setepat agar dapat merealisasikan pengadaan fasilitas, guru, dan tendik secara merata sesuai dengan kebutuhan wilayahnya masing-masing.

Untuk memudahkan pemerintah setempat dalam penanganan pemerataan kebutuhan sekolah, maka diperlukan pengklasteran atau pengelompokan sekolah. Pengelompokan sekolah berdasarkan minat siswa merupakan hal penting [3], salah satu dari penunjang minat siswa adalah fasilitas sebagaimana disebutkan Dinatus Solichah pada penelitiannya [4] terdapat hubungan signifikan antara fasilitas sekolah dan minat belajar dengan kebiasaan belajar. Fasilitas sekolah serta lingkungan sosial berpengaruh signifikan terhadap hasil belajar siswa [5]. selain itu, suasana belajar juga dipengaruhi dengan adanya fasilitas sekolah yang lengkap dan memadai sehingga dapat mendorong siswa untuk aktif mengembangkan potensinya [5]. Potensi setiap siswa beragam, sehingga perlu adanya berbagai upaya yang harus dilakukan dalam dunia pendidikan, baik dari penerapan konsep mengajar, mengevaluasi, bahkan melakukan penelitian.

Tinggi dan rendahnya keberhasilan siswa merupakan cermin dari kualitas pendidikan tersebut. Dunia pendidikan dituntut untuk memiliki kompetensi dalam bersaing dengan memanfaatkan fasilitas yang ada. Sehingga, diperlukan pengetahuan dan pendampingan kepada pihak-pihak sekolah yang dianggap masih berada pada posisi lebih rendah daripada sekolah yang lainnya. Untuk memudahkan dalam pendampingan diperlukan pengklasterisasian sekolah. Data mining merupakan bidang ilmu yang dapat membantu untuk menganalisis jumlah data yang cukup besar. Salah satunya metode data *mining* yaitu *Clustering*. Metode *Clustering* juga pernah dilakukan dalam penelitian pemetaan siswa berprestasi [6]. Hasil penelitian menunjukkan bahwa penerapan Algoritma *K-Means* dapat menetapkan siswa berprestasi Tinggi, menengah, dan cukup. Kemudian pada pemetaan penyebaran guru di provinsi banten diketahui bahwa hasil *Clustering* menunjukkan Kota Cilegon, Kota Serang, Kota Tangerang Selatan, Kabupaten Pandeglang dan Kabupaten Lebak memiliki kekurangan guru SD [7]. Pada penelitian pengelompokan perguruan tinggi berdasarkan provinsi dengan 2 *cluster* yaitu *cluster* fasilitas tingkat tinggi dan *cluster* fasilitas tingkat rendah, diketahui hasil penelitian bahwa Jawa Barat, Jawa Tengah dan Jawa Timur masuk pada *cluster* tinggi sedangkan 31 provinsi lainnya dengan *cluster* rendah [8].

Penelitian pada artikel ini memiliki keterkaitan dengan beberapa penelitian terdahulu yang dijadikan sebagai rujukan. Penelitian pada artikel ini juga memiliki beberapa kesenjangan dengan penelitian terdahulu sebagaimana disajikan pada Tabel 1 yang menunjukkan posisi penelitian pada artikel ini.

Tabel 1. Matriks Literatur Review dan Posisi Penelitian

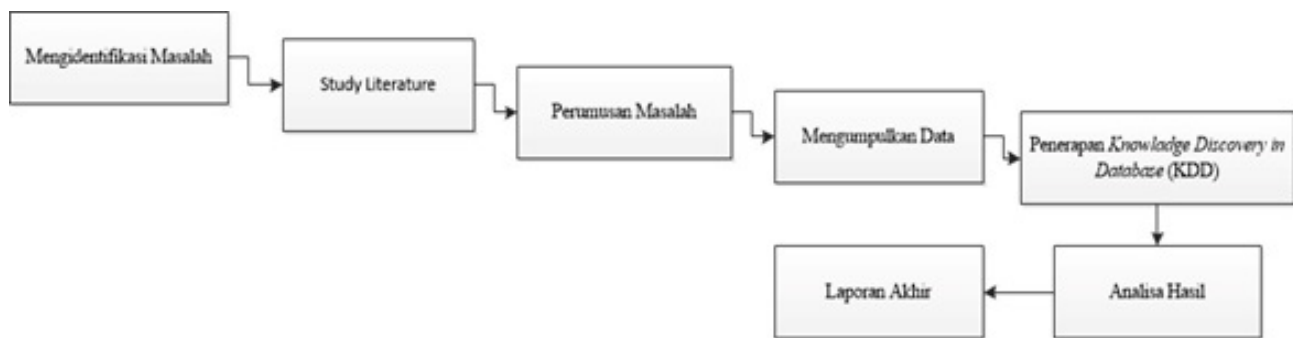
Tahun	Peneliti	Topik	Penelitian Terdahulu	Kesenjangan
2017	M. L. Sibuea dan A. Safta [6]	Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering	Penelitian ini melakukan klasifikasi siswa berprestasi berdasarkan kode, ekstrakurikuler, tugas, uts, uas, absensi, dan nilai sikap.	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2018	Y. A. Priambodo and S. Y. J. Prase-tyo [7]	Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Metode Spatial Clustering K-Means (Studi kasus: Wilayah Provinsi Banten).	Pada penelitian ini melakukan klasterisasi berdasarkan data jumlah sekolah, jumlah murid, dan jumlah guru.	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2018	F. Yunita [9]	Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru.	Pada penelitian ini melakukan klasterisasi berdasarkan Asal sekolah, Prodi/Jurusan yang dipilih, dan Nilai UAN	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2018	K. Handoko [10]	Pengelompokan Data Mining Pada Jumlah Penumpang di Bandara Hang Nadim.	Klasterisasi berdasarkan data penumpang berangkat, datang, dan transit,	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2018	K. Fatmawati and A. P. Windarto [11]	Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi	Pada penelitian ini melakukan klasterisasi berdasarkan jumlah rata-rata kabu-paten/kota dan jumlah rata-rata kabu-paten/kota yang terjangkit DBD	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2018	G. S. Nugraha, H. Hairani, and R. F. P. Ardi [12]	Aplikasi Pemetaan Kualitas Pendidikan di Indonesia Menggunakan Metode K-Means	Pada Penelitian ini merancang aplikasi untuk klasterisasi kualitas pendidikan berdasarkan Angka Partisipasi Kasar, Angka Partisipasi Murni, Angka Putus Sekolah, Angka Kelulusan dll.	Penelitian kami analisis klasterisasi berdasarkan data fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2019	M. A. Amri, A. P. Windarto, A. Wanto, and I. S. Damanik [8]	Analisis Metode K-Means Pada Pengelompokan Perguruan Tinggi Menurut Provinsi Berdasarkan Fasilitas yang dimiliki Desa	Klasterisasi pada penelitian ini dilakukan berdasarkan jumlah data fasilitas perguruan tinggi yang dimiliki desa pada tahun 2003-2018.	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2019	S. Suryadi [13]	Penerapan Metode Clustering K-Means untuk Pengelompokan Kelulusan Mahasiswa Berbasis Kompetensi,	Klasterisasi pada kelulusan mahasiswa dilakukan berdasarkan mhs ke-, ipk, kpt.	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2020	Suhefi Oktarian, Sarjon Defit, dan Sumijan [3]	Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering	Klasterisasi pengelompokan sekolah berdasarkan jumlah Rombel, Jumlah Peserta Didik, dan gabungan Jumlah guru + jumlah tenaga pendidik	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2020	I. F. Anshori and Y. Nuraini [14]	Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Tasikmalaya Menggunakan Algoritma K-Means	Klasterisasi pengelompokan data berdasarkan TKP, Waktu, Usia, Jenis Kelamin, Keadaan Korban, Profesi, dan Jenis Kendaraan.	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.
2019	A. Ali [15]	Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo	Klasterisasi data dilakukan berdasarkan jenis kelamin, usia, kecamatan dan diagnosa penyakitnya	Penelitian kami klasterisasi berdasarkan fasilitas (Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Jumlah Ruang Kelas), guru dan tenaga pendidik. Pada penelitian kami melakukan uji performa algoritma untuk mengetahui cluster yang optimal.

Berdasarkan beberapa penelitian terdahulu, maka diketahui terdapat kesenjangan yang diperoleh pada artikel penelitian kami adalah fokus penelitian yang berbeda dan atribut yang digunakan. Pada penelitian kami juga melakukan pencarian nilai performa dalam mengolah data. Nilai performa algoritma kami sajikan dengan mencari nilai *Davies Bouldin Index*. Dengan disajikan nilai *Davies Bouldin Index* maka dapat kita ketahui hasil dari tingkat keakuratan suatu algoritma dalam mengolah dan menganalisis data. Algoritma yang akan digunakan untuk menganalisis data adalah algoritma *K-Means*. Algoritma *K-Means* mampu mengelompokkan data pada kelompok yang sama dan data yang berbeda pada kelompok yang berbeda [9]. Algoritma ini menerapkan perhitungan dengan nilai *K* dan titik *centroid*. Tujuan dari penelitian ini adalah membantu pihak pemerintah dalam mengklaster sekolah tertentu menjadi beberapa *cluster*, sehingga memudahkan untuk melakukan pendampingan maupun pengadaan kebutuhan sekolah dilingkungan pemerintah Kabupaten Seruyan.

Artikel ini selanjutnya disusun dengan pembahasan metode penelitian, hasil dan pembahasan, dan kesimpulan. Pada bagian 2 metode penelitian membahas mengenai tahapan yang dilakukan dalam penelitian ini diantaranya seperti : identifikasi masalah, Studi literatur, perumusan masalah, pengumpulan data, penerapan metodologi *Knowledge Discovery in Database* (KDD), analisis hasil dan laporan akhir penelitian . Bagian 3 adalah hasil dan pembahasan, pada bagian ini menyajikan hasil penelitian. Bagian 4 kesimpulan hasil penelitian.

2. METODE PENELITIAN

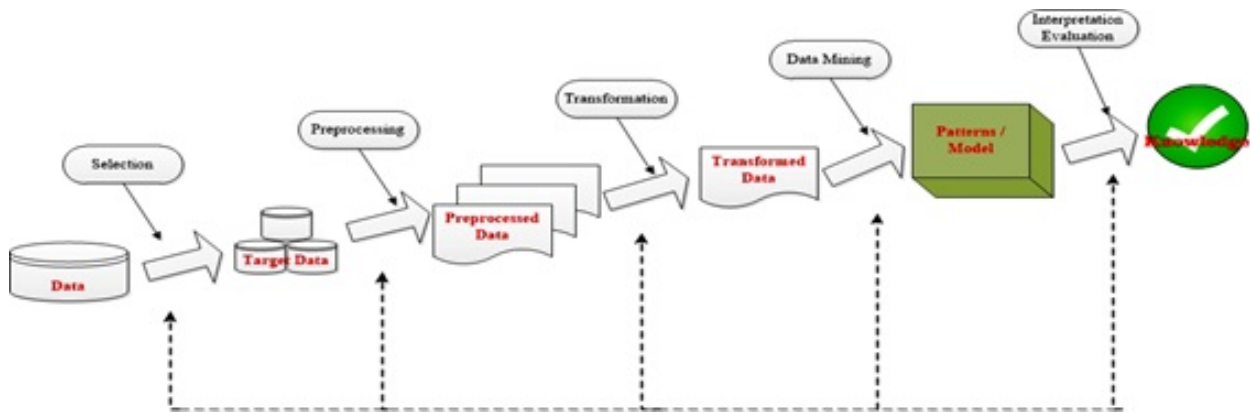
Metodologi Penelitian merupakan kerangka kerja berisi tentang tahapan-tahapan yang digunakan dalam penelitian. Tahapan penelitian yang dilakukan dalam penelitian ini adalah mengidentifikasi masalah, menganalisa masalah, studi literatur, mengumpulkan data, dan kemudian penerapan metodologi *Knowledge Discovery in Database* (KDD). Adapun tahapan-tahapan penelitian yang dilakukan dapat digambarkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berikut adalah deskripsi dari tahapan penelitian:

1. Identifikasi masalah bertujuan agar peneliti mendapatkan beberapa masalah yang kedepannya akan dicari bagaimana menyelesaikan masalah tersebut, dengan kata lain bisa disebutkan bahwa identifikasi masalah yang baik akan menggambarkan permasalahan dalam topik penelitian.
2. Pada penelitian ini studi literatur dilakukan dengan menelaah buku dan literatur yang berkaitan dengan masalah yang akan diselesaikan, studi literatur dilakukan juga untuk mengumpulkan data penunjang yang berkaitan dengan teori yang mendukung penelitian.
3. Setelah melakukan identifikasi masalah dan studi literatur, maka dilakukan selanjutnya adalah perumusan masalah. Hal ini dilakukan untuk mendapatkan perumusan masalah yang akan menjadi fokus dalam penelitian ini.
4. Proses pengumpulan data dilakukan dengan observasi di dinas pendidikan kabupaten seruyan. Kemudian observasi selanjutnya diarahkan oleh pihak dinas pendidikan kabupaten seruyan kepada pencarian data melalui aplikasi dapodik.
5. Penerapan metodologi *Knowledge Discovery in Database* (KDD) dilakukan setelah data terkumpul. Pada tahapan ini terdapat 5 langkah penting untuk dapat memperoleh suatu pengetahuan baru dari hasil pengolahan data. Lima langkah penting yang dilakukan dalam pengolahan data pada penelitian ini yaitu, *Selection, Preprocessing, Transformations, Data Mining, dan Interpretation Evaluation*. metodologi *Knowledge Discovery in Database* (KDD) dapat dilihat pada Gambar 2.



Gambar 2. Metodologi Knowledge Discovery in Database (KDD)

6. Setelah melakukan pengolahan data atau Penerapan metodologi *Knowledge Discovery in Database* (KDD) maka selanjutnya melakukan Analisa terhadap hasil pengolahan data. Analisis dilakukan dengan memahami hasil pengolahan data dengan meninjau beberapa penelitian terdahulu.
7. Penyusunan Laporan Akhir dilakukan dengan membuat sebuah artikel ilmiah berdasarkan pengamatan, pengolahan data dan analisis hasil.

2.1. Metodologi *Knowledge Discovery in Database* (KDD)

Penerapan *Knowledge Discovery In Database* (KDD) dalam penelitian ini dilakukan untuk mengetahui pengetahuan dalam pengambilan keputusan. Sebagaimana disebutkan oleh Yuli Mardi bahwa KDD adalah metode untuk mengetahui pengetahuan yang diperoleh dari sebuah basis data untuk keperluan pendukung keputusan [10]. Langkah-langkah Metodologi *Knowledge Discovery in Database* (KDD) pada penelitian ini sebagaimana ditunjukkan pada Gambar 2 dan dapat dijelaskan sebagai berikut:

1. Selection

Pada proses sebelum pengolahan data perlu dilakukan seleksi (*Selection*) data. Data diperoleh melalui aplikasi dapodik. Sebagaimana kita ketahui bahwa pada aplikasi Data Pokok Pendidikan (Dapodik) ini merupakan sekumpulan data besar yang menyimpan semua laporan data pendidikan secara nasional. Data diseleksi berdasarkan laporan-laporan sekolah yang ada pada kabupaten seruyan. Sekumpulan data yang diperoleh pada pengumpulan data kemudian diseleksi sesuai kebutuhan *Knowledge Discovery In Database* (KDD) yang diperlukan pada penelitian ini.

2. Preprocessing

Preprocessing dilakukan untuk memfokuskan terhadap data yang akan diproses. Penelitian ini melakukan langkah data reduction pada tahapan *preprocessing* data. Proses data *reduction* dilakukan untuk membuang data yang inkonsisten dan memperbaiki kesalahan data. *Preprocessing* dilakukan secara hati-hati hingga data yang dibutuhkan dapat sesuai dengan ketentuan dari *Knowledge Discovery In Database* (KDD).

3. Transformations

Transformations merupakan proses integrasi dari data yang telah dipilih pada penelitian sehingga data dapat sesuai dengan proses data *mining* [11]. Hal ini dilakukan untuk *mentransform* data-data yang dapat diolah dalam penerapan Algoritma yang ditetapkan pada penelitian. Sebagaimana diketahui Algoritma *K-Means* merupakan algoritma yang dapat memproses data dengan *type* data numerik atau integer. Sehingga Transformasi data dilakukan tujuannya agar data dapat diolah dengan menggunakan metode *K-Means Clustering* [12].

4. Data Mining

Data *Mining* atau biasa dikenal tahapan *mining proses* merupakan proses penambangann data untuk mendapatkan pengetahuan atau hubungan data-data yang dapat disimpulkan menjadi informasi yang baru [13]. *Mining* sendiri dapat diartikan untuk mendapatkan suatu berharga yang berjumlah sedikit dari material dasar yang berjumlah besar [14]. Pada tahapan ini diperlukan kepastian penggunaan metode yaitu dengan menentukan karakterisasi, klasifikasi, regresi, *clustering*, asosiasi, dan lain-lain [11]. Pemilihan metode dan algoritma pada tahapan ini sangat tergantung dengan penerapan KDD. Sehingga *mining proses* dapat disebut sebagai proses utama sedangkan tahapan yang lain disebut sebagai pendukung [15].

5. Interpretation/ Evaluation

Tahapan *Interpretation/ Evaluation* merupakan tahapan yang dilakukan untuk mengidentifikasi pola yang dihasilkan dari *mining proses*. Pola informasi yang dihasilkan dari *mining proses* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan [16]. *Evaluation* yang dilakukan akan meninjau pola-pola pada proses klasterisasi dengan menunjukkan nilai performa algoritma *K-Means*.

2.2. Algoritma K-Means

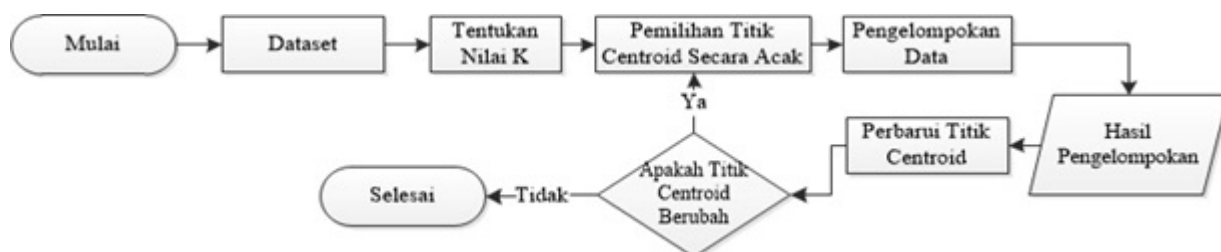
Dataset yang akan dianalisis pada penelitian ini tidak memiliki label sehingga metode yang cocok dalam penelitian ini adalah klusterisasi dengan menggunakan Algoritma *K-Means*. Algoritma *K-Means* adalah algoritma yang digunakan untuk analisis pengelompokan yang mengarah pada pemartisian n objek pengamatan kedalam k kelompok (*Cluster*) [17]. Pada Algoritma *K-Means* setiap objek pengamatan dimiliki oleh sebuah kelompok yang memiliki mean terdekat, dimana mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan oleh algoritma. Pada penelitian ini *process mining* dilakukan dengan menggunakan Algoritma *K-Means* yaitu dengan penerapan persamaan 1.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Proses klusterisasi dilakukan dengan mengidentifikasi data yang akan dikluster terlebih dahulu. Setelah melakukan indentifikasi data maka selanjutnya dapat menerapkan persamaan (1) untuk menghitung nilai jarak antara data x ke data y . Dimana diketahui bahwa x_i merupakan data testing ke- i sedangkan y_i adalah data *training* ke- i . Kemudian dilanjutkan dengan mencari nilai titik *centroid* dari *cluster* ke K . Untuk mencari nilai titik *centroid* dari *cluster* ke K dapat dilakukan dengan menggunakan persamaan (2).

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \quad (2)$$

Berdasarkan persamaan 2 diatas dapat dijelaskan bahwa μ_k merupakan nilai titik *centroid* dari *cluster* ke K . Kemudian simbol N_k digunakan untuk menunjukkan nilai banyaknya data pada *Cluster* ke K . Selanjutnya symbol x_i digunakan untuk menunjukkan nilai data ke i pada *Cluster* ke K . Gambar 3 adalah tahapan dari pengolahan data menggunakan Algoritma *K-Means*. *Dataset* yang telah final akan dilakukan pengolahan data dengan menggunakan Algoritma *K-Means*. Jumlah nilai K akan ditentukan untuk membentuk jumlah *cluster*, kemudian pencarian nilai titik *centroid*, pengelompokan data berdasarkan jumlah *cluster* yang telah ditentukan. Jika titik *centroid* berubah maka akan dilakukan lagi pemilihan titik *centroid* hingga diketahui tidak ada perubahan pada titik *centroid*.



Gambar 3. Tahapan Algoritma *K-Means*

Keterangan tahapan Algoritma *K-Means*:

1. Mulai dalam perhitungan untuk penerapan Algoritma *K-Means*
2. Menyiapkan *Dataset*: *Dataset* yang digunakan adalah data sekolah SD dan SMP dikabupaten seruyan.
3. Tentuan jumlah klaster yaitu nilai K : penentuan nilai K ini digunakan untuk menentukan kelompok atau *Cluster* sekolah yang akan dibentuk.
4. Pemilihan titik *centroid* secara acak dilakukan untuk memusatkan perhitungan dalam sebuah *cluster* pada setiap data yang diperhitungkan.
5. Kelompokan data sehingga terbentuk K buah klaster dengan titik *centroid* dari setiap *cluster*
6. Hasil Pengelompokan data berdasarkan iterasi yang dilakukan
7. Perbaruan nilai titik *centroid* untuk mengetahui adanya perubahan titik *centroid* atau tidak. Jika tidak terdapat perubahan maka perhitungan dihentikan.

2.3. Davies Boildin Index

Pengujian performa algoritma *K-Means* dilakukan dengan mencari nilai *Davies Boildin Index*. *Davies Boildin Index* berada dalam proses *mining* data. Pencarian nilai *Davies Boildin Index* setelah melakukan klusterisasi menggunakan Algoritma *K-Means* berdasarkan nilai K yang telah ditentukan. Klusterisasi yang telah ditemukan kemudian di uji menggunakan prinsip pengukuran pada *Davies Boildin Index*. Pengukuran *Davies Boildin Index* menggunakan prinsip memaksimalkan jarak *inter Cluster* serta meminimalkan jarak *intra cluster* [18]. Sehingga dapat dinyatakan bahwa semakin kecil nilai *Davies Boildin Index* maka menunjukkan

skema *cluster* yang paling optimal. Formula dari perhitungan *Davies Boildin Index* dapat disajikan pada persamaan (3).

$$DBI = \frac{1}{k} \sum_{i=1}^k \{R_i\}$$

dengan

$$R_i = \max_{j=1, \dots, k, i \neq j} R_{ij}, R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (3)$$

dan

$$S_i = \left[\frac{1}{n_i} \sum_{x \in n_i} d^2(x, v_i) \right]^{\frac{1}{2}}$$

Pada proses pencarian nilai *Davies Boildin Index* yang terdapat pada persamaan (3) bahwa simbol k menunjukkan nilai jumlah cluster dan R_{ij} adalah ukuran kemiripan antara nilai n_i dan n_j . Simbol S_i menunjukkan ukuran dispersi cluster ke- i , yang mana $i = 1, 2, \dots, k$. Simbol d_{ij} menunjukkan jarak antara centroid cluster ke- i dan centroid cluster ke- j ($d_{ij} = d_{ji}$). Simbol n_i menunjukkan bahwa banyaknya anggota cluster ke- i , $i = 1, 2, \dots, k$. Kemudian V_i menunjukkan nilai centroid cluster dari n_i .

3. HASIL DAN ANALISIS

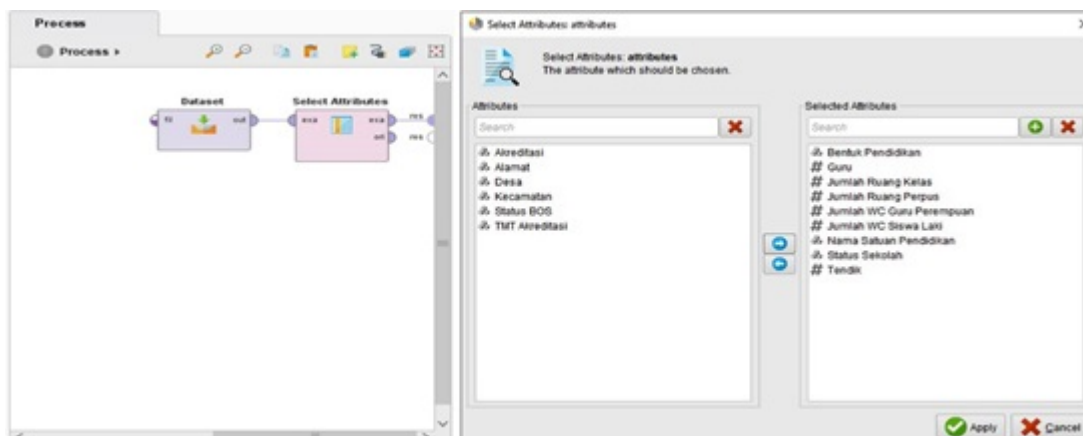
Penelitian dilakukan dengan tahapan yang telah di rencanakan dalam metodologi penelitian yaitu di awali dengan melakukan identifikasi masalah. Identifikasi masalah dilakukan dengan pengamatan terhadap dunia pendidikan di lingkungan kabupaten seruyan. Dalam pengamatan diketahui bahwa ditemukannya permasalahan terkait dengan tidak adanya pemerataan mengenai fasilitas sekolah dan kebutuhan lainnya yang di salurkan oleh pihak pemerintah. Kemudian untuk memperkaya informasi dalam pengamatan dilakukan juga *study literature* untuk mendukung kasus yang ditemukan secara ilmiah dan teoritis. Setelah diketahui kasus secara keseluruhan maka selanjutnya melakukan perumusan masalah dengan tujuan agar penelitian lebih fokus pada kasus yang diteliti yaitu mengenai pemerataan fasilitas pendidikan dan kebutuhan lainnya. Untuk dapat mengelompokan satuan pendidikan berdasarkan klasternya maka hasil dari *Study Literature* ditetapkan melakukan analisis data *mining* dengan metode *clustering*.

3.1. Dataset

Dataset diperoleh dari dinas pendidikan kabupaten Seruyan. Data awal hasil dari *Selection* pada aplikasi dapodik memiliki 15 atribut yaitu sebagai berikut : Nama Satuan Pendidikan, Bentuk Pendidikan, Status Sekolah, Alamat, Desa, Kecamatan, Status BOS, TMT Akreditasi, Akreditasi, Jumlah Ruang Kelas, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Guru, dan Tendik. *Dataset* memiliki rekam data sebanyak 232 *record*.

3.2. Selections

Dataset yang diperoleh kemudian dilakukan *Selection* sehingga sesuai dengan kebutuhan penelitian dan penerapan KDD yang direncanakan pada metodologi penelitian. Atribut yang diperoleh dari pihak dinas pendidikan kabupaten seruyan telah dipertimbangkan untuk tidak digunakan semua dalam proses pengolahan data. Sehingga diperlukan proses *selection* atribut. Proses *Selection* dilakukan dengan aplikasi *rapid miner* dengan menggunakan operator *Select Attributes*. Gambar 4 menunjukkan tampilan *Selection* pada aplikasi *rapid miner*.



Gambar 4. Desain Select Atribut

Gambar 4 menunjukkan proses seleksi atribut yang akan digunakan dalam pengolahan data. Atribut yang dipilih dalam pengolahan data terdiri dari 9 atribut yaitu : Nama Satuan Pendidikan, Bentuk Pendidikan, Status Sekolah, Jumlah Ruang Kelas, Jumlah Ruang Perpustakaan, Jumlah WC Guru Perempuan, Jumlah WC Siswa Laki, Guru, dan Tendik. Atribut Nama Satuan Pendidikan dijadikan *type id* sehingga pada pemrosesan data tidak akan diperhitungkan, karena atribut ini adalah atribut objek penelitian untuk mengetahui hasil dari *clustering*.

3.3. Preprocessing

Preprocessing dilakukan untuk memperbaiki data rusak dan data tidak konsisten menjadi data yang dapat olah. Pada *Preprocessing* juga dilakukan data *reduction* untuk data yang sulit dimengerti. Proses *reduction* dapat dilakukan dengan membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Gambar 5 menunjukkan *Preprocessing Data*.

Row No.	Nama Sat...	Jumlah Ru...	Jumlah WC ...	Jumlah WC ...	Bentuk Pen...	Status Seko...	Jumlah Ru...	Guru	Tendik
211	SMP SAMP 3.	1	1	0	SMP	Negeri	3	5	3
212	SMP SARU A.	0	0	0	SMP	Negeri	2	3	3
213	SMP SARU A.	0	0	0	SMP	Negeri	3	5	3
214	SMP SARU A.	0	0	0	SMP	Negeri	1	3	3
215	SMP SARU A.	0	0	0	SMP	Negeri	2	5	1
216	SMP SIAST...	1	1	1	SMP	Swasta	6	6	2
217	SMP SIAST...	1	1	1	SMP	Swasta	6	6	2
218	SMP SIAST...	0	0	1	SMP	Swasta	6	6	1
219	SMP TUNAS...	1	1	1	SMP	Swasta	9	15	3
220	SMP ALJAH...	1	1	1	SMP	Swasta	9	14	4
221	SMPN SATU...	0	0	1	SMP	Negeri	3	9	3
222	SMPN -4 BE...	1	1	1	SMP	Swasta	10	13	2
223	SMPN BINA B.	1	6	5	SMP	Swasta	9	9	1
224	SMPN EVA T.	1	9	1	SMP	Swasta	6	12	2

Gambar 5. Preprocessing

Pada Gambar 5 *preprocessing* data dilakukan dengan operator *replace missing value* dan operator *filter examples*. Proses ini melakukan perbaikan data yang tidak konsisten dan penghapusan *record* yang tidak dapat ditemukan tujuan dari data atau data tidak dapat dimengerti. *Dataset* semula berjumlah 232 setelah tahapan *preprocessing* selesai maka data yang *valid* saat ini berjumlah 224 *record*.

3.4. Transformations

Tahapan *transformations* Data merupakan tahapan untuk merubah *value* dari *polynomial* menjadi integer dan tahapan ini merupakan akhir dari persiapan data. Hal yang sama juga dilakukan Tuti Khotimah [19] untuk mengubah *value* bertipe *time* menjadi tipe integer. Data-data berjenis nominal harus dilakukan proses inialisasi data kebentuk angka atau numerikal [20]. Persiapan data dilakukan mulai *dataset* yang tidak *valid*, kemudian dilakukan *select attributes* hingga *transformations* data. *Transformations* data yang dilakukan dalam penelitian ini mengubah data dari *polynomial* menjadi data *numeric* yaitu pada atribut status sekolah dan atribut bentuk pendidikan. Nilai status sekolah swasta dirubah menjadi nilai 1 sedangkan nilai status sekolah negeri dirubah menjadi nilai 2. Kemudian nilai bentuk pendidikan SD dirubah menjadi nilai 1 sedangkan nilai bentuk pendidikan SMP dirubah menjadi 2. Gambar 6 merupakan tahapan *transformations* Data.

Row No.	Nama Sat...	Bentuk Pen...	Status Seko...	Jumlah Ru...	Jumlah WC ...	Jumlah WC ...	Jumlah Ru...	Guru	Tendik
211	SMP SAMP 3.	2	2	1	1	0	3	5	3
212	SMP SARU A.	2	2	0	0	0	2	3	3
213	SMP SARU A.	2	2	0	0	0	3	5	3
214	SMP SARU A.	2	2	0	0	0	1	3	3
215	SMP SARU A.	2	2	0	0	0	2	5	1
216	SMP SIAST...	2	1	1	1	1	6	6	2
217	SMP SIAST...	2	1	1	1	1	6	6	2
218	SMP SIAST...	2	1	0	0	1	6	6	1
219	SMP TUNAS...	2	1	1	1	1	9	15	3
220	SMP ALJAH...	2	1	1	1	1	9	14	4
221	SMPN SATU...	2	2	0	0	1	3	9	3
222	SMPN -4 BE...	2	1	1	1	1	10	13	2
223	SMPN BINA B.	2	1	1	6	5	9	9	1
224	SMPN EVA T.	2	1	1	0	1	6	12	2

Gambar 6. Transformations Data

Gambar 6 menunjukkan desain proses *transformations* data dengan hasil proses diketahui setiap atribut memiliki *value numeric*. Data yang telah *ditransformaions* semua siap di *mining* menggunakan metode *Clustering* dengan algoritma *K-Means*.

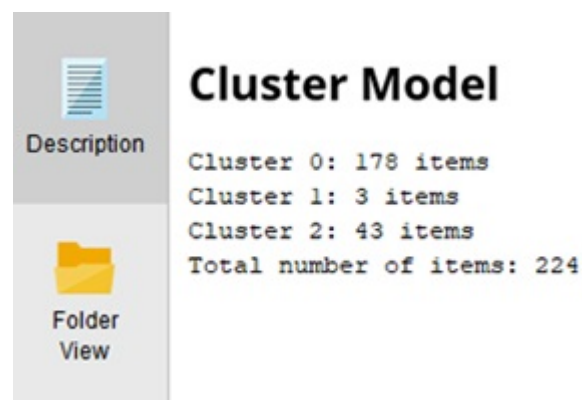
3.5. Data Mining

Dataset yang telah melalui tahapan *preprocessing* dan *transformations* maka siap untuk tahapan berikutnya yaitu *mining process*. Jumlah *dataset* yang *valid* saat ini sebanyak 224 *record* dengan 9 *attributes*. Tabel 2 menunjukkan *dataset* yang siap di *mining process* atau *dataset valid*.

Tabel 2. Dataset Valid

Nama Satuan Pendidikan	Bentuk Pendidikan	Status Sekolah	Jumlah Ruang Perpus	Jumlah WC Guru Perempuan	Jumlah WC Siswa Laki	Jumlah Ruang Kelas	Guru	Tendik
SD Asseruyaniyyah	1	1	1	1	1	10	14	3
SD Eka Tjipta Bukit Tiga	1	1	1	0	3	6	10	2
SD Eka Tjipta Hanau	1	1	2	0	1	6	10	2
SD Eka Tjipta Muara Dua	1	1	1	0	1	6	8	1
SD Eka Tjipta Nahiyang	1	1	1	1	1	6	8	2
SD Eka Tjipta Rungau	1	1	1	1	1	9	12	4
SD Eka Tjipta Semandau	1	1	1	0	1	6	10	2
SD Eka Tjipta Semilar	1	1	1	1	3	12	16	2
SD Eka Tjipta Seruyan	1	1	1	1	3	10	10	2
SD Eka Tjipta Sungai Ayawan	1	1	1	0	1	6	7	1
SD Eka Tjipta Tangar	1	1	1	1	1	6	9	1
...
...
SMPS Bina Bangsa 03	2	1	1	6	5	9	8	1
SMPS Eka Tjipta Perdana	2	1	1	0	1	6	12	2

Tabel 2 *dataset valid* menunjukkan data telah dirubah ke bilangan *numeric* semua, sehingga penerapan algoritma *K-Means* dapat dilakukan dengan menggunakan *tools rapid minner*. *Dataset* yang diolah dibagi menjadi 3 *Cluster* yaitu dengan menentukan nilai $k=3$. Pembagian dari 3 *Cluster* menghasilkan *cluster* model sebagaimana disajikan pada Gambar 7.



Gambar 7. Cluster Model Pengelompokan Sekolah di Kabupaten Seruyan

Gambar 7 menampilkan keanggotaan dari masing-masing *Cluster*. Pada *Cluster 0* keanggotaan terdiri dari 178 *items* artinya dalam *Cluster 0* memiliki 178 sekolah yang dikategorikan dalam satu kelompok. *Cluster 1* memiliki 3 *items* artinya dalam *Cluster 1* terdapat 3 sekolah. Selanjutnya *Cluster 2* memiliki 43 *items* artinya dalam *Cluster 2* terdapat 43 sekolah. Setelah penentuan nilai *cluster* berikutnya pemilihan titik *centroid*. Titik *centroid* dipilih secara acak pada setiap atribut dimasing-masing *Cluster*. Pemilihan titik *centroid* dapat disajikan pada Tabel 3.

Tabel 3. Titik Centroid

Attribute	Cluster_0	Cluster_1	Cluster_2
Bentuk Pendidikan	1.258	1.667	1.279
Status Sekolah	1.792	1.667	1.698
Jumlah Ruang Perpus	0.831	1	1.023
Jumlah WC Guru Perempuan	0.489	1	0.767
Jumlah WC Siswa Laki	0.938	2	1.395
Jumlah Ruang Kelas	6.034	16	11.268
Guru	7.652	32.667	15.140
Tendik	1.466	7.333	3.070

Dari penentuan nilai *cluster* dan pemilihan titik *centroid* menghasilkan keanggotaan pada setiap *cluster*. Diketahui bahwa pada *cluster-2* berjumlah 43 sekolah yang terdiri dari SD Asseruyaniyyah, SD Eka Tjipta Rungau, SD Eka Tjipta Semilar, SD Erna

Djuliawati, SD Negeri 1 Kuala Pembuang I, SD Negeri 1 Kuala Pembuang II, SD Negeri 1 Pembuang Hulu I, SD Negeri 1 Persil Raya, SD Negeri 1 Rantau Pulut, SD Negeri 1 Sembuluh II, SD Negeri 1 Telaga Pulang, SD Negeri 1 Terawan, SD Negeri 1 Tumbang Langkai, SD Negeri 1 Tumbang Manjul, SD Negeri 2 Kartika Bhakti, SD Negeri 2 Kuala Pembuang I, SD Negeri 2 Lanpasa, SD Negeri 2 Pematang Panjang, SD Negeri 2 Pembuang Hulu II, SD Negeri 2 Persil Raya, SD Negeri 2 Rantau Pulut, SD Negeri 2 Sembuluh I, SD Negeri 2 Tumbang Manjul, SD Negeri 3 Kuala Pembuang I, SD Negeri 3 Kuala Pembuang II, SD Swasta Bina Bangsa 03, SD Swasta Wijaya Kusuma 2, SD Wijaya Kusuma, SDIT An Najiyah, SDS Bina Bangsa 04, SDS Bina Bangsa 05, SMP Eka Tjipta Katayang, SMP Negeri 1 Danau Sembuluh, SMP Negeri 1 Hanau, SMP Negeri 1 Seruyan Hilir Timur, SMP Negeri 1 Seruyan Hulu, SMP Negeri 1 Seruyan Raya, SMP Negeri 1 Seruyan Tengah, SMP Negeri 2 Danau Sembuluh, SMP Negeri 2 Seruyan Tengah, SMP Negeri 3 Seruyan Tengah, SMP Tunas Agro, dan SMP Wijaya Kusuma. *Cluster-1* berjumlah 3 sekolah yang terdiri dari SD Qurrata A Yun, SMP Negeri 1 Kuala Pembuang, dan SMP Negeri 2 Kuala Pembuang. Sedangkan pada *Cluster-0* berjumlah 178 sekolah dengan data sekolah yang tidak tercantum pada *Cluster-2* dan *Cluster-1* dari jumlah sekolah sebanyak 224 sekolah dikabupaten seruyan.

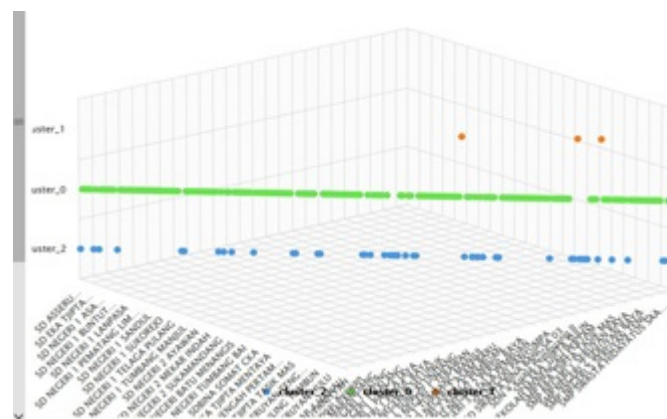
3.6. Analisis Cluster

Clustering menggunakan Algoritma *K-Means* dilakukan dengan mengklasterisasi keanggotaan berdasarkan kemiripannya. Keanggotaan dari *Cluster-1* memiliki karakteristik bentuk pendidikan SD atau SMP, status sekolah swasta atau negeri, Jumlah Ruang Perpustakaan 0 sampai 2, Jumlah WC Guru Perempuan 0 sampai 2, Jumlah WC Siswa Laki 0 sampai 5, Jumlah Ruang Kelas 6 sampai 22, Jumlah Guru 10 sampai 22, dan Jumlah Tendik 1 sampai 7. Keanggotaan *Cluster-1* memiliki karakteristik bentuk pendidikan SD atau SMP, status sekolah swasta atau negeri, Jumlah Ruang Perpustakaan 1, Jumlah WC Guru Perempuan 1, Jumlah WC Siswa Laki 0 sampai 5, Jumlah Ruang Kelas 12 sampai 19, Jumlah Guru 29 sampai 37, dan Jumlah Tendik 2 sampai 12. Selanjutnya keanggotaan *Cluster-0* memiliki karakteristik bentuk pendidikan SD atau SMP, status sekolah swasta atau negeri, Jumlah Ruang Perpustakaan 0 sampai 2, Jumlah WC Guru Perempuan 0 sampai 2, Jumlah WC Siswa Laki 0 sampai 9, Jumlah Ruang Kelas 1 sampai 13, Jumlah Guru 1 sampai 13, dan Jumlah Tendik 0 sampai 5. Analisis *Cluster* juga dapat disajikan sebagaimana pada Tabel 4.

Tabel 4. Karakteristik Keanggotaan Pada Cluster

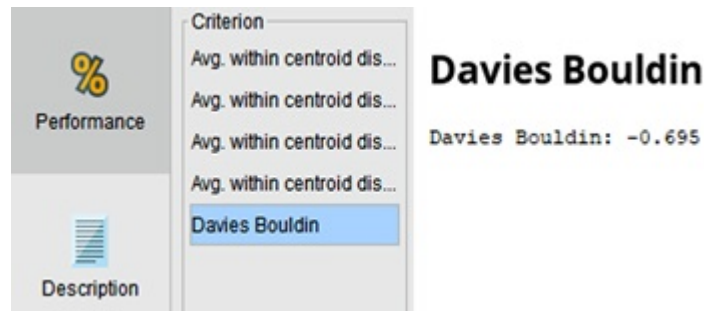
Cluster	Bentuk Pendidikan	Status Sekolah	Jumlah Ruang Perpustakaan	Jumlah WC Guru Perempuan	Jumlah WC Siswa Laki	Jumlah Ruang Kelas	Guru	Tendik
Cluster 0	SD/ SMP	Swasta/Negeri	0-2	0-2	0-9	Jan-13	Jan-13	0-5
Cluster 1	SD/ SMP	Swasta/Negeri	1	1	0-5	Dec-19	29-37	02-Dec
Cluster 2	SD/ SMP	Swasta/Negeri	0-2	0-2	0-5	Jun-22	Oct-22	01-Jul

Tabel 4 karakteristik keanggotaan pada *cluster* menunjukkan bahwa pemerataan kebutuhan sekolah sangat tidak seimbang terlihat pada atribut guru. Sehingga hal tersebut perlu dijadikan pertimbangan oleh pihak dinas pendidikan kabupaten seruyan agar dapat di realisasikan pengadaan guru secara seimbang. Kemudian jika dilihat dari pemerinkatan *cluster* maka dapat disajikan pada gambar 8.



Gambar 8. Peringkat Cluster

Gambar 8 menunjukkan hasil dari visualisasi pemerinkatan *clustering*. Diketahui bahwa *orange* berada pada peringkat pertama. Kemudian warna *green* peringkat kedua serta warna *blue* berada pada peringkat ketiga. Sehingga berdasarkan hasil visualisasi Gambar 8 dan Gambar 7 dapat dijelaskan bahwa warna *orange* berada pada posisi *Cluster-1* artinya bahwa *cluster 1* berada peringkat pertama sebanyak 3 sekolah. Warna *green* menunjukkan *cluster-0* artinya *cluster 0* berada pada peringkat kedua sebanyak 178 sekolah. Selanjutnya *blue* berada pada *cluster-2* artinya *cluster 2* berada pada peringkat ketiga sebanyak 43 sekolah.



Gambar 9. Davies Bouldin Index

Performance suatu algoritma dari hasil analisis dapat disajikan dalam berbagai cara. Penelitian ini menyajikan *performance* algoritma dengan menunjukkan nilai dari *Davies Bouldin Index*. Pada Gambar 9 *Davies Bouldin Index* yang dihasilkan dengan nilai -0,695 menggunakan nilai $K=3$. Tabel 5 adalah nilai perbandingan performa dari algoritma dalam mengolah data.

Tabel 5. Nilai Perbandingan Performance

Nilai K	<i>Davies Bouldin Index</i>
2	-0.738
3	-0.695
4	-0.920
5	-1.008
6	-1.041
7	-1.027
8	-0.989
9	-0.898
10	-0.869
11	-0.893
12	-0.967
13	-0.830

Pengujian performa Algoritma dalam mengolah data dilakukan sebanyak 12 kali. Pengujian pertama dilakukan dengan menentukan nilai $k=2$ menghasilkan nilai *Davies Bouldin Index* -0,738. Kemudian dilanjutkan pengujian performa algoritma dengan menentukan nilai $k=3$ dan seterusnya hingga nilai $k=13$. Untuk mengetahui nilai performa terbaik dalam pengolahan data maka dipilih nilai *Davies Bouldin Index* yang terkecil (mendekati nilai 0). Dari hasil nilai *Davies Bouldin Index* pada tabel 4 dapat dibandingkan bahwa nilai terkecil *Davies Bouldin Index* terdapat pada nilai $k=3$ dengan nilai *Davies Bouldin Index* = -0,695. sehingga dapat disimpulkan bahwa membagi data menjadi 3 *cluster* memiliki nilai keakuratan lebih baik dibandingkan dengan *cluster* yang lainnya.

Hasil penelitian diketahui bahwa membentuk data menjadi 3 *cluster* sejalan dengan penelitian [3, 6, 7, 12, 15–17, 21]. Beberapa penelitian tersebut melakukan pengolahan data menjadi 3 *cluster*. Sebagaimana disarankan pada penelitian [21] perlu dilakukan analisis jumlah *cluster* untuk memperoleh hasil *clustering* yang optimal. Oleh sebab itu, keterbaruan pada penelitian ini melakukan pengujian dengan mendapatkan nilai *Davies Bouldin Index* pada setiap pengujian *cluster*. Pengujian *cluster* sebanyak 12 kali dilakukan mulai dari nilai $K=2$ sampai nilai $K=13$. Hasil pengujian menunjukkan bahwa jumlah *cluster* sebanyak 3 kelompok dengan nilai -0,695 merupakan performa terbaik dibandingkan dengan pengujian menggunakan *cluster* yang lainnya.

4. KESIMPULAN

Pada penelitian ini data awal berjumlah 232 *record* setelah *preprocessing* Data diperoleh sebanyak 224 *record* yang dapat dianalisis. Hasil penelitian memiliki keterbaruan pada pengujian Algoritma *K-Means* dalam mengolah data dengan menunjukkan nilai *Davies Bouldin Index*. Keterbaruan penelitian ini dapat dilihat dari beberapa penelitiannya sebelumnya tidak melakukan pengujian performa Algoritma *K-Means* dalam mengolah data dan diperkuat oleh penelitian [21] tahun 2018 yang menyarankan untuk melakukan analisis dalam penentuan jumlah *cluster*. Berdasarkan hasil pengujian, Algoritma *K-means* memiliki performa terbaik dengan membagi data menjadi 3 *cluster* sehingga menghasilkan nilai *Davies Bouldin Index* = -0,695. Diketahui dari hasil pembagian *cluster* bahwa *cluster* 0 sebanyak 178 sekolah, *cluster* 1 sebanyak 3 sekolah, dan *cluster* 2 sebanyak 43 sekolah. Kemudian dari visualisasi algoritma *K-Means* menunjukkan pemeringkatan *cluster* secara berurutan adalah *cluster* 1, *cluster* 0 dan terakhir adalah *cluster* 2. Sehingga dapat diketahui bahwa terdapat 43 sekolah dalam *cluster* rendah. Untuk itu, penelitian ini merekomendasikan perlu dilakukan pendampingan dan pengadaan kebutuhan sekolah terhadap sekolah *cluster* rendah oleh dinas pendidikan pada pemerintah kabupaten Seruyan. Berdasarkan hasil penelitian yang telah dilakukan, maka penulis menyarankan agar penelitian ini dapat dilanjutkan dengan menganalisis penentuan *cluster* kembali sehingga mendapatkan nilai *Davies Bouldin Index* lebih kecil (mendekati 0).

UCAPAN TERIMA KASIH

Peneliti mengucapkan terimakasih kepada pihak Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Darwan Ali yang telah memberikan izin kepada kami untuk melakukan penelitian mengenai klasterisasi pendidikan SD dan SMP di Kabupaten Seruyan. Kami ucapkan terima kasih pula kepada Dinas Pendidikan Kabupaten Seruyan yang telah mendukung penelitian ini sehingga dapat berjalan hingga selesai. Kemudian kami ucapkan terima kasih pula kepada rekan-rekan dosen Universitas Darwan Ali yang telah memberikan dukungan moril sehingga penelitian ini dapat diselesaikan dengan baik.

REFERENSI

- [1] Badan Pusat Statistik, "Hasil Sensus Penduduk 2020 Indonesia," *Bps.Go.Id*, no. 27, 2021.
- [2] M. Zein, "Peran Guru dalam Pengembangan Pembelajaran," *Jurnal Inspiratif Pendidikan*, vol. 5, no. 2, 2016.
- [3] S. Oktarian, S. Defit, and Sumijan, "Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering," *Jurnal Informasi dan Teknologi*, vol. 2, no. 3, pp. 68–73, 2020.
- [4] D. Solichah, "Hubungan Antara Fasilitas Sekolah, Minat Belajar, Dan Kebiasaan Belajar Siswa MI Al-Huda," *Tech. Rep.*, 2018.
- [5] N. Amah and A. D. Nugroho, "Pengaruh Fasilitas Sekolah Terhadap Hasil Belajar Akuntansi dengan Lingkungan Sosial Sebagai Pemoderasi," *Journal of Accounting and Business Education*, vol. 2, no. 4, 2016.
- [6] M. L. Sibuea and A. Safta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 4, no. 1, 2017.
- [7] Y. A. Priambodo and S. Y. J. Prasetyo, "Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Metode Spatial Clustering K-Means (Studi kasus : Wilayah Provinsi Banten)," *Indonesian Journal of Computing and Modeling*, vol. 1, no. 1, 2018.
- [8] M. A. Amri, A. P. Windarto, A. Wanto, and I. S. Damanik, "Analisis Metode K-Means Pada Pengelompokan Perguruan Tinggi Menurut Provinsi Berdasarkan Fasilitas yang Dimiliki Desa," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, 2019.
- [9] F. Nasari and S. Darma, "Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama)," *Semasteknomedia Online*, vol. 3, no. 1, 2015.
- [10] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, vol. 2, no. 2, 2017.
- [11] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila," *Jurnal Media Infotama*, vol. 11, no. 2, 2015.
- [12] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *SIS-TEMASI : Jurnal Sistem Informasi*, vol. 7, no. 3, 2018.
- [13] T. Baraas, A. Juliansyah, and A. A. Rizal, "Klasifikasi Data Log Intrusion Detection Sistem (Ids) Dengan Decision Tree C4.5," *Jurnal Bumigora Information Technology (BITE)*, vol. 1, no. 2, 2019.
- [14] A. S. Suweleh, D. Susilowati, and Hairani Hairani, "Aplikasi Penentuan Penerima Beasiswa Menggunakan Algoritma C4.5," *Jurnal BITE*, vol. 2, no. 1, pp. 12–21, 2020.
- [15] K. Handoko, "Pengelompokkan Data Mining Pada Jumlah Penumpang di Bandara Hang Nadim," *CBIS (Computer Based Information System Journal)*, vol. 6, no. 2, 2018.
- [16] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi," *Computer Engineering, Science and System Journal*, vol. 3, no. 2, 2018.
- [17] S. Suryadi, "Penerapan Metode Clustering K-Means untuk Pengelompokan Kelulusan Mahasiswa Berbasis Kompetensi," *Jurnal Informatika*, vol. 6, no. 1, 2019.
- [18] W. Gie and D. Jollyta, "Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index : Status Covid-19 Wilayah Riau," 2020.
- [19] T. Khotimah, "Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means," *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 5, no. 1, pp. 83 – 88, 2014.

- [20] I. F. Anshori and Y. Nuraini, "Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Tasikmalaya Menggunakan Algoritma K-Means," *Jurnal Responsif: Riset Sains dan Informatika*, vol. 2, no. 1, 2020.
- [21] G. S. Nugraha, H. Hairani, and R. F. P. Ardi, "Aplikasi Pemetaan Kualitas Pendidikan Di Indonesia Menggunakan Metode K-Means," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 17, no. 2, pp. 13–23, 2018.

