



Pemetaan dan Klasterisasi Sekolah Muhammadiyah di Kabupaten PPU Berdasarkan Fasilitas, Pendidik dan Tenaga Pendidik Menggunakan Metode K-Means Clustering

Muhammad Riadhus Sholihin¹, Rudiman²

^{1,2} Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur

email: 1811102441068@umkt.ac.id¹), rud959@umkt.ac.id²)

INFO ARTIKEL

Sejarah Artikel:

Diterima 29 Oktober 2022
Direvisi 03 Desember 2022
Disetujui 22 Desember 2022
Dipublikasi 30 Desember 2022

Katakunci:

Pemetaan
Pengelompokan
Muhammadiyah

ABSTRAK

Fasilitas Pendidikan merupakan suatu bagian untuk tercapainya pembelajaran di sekolah. Fasilitas yang umumnya berada di sekolah adalah ruang kelas, perpustakaan dan wc yang diperuntukkan untuk guru/tenaga pendidik dan murid. Dengan adanya fasilitas, guru, dan tenaga pendidik sangat penting untuk mendukung berjalannya kegiatan belajar mengajar. Untuk memudahkan pemerintah dan sekolah Muhammadiyah dalam mengklaster sekolah tertentu menjadi beberapa cluster, sehingga memudahkan dalam pendampingan maupun pengadaan kebutuhan sekolah dilingkungan pemerintah Kabupaten Penajam Paser Utara. Klasterisasi dilakukan dengan menggunakan algoritma K-Means. Penerapan Algoritma K-Means dengan menentukan nilai Cluster yaitu 3. Hasil penelitian menunjukkan setiap Cluster memiliki jumlah anggota-anggotanya masing-masing. Pada Cluster 0 terdiri 5 sekolah, Cluster 1 terdiri dari 8 sekolah, dan Cluster 2 terdiri 1 sekolah. Kemudian hasil pengujian performance algoritma K-Means dengan membagi 3 Cluster diperoleh nilai Davies Bouldin Index senilai 0,446. Dari hasil analisis dan pengolahan data terdapat 8 sekolah dalam cluster rendah, sehingga pada penelitian ini merekomendasikan perlu adanya pendampingan dan pengadaan kebutuhan sekolah terhadap cluster rendah oleh dinas Pendidikan pemerintahan Kabupaten Penajam Paser Utara.

ABSTRACT

Educational facilities are a part of achieving learning in schools. Facilities that are generally located in schools are classrooms, libraries and toilets that are intended for teachers/educational staff and students. With the existence of facilities, teachers and teaching staff it is very important to support the running of teaching and learning activities. To make it easier for the government and Muhammadiyah schools to cluster certain schools into several clusters, so as to make it easier to assist and procure school needs in the North Penajam Paser District government environment. Clustering is done 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 members. Cluster 0 consists of 5 schools, Cluster 1 consists of 8 schools, and Cluster 2 consists of 1 school. Then the results of testing the performance of the K-Means algorithm by dividing into 3 clusters obtained a Davies Bouldin Index value of 0.446. From the results of analysis and data processing there were 8 schools in low clusters, so this study recommended the need for assistance and provision of school needs for low clusters by the North Penajam Paser District education office.

©2022 diterbitkan oleh Prodi Teknik Informatika Universitas Yudharta Pasuruan

1. Pendahuluan

Muhammadiyah merupakan salah satu persyarikatan islam yang bergerak dibidang sosial, Pendidikan, dan juga sangat aktif dalam menyebarkan agama islam melalui pendidikan, termasuk sekolah dan madrasah untuk mencerdaskan pengetahuan bangsa [1]. Indonesia mempunyai beberapa jenjang pendidikan yaitu taman kanak-kanak, sekolah dasar, sekolah menengah pertama, dan sekolah menengah kejuruan. Kebutuhan fasilitas di setiap sekolah sangat penting dan diperlukan untuk memberikan pendidikan sekolah yang layak. Diketahui pada bulan September tahun 2020 hasil sensus penduduk Indonesia tercatat mempunyai penduduk berjumlah 270,20 juta jiwa Pada Kalimantan Timur mencapai 3,77 juta jiwa Tahun 2021 pada Kabupaten Penajam Paser Utara tercatat jumlah penduduk 185.022 jiwa [2].

Fasilitas Pendidikan merupakan salah satu indikator tercapainya capaian pembelajaran di sekolah. Pada intinya fasilitas harus dimanfaatkan secara efisiensi terhadap pemakaian fasilitas pembelajaran agar tujuan pembelajaran di sekolah bisa meningkatkan mutu sekolah. Lebih dari itu, yaitu mengarahkan dan memberikan fasilitas belajar agar proses belajar lebih memadai. Dengan begitu, hadirnya fasilitas dan guru sangat diperlukan pada lingkungan sekolah. Bahkan keberadaan tenaga kependidikan juga akan dibutuhkan dalam proses berjalannya Pendidikan di lingkungan sekolah [3].

Dinas Pendidikan Kabupaten Penajam Paser Utara menjadi salah satu penanggung jawab berkaitan pelayanan Pendidikan. Hal ini tercantum pada UU Nomor 23 tahun 2014 pasal 12. Diketahui, melihat perekaman data dari dapodik bahwa pengadaan fasilitas di setiap sekolah Muhammadiyah di Kabupaten Penajam Paser Utara. Dari beberapa sekolah terekam hanya memiliki 2 bahkan 1 guru pada satu sekolah. Tentu ini menjadi perhatian untuk pihak pemerintah setempat agar bisa mengadakan kembali fasilitas, guru, dan tendik secara merata sesuai dengan kebutuhan daerahnya.

Agar membantu pemangku berkepetingan, pengadaan kebutuhan sekolah, maka perlu dilakukan pengelompokan sekolah dengan membuat sistem informasi pemetaan sekolah Muhammadiyah berdasarkan pendidik, dan tenaga pendidik di Kabupaten Penajam Paser Utara, dari hasil pengelompokan data menggunakan *K-Means Clustering* yang nantinya mampu melakukan pengklasifikasian terhadap sekolah Muhammadiyah di Kabupaten Penajam Paser Utara dan mampu menampilkan hasil visualisasi terkait pengklasifikasian tersebut.

2. Kajian Teori

2.1. Data Mining

Data mining adalah perilaku atau proses analisis terhadap proses penemuan pengetahuan di dalam sebuah basis data atau yang *knowledge discovery in databases* (KDD) [4]. Produk pengetahuan dapat berupa pola data atau relasi antar daya yang valid. *Data mining* merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan data yang sangat besar, meliputi metode – metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistics*, dan *database system* [5]. *Data mining* ditujukan untuk mengestrak (mengambil intisari) pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basis data, menejemen data, pemrosesan data, pertimbangan model dan inferensi, ukuran ketertarian, perimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi, dan *online updating* [6]. Tahapan awal dilakukan ada proses *data mining* adalah menyeleksi data dari sumber data ke data target, tahap ke dua yaitu *pre-processing* yang ditunjukkan untuk memperbaiki kualitas data, transformasi, *data mini* serta tahap interpretasi dan aveluasi yang akan menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik [7].

2.2. K.Means

K-means merupakan salah satu metode yang paling mudah digunakan dalam proses pengelompokan. *Centroid* ditentukan dengan menggunakan data pertama yang kemudian akan didefinisikan sebagai *centroid* pertama, hal tersebut juga berlaku untuk data kedua dan seterusnya hingga data ke n dan *centroid* ke n [8]. setelah melalui proses pengelompokan maka tahapan pertama dinyatakan selesai. Tahap selanjutnya adalah menghitung *k centroid* baru yang digunakan sebagai *baycenters* pada kelompok yang dihasilkan [9].

K-means pada dasarnya merupakan konsep aligaratima yang terhitung sederhana, yaitu dengan memperkecil *Sum of Squared Error* (SSE) diantara objek-objek data dengan sejumlah *k centroid* yang sudah di tetapkan di awal. Adapun prinsip utama dari algoritma yaitu menyusun k buah partisi/pusat (*centroid*)/ rata-rata (*mean*) dari sekumpulan data [10]. Langkah- langkah algoritma *k-means* mencakup beberapa poin yaitu :

1. Menentukan jumlah *cluster* (*k*) pada data.
2. Menentukan secara random titik pusat(*centroid*) sebanyak data *k*.
3. Rumus Euclidean Distance di gunakan untuk menghitung jarak terdekat dengan titik pusat(*centroid*) yaitu :

$$D = \sqrt{(x_1 + s_i) + (y_1 + t_i)} \quad (1)$$

Keterangan :

D = *Euclidean Distance*

i = Banyak Objek

x,y = Titik Kordinat Objek

s,t = Titik Kordinat Centroid

4. Perhitungan anggota *cluster* yang sekarang dengan pusat *cluster*. Pusat *cluster* merupakan rata-rata dalam suatu *cluster*. Adapun rumus yang dapat di gunakan yaitu :

$$v_{ij} = \frac{1}{n} \sum_{k=0}^{n_i} x_{kj} \quad (2)$$

Keterangan :

v_{ij} = *Centorid* rata-rata *cluster* ke-I, untuk variabel ke-j

n_i = Jumlah anggota *cluster* ke-i

i, K = Index dari *cluster*

j = Index dari variabel

x_{kj} = Nilai data ke-k dari variabel ke-j. untuk *cluster* tersebut

5. Kemudian lakukan langkah ke 3 dan langkah ke 4 kembali dan itaerasi mencapai *centroid* bernilai optimal.

Metode *k-means clustering* dapat mengolah data dengan bentuk data angka atau integer dari hal tersebut maka data dengan bentuk huruf harus di lakukan inialisasi ke dalam bentuk angka dengan langkah sebagai berikut [11]:

1. Melakukan pengurutan data berdasarkan kemunculan data.
2. Setelah malakan pengurutan berdasarkan frekuensi maka data tersebut dapat di inialisasi dengan frekuensi tertinggi bernilai 1 kemudian 2, 3 dan seterusnya.

2.3. Sistem Informasi Geografis (GIS)

Sistem informasi geografis (SIG) adalah sistem komputer yang di tujuan guna pemeriksaan, pengumpulan, pemanduan dan analisis informasi yang berkaitan dengan permukaan bumi [12]. Terdapat 5 komponen dalam SIG antara lain:

1. *People* yaitu seseorang yang menjalankan, mengoperasikan, mengembangkan dan memperoleh sebuah manfaat dari sistem.
2. *Procedure* yaitu prosedur yang telah dikumpulkan untuk mengelolah data tersebut menjadi informasi.
3. *Software* SIG yaitu *software* yang memiliki kemampuan untuk pengelolaan, penyimpanan, pemrosesan, menganalisis dan penayangan data spesial.
4. *Hardware* SIG yaitu perangkat yang mendukung untuk mengoperasikan perangkat lunak yang akan dipergunakan.

2.4. Quantum GIS

Quantum GIS merupakan perangkat lunak yang termasuk dalam GIS yang memiliki sifat *open source* yang dipopulerkan oleh komunitas yang tergabung dalam qgis.org yang diinisiasi oleh Gary Sherman. Dalam memproses kebutuhan analisis spasial, Quantum GIS dibekali dengan fitur yang dapat mengolah data dengan berbagai macam format dan fungsioanl vektor, raster, dan juga database. Fitur yang terdapat dalam perangkat lunak ini

telah cukup lengkap karena telah diintegrasikan dengan perangkat lunak GRASS. *User Interface* Quantim GIS juga sudah dibuat agar lebih mudah digunakan terutama bagi mereka yang sudah sering menggunakan perangkat lunak ini [13].

2.5. Rapid Miner

Rapid Miner merupakan perangkat yang berbasis *open source*. Rapid Miner digunakan untuk melakukan analisis data dan sebagai alat bantu dalam proses data mining yang dapat terintegrasi pada perangkat lunak itu sendiri [14]. Dengan menggunakan bahasa Java perangkat lunak ini mampu bekerja dan *compatible* dalam semua jenis sistem operasi. Rapid Miner atau yang pada masanya dikenal dengan nama YALE dikembangkan pertama kali oleh Ralf Klöckner, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund pada tahun 2001 yang kemudian di pasarkan dibawah lisensi dari AGPL Ver.3. saat ini ribuan aplikasi telah dikembangkan menggunakan Rapid Miner yang telah tersebar lebih dari 40 negara. Rapid Miner menyediakan GUI untuk merancang sebuah pipeline analitis [15].

2.6. Centroid

Dalam proses pemetaan, *centroid* merupakan bagian penting untuk menentukan titik koordinat dan menyusynnya menjadi data awal dari keberadaan objek. Titik koordinat yang ditentukan pada peta yang merupakan pertemuan antara garis horizontal dan garis vertikal. Dua jenis koordinat yang terdapat pada peta adalah koordinat geografis dan koordinat grid/UTM [16].

2.7. Davies Bouldin Index

Davies-Bouldin Index (DBI) adalah sebuah metode yang digunakan untuk melakukan validasi terhadap *cluster* secara kuantitatif dari hasil *clustering*. Metode ini memiliki tujuan untuk memaksimalkan jarak *intercluster* antara satu *cluster* terhadap *cluster* yang lain. Dalam penelitian ini DBI akan digunakan untuk mendeteksi *outlier* pada masing masing *cluster* yang terbentuk [17].

Sum of Square Between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan:

$$SSB_{ij} = ||C_i - C_j||^2 \quad (3)$$

Sum of Square Within cluster (SSW) merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah *cluster* ke-I yang dirumuskan sebagai berikut:

$$SSW_{ij} = \frac{1}{M_i} \sum_{j=1}^{M_i} d(X_j, C_{pi}) \quad (4)$$

Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran rasio (Rij) untuk mengetahui nilai perbandingan antar *cluster* ke-I dan *cluster* ke-j

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai Davies Bouldin Index (DBI) dari persamaan berikut:

$$DBI = \frac{1}{N} \sum_{i=1}^N \max_{i \neq j} (R_{i,k}) \quad (6)$$

Dari persamaan tersebut, semakin kecil nilai DBI yang diperoleh (non-negatif >=0), maka cluster yang diperoleh semakin bagus/ valid.

3. Metodologi Penelitian

3.1. Metode Dasar Penelitian

Metode ini untuk membantu mengolah data khususnya mengelompokkan data yang memiliki karakteristik yang serupa berdasarkan fasilitas, pendidik, dan tenaga pendidik di kabupaten Penajam Paser Utara yang akan dibagi menjadi 3 *cluster* yaitu lengkap, kurang lengkap, dan sangat kurang lengkap.

3.2. Objek Dan Sumber Data Penelitian

Objek penelitian ini adalah sekolah Muhammadiyah yang berada di Kabupaten Penajam Paser Utara. Dengan menggunakan dua jenis data yaitu data primer yang diperoleh dari kuesioner yang didistribusikan melalui *google form* dan diisi oleh responden dari Perwakilan Wilayah Muhammadiyah dan sekolah Muhammadiyah di Kabupaten Penajam Paser Utara. Dan menggunakan data sekunder yang di peroleh dari berbagai sumber. Berikut adalah atribut yang digunakan pada penelitian ini.

Tabel 1. Form Atribut

No.	Atribut	Text	Numerik
1	Nama Sekolah	√	
2	Kecamatan	√	
3	Kabupaten	√	
4	Alamat	√	
5	Akreditasi	√	
6	Kepala Sekolah	√	
7	Status Sekolah	√	
8	Jumlah Bangunan Perpustakaan		√
9	Jumlah TOILET Guru		√
10	Jumlah MCK Siswa		√
11	Jumlah Kelas		√
12	Jumlah Guru		√
13	Jumlah Tenaga Kependidikan		√
14	Jumlah Siswa		√

3.3. Tahapan Penelitian

Penelitian ini dilakukan tahapan-tahapan sebagai berikut :

1. Melakukan wawancara dengan pimpinan Wilayah Muhammadiyah Kalimantan timur terkait data sekolah yang berada di Kabupaten Penajam Paser Utara.
2. Mengumpulkan data sekolah Muhammadiyah dan data spasial peta administrasi Kabupaten Penajam Paser Utara.
3. Melakukan analisis dan penyempurnaan data yang telah dilakukan sebelum pengelompokan data dengan menggunakan metode *k-means clustering*.
4. Melakukan evaluasi *clustering*.
5. Melakukan perancangan peta digital

4. Hasil Uji Coba Dan Pembahasan

4.1. Pengolahan Dan Analisa Data

Data yang diperoleh dari proses wawanca dengan Pimpinan Wilayah Muhammadiyah yang berkaitan dengan alamat sekolah kemudian dirubah menjadi koordinat untuk mempermudah dalam proses pelacakan pada *Google Maps*. Maka dihasilkan koordinat sebagai berikut :

Tabel 2. Koordinat Atribut

No.	Jenis Sekolah	X	Y
1	TK ABA 1 PPU	116.75400188134800	-1.275552209861100
2	TK ABA 1 PPU	116.46259867658700	-1.5033653355121500
3	SD Muh 1 PPU	116.75115039965200	-1.2761845568290800
4	SD Muh 1 PPU	116.45833954493000	-1.4931156848965000
5	SMP Muh 1 PPU	116.76108174732500	-0.9187669375555370
6	SMP Muh 2 PPU	116.45564616324000	-1.5047275500152100
7	SMK Muh 2 PPU	116.66414402113200	-1.0357559912159000
8	SMK Muh 1 PPU	116.76362069875000	-126232423775170000

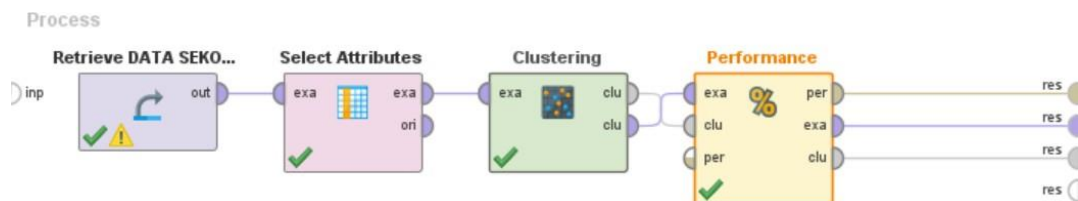
Proses ini merupakan proses penyempurnaan dan penyatuan dataset yang akan digunakan pada proses *clustering*. Pada tahap ini dataset diolah dengan data tranformasi dimana atribut yang ditranformasi merupakan data sekolah yang akan ditambahkan atribut baru yaitu jumlah guru dan jumlah siswa. Berikut adalah hasil data yang telah ditranformasi.

Tabel 3. Data Sekolah

No.	Nama Sekolah	X	Y
1	TK ABA 1 PPU	3	19
2	TK ABA 1 PPU	5	60
3	SD Muh 1 PPU	8	170
4	SD Muh 1 PPU	8	139
5	SMP Muh 1 PPU	13	201
6	SMP Muh 2 PPU	8	131
7	SMK Muh 2 PPU	14	143
8	SMK Muh 1 PPU	23	301

Nilai atribut diatas didapatkan dengan menyamakan setiap baris data sekolah yang berada pada tabel 2 dan menghitung jumlah kelurahan dan kecamatan secara otomatis.

4.2. Proses Clustering



Gambar 1. Desain Proses Clustering

Gambar 1 merupakan seluruh proses tahapan *clustering* yang dilaksanakan dengan 4 tahapan

1. **Import Data** merupakan proses import data kedalam rapid miner, data yang ada merupakan satu data dengan 14 atribut seperti yang ditampilkan pada tabel 1.
2. **Pemilihan Data** dimana data yang akan diorises diseleksi hingga menghasilkan enam atribut yaitu jumlah guru, jumlah kelas, jumlah bangunan perpustakaan, jumlah wc guru, jumlah wc siswa, dan jumlah tenaga kependidikan.
3. **Pengelompokan Data** dengan menerapkan alogaritma *K-Means*, yaitu jumlah data set sebanyak n dan jumlah insialisasi *centroid k* yang ditentukan secara acak sehingga membentuk 14 data penelitian. Paramater yang digunakan dalam proses ini adalah :
 Jumlah Cluster : 3
 Jumlah Data : 14
 Jumlah Atribut : 14
4. **Cluster Distance Performance** yang bertujuan mengambail cluster centroid dan cluster set untuk input dan meilhat kinerja model dengan centroid.

Setelah semua proses dijalankan, maka *rapid miner* menghasilkan sebagai berikut :

Row No.	id	label	Jumlah Rua...	Jumlah WC ...	Jumlah WC ...	Jumlah Kelas	Jumlah Guru	Tendik
1	1	cluster_1	1	1	1	4	3	1
2	2	cluster_1	0	1	1	3	5	1
3	3	cluster_1	0	1	1	2	2	1
4	4	cluster_1	0	1	1	2	2	1
5	5	cluster_1	1	1	1	2	2	1
6	6	cluster_1	0	1	1	2	2	1
7	7	cluster_1	0	1	1	3	2	1
8	8	cluster_1	1	2	1	6	8	2
9	9	cluster_0	1	1	1	10	8	4
10	10	cluster_0	1	1	1	16	13	5
11	11	cluster_0	1	1	2	8	8	5
12	12	cluster_0	1	1	1	16	13	5
13	13	cluster_0	1	1	8	14	14	4
14	14	cluster_2	2	2	11	23	23	8

Gambar 2. Example Set Result

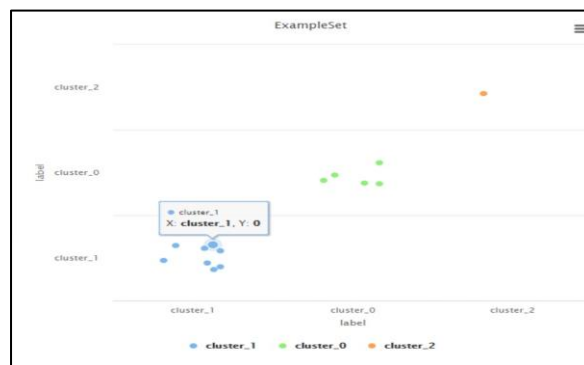
Example Set Result memperlihatkan hasil data klaster yang diberi label dan dibagi menjadi tiga yaitu cluster_0, cluster_1 dan cluster_2. berdasarkan hasil data dengan jarak terdekat

Cluster Model

```
Cluster 0: 5 items
Cluster 1: 8 items
Cluster 2: 1 items
Total number of items: 14
```

Gambar 3. Cluster Model

Dari gambar 3 selanjutnya dapat dihasilkan pembagian data dimana cluster_0 memiliki 5 sekolah, cluster_1 memiliki 8 sekolah dan cluster_2 memiliki 1 sekolah. semua dataset di cluster terdapat 14 data yang ditampilkan dalam scatter plot sebagai berikut :



Gambar 4. Plot Cluster Model

Dari hasil scatter plot dapat dilihat bahwa cluster_0 merupakan fasilitas yang kurang lengkap, cluster_2 merupakan sekolah dengan fasilitas lengkap, cluster_1 merupakan cluster dengan fasilitas yang sangat kurang lengkap.

Tabel 4. Data Hasil Cluster

No.	Nama Sekolah	Label
1	TK ABA 1 PPU	cluster_1
2	TK ABA 1 PPU	cluster_1
3	TK ABA 2 PPU	cluster_1
4	TK ABA 3 PPU	cluster_1
5	TK ABA 1 PPU	cluster_1
6	TK ABA 2 PPU	cluster_1
7	TK ABA 3 PPU	cluster_1
8	SD Muhammadiyah 1 PPU	cluster_1
9	SD Muhammadiyah 1 PPU	cluster_0
10	SMP Muhammadiyah 1 PPU	cluster_0
11	SMP Muhammadiyah 2 PPU	cluster_0
12	SMP Muhammadiyah 1 PPU	cluster_0
13	SMK Muhammadiyah 2 PPU	cluster_0
14	SMK Muhammadiyah 1 PPU	cluster_2

Dari tabel diatas bisa dilihat jumlah sekolah berdasarkan jumlah ruangan kelas sekolah, jumlah guru perempuan, jumlah wc murid pria, jumlah perpustakaan, jumlah jumlah kependidikan dicluster menjadi 3 cluster. Paling lengkap fasilitas cluster_2, yaitu SMK Muhammadiyah 1 PPU, kurang

lengkap cluster_0, yaitu SD Muhammadiyah 1 PPU, SMP Muhammadiyah 1 PPU, SMP Muhammadiyah 2 PPU, SMP Muhammadiyah 1 PPU, dan SMK Muhammadiyah 2 PPU dan sangat kurang lengkap cluster_1 yaitu TK ABA 1 PPU, TK ABA 1 PPU, TK ABA 2 PPU, TK ABA 3 PPU, TK ABA 1 PPU, TK ABA 2 PPU, dan TK ABA 3 PPU.

Selanjutnya dilakuakn proses Davies Bouldin (DBI) pada data sekolah yang telah diproses dan mendapatkan hasil sebesar 0,446. Dari hasil yang telah ditampilkan oleh performancevector dari cluster_2, cluster_0 dan cluster_1 dapat dilihat pada gambar dibawah ini.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 12.652
Avg. within centroid distance_cluster_0: 25.200
Avg. within centroid distance_cluster_1: 6.391
Avg. within centroid distance_cluster_2: 0.000
Davies Bouldin: 0.446
```

Gambar 5. Hasil Performancevector

validasi DBI pada sekolah Muhammadiyah berdasarkan fasilitas, pendidik, dan tenaga pendidik. Didapatkan hasil nilai 0.446 dari hasil validasi yang ditampilkan nilai ini bisa disimpulkan baik karena nilainya adalah positif mendekati nilai 0. Cluster yang baik adalah cluster yang memiliki nilai kohesi sekecil mungkin dan separasi yang besar. Semakin kecil nilai DBI yang diperoleh (non-negatif >=0), maka cluster yang diperoleh semakin bagus/ valid.

4.3. Perancangan Peta Digital

Pada tahapan ini proses perancangan peta digital menggunakan alat bantu Quantum GIS untuk menampilkan hasil peda dan mengolah data peda administrasi dengan tahapan sebagai berikut :

1. Import data file SHP ke dalam Quantum GIS dengan menggunakan vector.
2. Masukan data SHP dengan data atribut untuk melakukan seleksi agar dapat difokuskan pada Kabupaten Penajam Paser Utara.
3. Selanjutnya semua data dan vektor dikelompokan berdasarkan kelurahan yang ada di Kabupaten Penajam Paser Utara yang divisualisasikan dengan menggunakan perbedaan warna.



Gambar 6. Pembagian Desa/Kelurahan

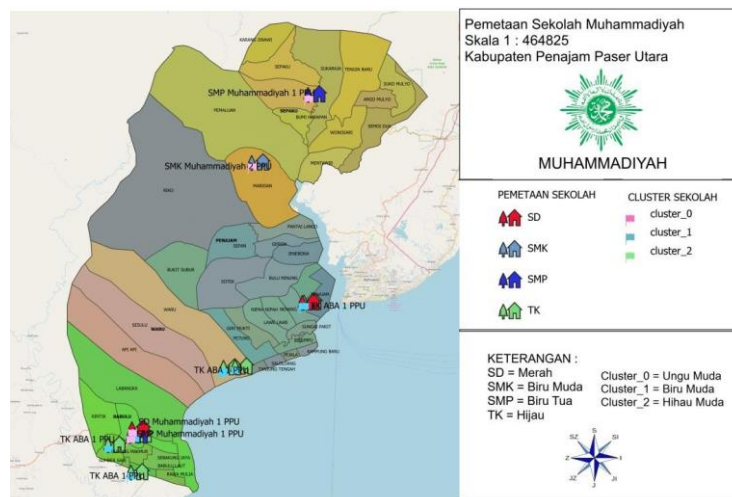
4.4. Hasil Pemetaan dan Clustering

Hasil cluster yang terdiri dari 3 cluster dari tingkat penyebarannya. Data tersebut kemudian digabungkan berdamaan dengan menggunakan tools Quantum GIS yang menggunakan atribut yang nilainya bisa menghasilkan peta dan cluster. Berikut adalah hasil data yang telah ditambahkan atribut dan label.

Juru Peremimlah WC Siswa La	Jumlah Kelas	Jumlah Guru	Tendik	2017/2018	2018/2019	2019/2020	2020/2021	2021/2022	X	Y	label	
1	1	4	3	1	7	10	20	15	15	116.7540019	-1.27555221	cluster_1
1	1	3	5	1	15	20	57	40	60	116.4625987	-1.50336336	cluster_1
1	1	2	2	1	22	34	56	30	48	116.4565465	-1.567870139	cluster_1
1	1	2	2	1	6	12	15	20	21	116.4172739	-1.527004478	cluster_1
1	1	2	2	1	9	31	27	32	37	116.6130602	-1.391021591	cluster_1
1	1	2	2	1	5	8	10	16	21	116.6357214	-1.386260436	cluster_1
1	1	3	2	1	14	23	26	37	35	116.6156347	-1.38770451	cluster_1
2	1	6	8	2	120	177	196	156	170	116.7511504	-1.276184557	cluster_1
1	1	10	8	4	101	111	155	120	139	116.4583395	-1.493115685	cluster_0
1	1	16	13	5	140	167	123	153	201	116.7610817	-0.918766938	cluster_0
1	2	8	8	5	152	132	122	139	141	116.4556462	-1.50472755	cluster_0
1	1	16	13	5	120	149	134	167	201	116.7610817	-0.918766938	cluster_0
1	8	14	14	4	119	132	102	130	143	116.664144	-1.035755991	cluster_0
2	11	23	23	8	180	290	254	288	301	116.7636207	-1.26232e+16	cluster_2

Gambar 7. Data Sekolah Setelah Penambahan Atribut dan Label

Hasil akhir dari proses ini akan menghasilkan peta penyebaran sekolah Muhammadiyah di Kabupaten Penajam Paser Utara yang ditampilkan sebagai berikut :



Gambar 8. Penyebaran Sekolah Muhammadiyah di Kabupaten Penajam Paser Utara

Dari penelitian yang telah diselesaikan keluarlah hasil pada sekolah Muhammadiyah yang ada di Kabupaten Penajam Paser Utara memiliki 14 sekolah yang tersebar di Kabupaten Penajam Paser Utara dengan persebaran yang terbagi dari kecamatan dan kelurahan kemudian divisualisasikan dengan menggunakan aplikasi Rapid Miner dan Quantum GIS, maka dengan itu terlihatlah semua pembagian sekolah berdasarkan fasilitas, pendidik, dan tenaga pendidik yang telah di cluster dimasing-masing setiap sekolah.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dikerjakan, dapat disimpulkan sebagai berikut :

1. Pemetaan sekolah Muhammadiyah di Kabupaten Penajam Paser Utara menggunakan algoritma k-means clustering berdasarkan fasilitas, pendidik, tenaga pendidik menghasilkan 3 cluster : cluster 2 tingkat terbanyak terdapat 1 sekolah, cluster 1 tingkat sedang terdapat 8 sekolah, cluster 0 terdapat 5 sekolah.
2. Berdasarkan hasil pengujian, Algoritma K-Means memiliki performa terbaik dengan membagi data menjadi 3 cluster sehingga menghasilkan nilai davies bouldin index = 0.446. Diketahui hasil dari pembagian cluster 2 tingkat terbanyak fasilitasnya adalah SMK Muhammadiyah 1 PPU Kec. Babulu, Cluster 0 tingkat sedang adalah SD Muhammadiyah 1 PPU Kec. Babulu, SMP Muhammadiyah 1 PPU Kec. Babulu, SMP Muhammadiyah 2 PPU Kec. Babulu, SMP Muhammadiyah 1 PPU Kec. Sepaku, SMK Muhammadiyah 2 PPU Kec. Penajam, dan cluster 1 tingkat sedikit adalah TK ABA 1 PPU Kec. Penajam, TK ABA 1 PPU Kec. Babulu, TK ABA 2 PPU Kec. Babulu, TK ABA 3 PPU Kec. Babulu, TK ABA 1 PPU Kec. Waru, TK ABA 2 PPU Kec. Waru, TK ABA 3 PPU Kec. Waru, SD Muhammadiyah 1 PPU Kec. Penajam.

6. Daftar Pustaka

- [1] Setiawan, F. Setiawan (2021) *Muhammadiyah mencerdaskan bangsa*. Yogyakarta, UAS PRESS.
- [2] Badan Pusat Statistik (2021). *BPS: 270,20 juta Penduduk Indonesia Hasil SP2020*. Badan Pusat Statistik. <https://www.bps.go.id/news/2021/01/21/405/bps--270-20-juta-penduduk-indonesia-hasil-sp2020.html>
- [3] HM, M. A. (2020). Peran Guru dalam Pengembangan Pembelajaran. *el-Idarah: Jurnal Manajemen Pendidikan Islam*, 5(1), 44-59.
- [4] Agushinta, D., & Murniyati, M. (2022). Penerapan Algoritma K-Means Clustering Pada Data Keluhan Pelanggan PT. PLN (Persero). *Journal of Information System, Applied, Management, Accounting and Research*, 6(2), 327-340.
- [5] Rahmawati, L., Widya Sihwi, S., & Suryani, E. (2016). *Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret)*. *Jurnal Teknologi & Informasi IT Smart*, 3(2), 66. <https://doi.org/10.20961/its.v3i2.654>
- [6] Suyanto, 2017, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.
- [7] Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Ddb) Berdasarkan Provinsi. *Computer Engineering, Science And System Journal*, 3(2), 173.
- [8] Nurahman, N., Purwanto, A., & Mulyanto, S. (2022). Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(2), 337-350. <https://doi.org/10.30812/matrik.v21i2.1411>
- [9] Sibuea, F. L., & Sapta, A. (2017). *Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering*. *Jurnal Teknologi Dan Sistem Informasi*, 1, 10.
- [10] 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.
- [11] Hakim, L., & Prayoga, R. H. (2017). PENGENALAN IRIS MATA METODE TIGA KELAS DENGAN ADAPTIVE K-MEANS CLUSTERING. *Explore IT! : Jurnal Keilmuan Dan Aplikasi Teknik Informatika*, 9(2). <https://doi.org/10.35891/explorit.v9i2.791>
- [12] Rhind, D. (1988). A GIS research agenda. 2.
- [13] Budiyanto, E. (2016). Sistem informasi geografis dengan Quantum GIS.
- [14] 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.
- [15] Hamzah, R., Marpaung, S., & Prayogo, T. (2017). Metode Penentuan Titik Koordinat Zona Potensi Penangkapan Ikan Pelagis Berdasarkan Hasil Deteksi Termal Front Suhu Permukaan Laut. *Jurnal Penginderaan Jauh Dan Pengolahan Data Citra Digital*, 13(2). <https://doi.org/10.30536/j.pjpdcd.2016.v13.a2364>
- [16] Dimas, D., Nurjayadi, N., & Haryono, D. (2019). Penerapan Augmented Reality Pada Informasi Data Peta Kawasan Hutan Lindung Menggunakan Metode Marker. *SATIN - Sains Dan Teknologi Informasi*, 4(2), 100.
- [17] Rahmawati, L., Widya Sihwi, S., & Suryani, E. (2016). *Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret)*. *Jurnal Teknologi & Informasi IT Smart*, 3(2), 66. <https://doi.org/10.20961/its.v3i2.654>