

ANALISIS CLUSTER FAKTOR PENUNJANG PENDIDIKAN MENGGUNAKAN ALGORITMA K-MEANS (STUDI KASUS: KABUPATEN KARAWANG)

Abdussalam Amrullah¹, Intan Purnamasari², Betha Nurina Sari³, Garno⁴, Apriade Voutama⁵

¹²³⁴⁵Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang

Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Jawa Barat 41361

¹abdussalam.amrullah18179@student.unsika.ac.id,

²intan.purnamasari@staff.unsika.ac.id, ³betha.nurina@staff.unsika.ac.id, ⁴garno@staff.unsika.ac.id,

⁵apriade.voutama@staff.unsika.ac.id

Abstract

Education is one of the means to create and improve the quality of better human resources. This is expected to improve human welfare. Based on data from the Ministry of Education and Culture, there are 4 sub-districts in Karawang Regency that do not have state high schools. This can result in difficulties for students who have financial deficiencies which can ultimately lead to dropping out of school. Then apart from that distance can also be an obstacle. The purpose of this study is to apply the clustering method for the distribution of educational supporting factors in Karawang District so that later this research can improve the quality of education evenly in Karawang District, not only concentrated in certain areas by paying attention to educational supporting factors. The algorithm used in this research is K-Means. The elbow method used in determining the optimal cluster supported by the silhouette method resulted in the best number of clusters being 2 clusters. The results of the clustering evaluation resulted in Davies-Bouldin Index (DBI) value of 0.408 and Silhouette Coefficient value of 0.747 (strong structure). Cluster 1 consists of 7 sub-districts and Cluster 2 consists of 23 sub-districts. Based on the results of clustering analysis, Cluster 1 has the average number of attributes of schools, teachers, classes, laboratories, libraries at all levels of education (SD, SMP, SMA, SMK, and SLB) with state status and the population shows higher results if compared with the average number of each attribute in Cluster 2. So it can be concluded that Cluster 1 is a high category and Cluster 2 is a low category.

Keywords : *education, clustering, k-means, karawang district*

Abstrak

Pendidikan merupakan salah satu sarana untuk menciptakan dan meningkatkan kualitas sumber daya manusia yang lebih baik. Hal ini diharapkan dapat meningkatkan kesejahteraan manusia. Berdasarkan data Kementerian Pendidikan dan Kebudayaan, terdapat 4 kecamatan di Kabupaten Karawang yang belum memiliki SMA negeri. Hal ini dapat mengakibatkan kesulitan bagi siswa yang memiliki kekurangan keuangan yang pada akhirnya dapat menyebabkan putus sekolah. Kemudian selain dari itu jarak juga bisa menjadi kendala. Tujuan dari penelitian ini adalah menerapkan metode klasterisasi untuk persebaran faktor pendukung pendidikan di Kabupaten Karawang agar nantinya penelitian ini dapat meningkatkan mutu pendidikan secara merata di Kabupaten Karawang, tidak hanya terkonsentrasi pada wilayah tertentu dengan memperhatikan faktor pendukung pendidikan. Algoritma yang digunakan dalam penelitian ini adalah *K-Means*. Metode elbow yang digunakan dalam menentukan cluster optimal yang didukung dengan metode *Silhouette* menghasilkan jumlah *cluster* terbaik menjadi 2 cluster. Hasil evaluasi clustering menghasilkan nilai *Davies-Bouldin Index* (DBI) sebesar 0,408 dan nilai *Silhouette Coefficient* sebesar 0,747 (struktur kuat). Klaster 1 terdiri dari 7 kecamatan dan Klaster 2 terdiri dari 23 kecamatan. Berdasarkan hasil analisis clustering, cluster 1 rata-rata memiliki jumlah atribut sekolah, guru, kelas, laboratorium, perpustakaan pada semua jenjang pendidikan (SD, SMP, SMA, SMK, dan SLB) dengan status status dan populasi menunjukkan hasil yang lebih tinggi jika dibandingkan dengan jumlah rata-rata masing-masing atribut pada Cluster 2. Sehingga dapat disimpulkan bahwa Cluster 1 termasuk kategori tinggi dan Cluster 2 termasuk kategori rendah.

Kata kunci : pendidikan, klasterisasi, k-means, kabupaten karawang

tion is one of the means to create and improve the quality of better human resources. This is expected to improve human welfare. Based on data from the Ministry of Education and Culture, there are 4 sub-districts in Karawang Regency that do not have state high schools. This can result in difficulties for students who have financial deficiencies which can ultimately lead to dropping out of school. Then apart from that distance can also be an obstacle. The purpose of this study is to apply the clustering method for the distribution of educational supporting factors in Karawang District so that later this research can improve the quality of education evenly in Karawang District, not only concentrated in certain areas by paying attention to educational supporting factors. The algorithm used in this research is K-Means. The elbow method used in determining the optimal cluster supported by the silhouette method resulted in the best number of clusters being 2 clusters. The results of the clustering evaluation resulted in Davies-Bouldin Index (DBI) value of 0.408 and Silhouette Coefficient value of 0.747 (strong structure). Cluster 1 consists of 7 sub-districts and Cluster 2 consists of 23 sub-districts. Based on the results of clustering analysis, Cluster 1 has the average number of attributes of schools, teachers, classes, laboratories, libraries at all levels of education (SD, SMP, SMA, SMK, and SLB) with state status and the population shows higher results if compared with the average number of each attribute in Cluster 2. So it can be concluded that Cluster 1 is a high category and Cluster 2 is a low category.

Keywords : *education, clustering, k-means, karawang district*

Abstrak

Pendidikan merupakan salah satu sarana untuk menciptakan dan meningkatkan kualitas sumber daya manusia yang lebih baik. Hal ini diharapkan dapat meningkatkan kesejahteraan manusia. Berdasarkan data Kementerian Pendidikan dan Kebudayaan, terdapat 4 kecamatan di Kabupaten Karawang yang belum memiliki SMA negeri. Hal ini dapat mengakibatkan kesulitan bagi siswa yang memiliki kekurangan keuangan yang pada akhirnya dapat menyebabkan putus sekolah. Kemudian selain dari itu jarak juga bisa menjadi kendala. Tujuan dari penelitian ini adalah menerapkan metode klusterisasi untuk persebaran faktor pendukung pendidikan di Kabupaten Karawang agar nantinya penelitian ini dapat meningkatkan mutu pendidikan secara merata di Kabupaten Karawang, tidak hanya terkonsentrasi pada wilayah tertentu dengan memperhatikan faktor pendukung pendidikan. Algoritma yang digunakan dalam penelitian ini adalah *K-Means*. Metode elbow yang digunakan dalam menentukan cluster optimal yang didukung dengan metode *Silhouette* menghasilkan jumlah *cluster* terbaik menjadi 2 cluster. Hasil evaluasi clustering menghasilkan nilai *Davies-Bouldin Index* (DBI) sebesar 0,408 dan nilai *Silhouette Coefficient* sebesar 0,747 (struktur kuat). Klaster 1 terdiri dari 7 kecamatan dan Klaster 2 terdiri dari 23 kecamatan. Berdasarkan hasil analisis clustering, cluster 1 rata-rata memiliki jumlah atribut sekolah, guru, kelas, laboratorium, perpustakaan pada semua jenjang pendidikan (SD, SMP, SMA, SMK, dan SLB) dengan status status dan populasi menunjukkan hasil yang lebih tinggi jika dibandingkan dengan jumlah rata-rata masing-masing atribut pada Cluster 2. Sehingga dapat disimpulkan bahwa Cluster 1 termasuk kategori tinggi dan Cluster 2 termasuk kategori rendah.

Kata kunci : *pendidikan, klusterisasi, k-means, kabupaten karawang*

1. PENDAHULUAN

Pendidikan merupakan salah satu sarana untuk menciptakan dan meningkatkan kualitas sumber daya manusia yang lebih baik. Hal ini diharapkan dapat meningkatkan kesejahteraan manusia. Pembangunan nasional atau khususnya peningkatan pembangunan ekonomi juga salah satu peranan pendidikan [1]. Cita-cita tersebut juga merupakan harapan bangsa yang bertujuan untuk meningkatkan kualitas sumber daya manusia yang berdaya guna demi terciptanya masyarakat yang makmur, adil dan sejahtera salah satu jalannya yaitu melalui jalur pendidikan.

Terdapat banyak faktor yang dapat mempengaruhi pendidikan salah satunya sarana dan prasarana. Penelitian ini bertujuan untuk membuktikan apakah peningkatan kualitas sarana dan prasarana dapat mempengaruhi minat belajar siswa dalam pembelajaran pendidikan agama islam. Berdasarkan hasil penelitian tersebut diperoleh nilai korelasi sebesar $r = 0,405$ atau sebesar 40.5%, dan angka tersebut terletak pada angka 0,410-0,60, yang membuktikan bahwa terdapat korelasi antara variabel X (sarana dan prasarana) dan variabel Y (minat siswa) itu sedang. Sedangkan sisanya yaitu 59.5% minat siswa dipengaruhi oleh faktor lain [2]. Dengan mengatasi faktor-faktor penunjang tersebut diharapkan dapat

meningkatkan kualitas pendidikan di Indonesia khususnya di Kabupaten Karawang dengan melakukan pemerataan kualitas pendidikan di setiap daerahnya.

Berdasarkan data dari kemdikbud terdapat 4 kecamatan di Kabupaten Karawang yang tidak memiliki sekolah menengah atas negeri, antara lain Kecamatan Rengasdengklok, Purwasari, Cilamaya Kulon, dan Cilebar. Selain hanya terdapat 2 sekolah luar biasa negeri.

Data Mining merupakan teknik pencarian dalam suatu *big data* yang bertujuan menemukan suatu pengetahuan yang berdasarkan pola yang ada [3]. Fungsi *data mining* adalah mengidentifikasi kesimpulan yang diperoleh dalam proses pencarian pola yang terdapat pada suatu kumpulan data [4]. *Clustering* merupakan suatu metode pada *data mining* yang bertujuan untuk mengelompokkan objek-objek kedalam *cluster-cluster* yang memiliki kemiripan karakteristik satu dengan yang lainnya di dalam *cluster* yang sama dan memiliki perbedaan karakteristik pada objek yang berada pada *cluster* lain [5]. Terdapat banyak algoritma yang digunakan dalam *clustering*, yang sering digunakan yaitu *K-Means* dan *K-Medoids*. Terdapat banyak penelitian yang membandingkan antara *K-Means* dan *K-Medoids*. Penelitian yang membandingkan kedua algoritma tersebut dengan menggunakan dataset Covid-19 menghasilkan nilai DBI sebesar 0,064 dan nilai DBI sebesar 0,411 pada algoritma *K-Medoids* [6].

Terdapat berbagai penelitian *clustering* di bidang pendidikan yang menggunakan algoritma *K-Means*. Penelitian yang dilakukan dengan memetakan penyebaran guru yang ada di Provinsi Banten. Berdasarkan penelitian tersebut, terdapat 1 kabupaten/kota yang memiliki kelebihan jumlah guru SD yaitu Tangerang. Sedangkan kabupaten yang memiliki kekurangan guru di tingkat SD yaitu berjumlah 5 kabupaten/kota. Kota Tangerang juga memiliki kelebihan guru pada tingkat SMP dan 5 kabupaten/kota mengalami kekurangan guru di tingkat SMP. Dan yang terakhir terdapat 2 daerah yang memiliki kelebihan guru di tingkat SMA, yaitu Kabupaten Tangerang dan Kota Tangerang [7]. Selanjutnya pada penelitian dengan menggunakan algoritma *K-Means* dalam mengelompokkan utamanya yaitu siswa. Berdasarkan penelitian tersebut, rata-rata murid mengambil jurusan rekayasa perangkat lunak dan sebagian siswa gagal dikelompokkan [8].

Oleh karena itu pada penelitian ini akan melakukan *clustering* faktor penunjang Pendidikan yang berada di Kabupaten Karawang

menggunakan algoritma *K-Means*. Penelitian ini diharapkan dapat membantu peningkatan kualitas Pendidikan secara merata tidak hanya terfokus pada daerah tertentu saja

2. TINJAUAN PUSTAKA DAN TEORI

2.1. Tinjauan Pustaka

Penelitian yang telah dilakukan sebelumnya yaitu terkait objek penelitian pengelompokan faktor pendukung pendidikan. Penelitian ini menggunakan algoritma *K-Means* dengan menggunakan 6 atribut, antara lain jumlah sekolah dan jumlah guru. Penelitian tersebut menghasilkan 2 *cluster*, yaitu *cluster* 1 yang terdiri dari 21 kecamatan dan *cluster* 2 terdiri dari 11 kecamatan. Hasil ini didapatkan dilakukan 5 kali iterasi. Pada penelitian ada terdapat adanya metode evaluasi dari hasil *clustering* tersebut [1]. Penelitian selanjutnya yang serupa yaitu memetakan penyebaran guru di Provinsi Banten [7]. Penelitian tersebut menggunakan metode *Spatial Clustering K-Means*. Hasil dari penelitian ini yaitu berupa pemetaan daerah yang memiliki kekurangan, kecukupan dan kelebihan.

Berdasarkan penelitian-penelitian yang tersebut maka pada penelitian ini akan dilakukan analisis *cluster* pemerataan faktor penunjang pendidikan di Kabupaten Karawang. Dalam Penelitian ini menggunakan algoritma *K-Means* dengan jumlah atribut yang digunakan yaitu 26 atribut. Atribut tersebut antara lain jumlah sekolah, jumlah guru, jumlah kelas, jumlah laboratorium, dan jumlah perpustakaan pada jenjang SD, SMP, SMA, SMK, dan SLB berstatus negeri dan jumlah penduduk di Kabupaten Karawang pada tahun 2021.

2.2. Knowledge Discovery on Database (KDD)

KDD merupakan suatu metodologi penelitian yang biasa digunakan dalam proses *data mining*. KDD merupakan suatu metode penelitian yang bertujuan mendapatkan informasi dalam suatu *database*[9]. Metode KDD terdiri dari *Selection*, *Preprocessing*, *Transformation*, *Data mining*, dan *Interpretation/Evaluation* [10].

2.3. K-Means

K-Means merupakan metode *partitional clustering* atau pengelompokan non-hirarki (*clustering non-hierarchical*) yang digunakan untuk menganalisis data atau metode data mining yang bersifat *unsupervised learning* dan mengelompokkan data dengan sistem partisi

[11]. Tahap pertama pada algoritma *K-Means* yaitu menentukan jumlah kluster dan inisiasi titik *cluster*. Selanjutnya tahap kedua yaitu Memasukan semua data ke cluster terdekat. Jarak kedua objek dapat menentukan kedekatan tersebut. Untuk menghitung kedekatan objek tersebut yaitu menggunakan teori jarak Euclidean yang dirumuskan [12] sebagai berikut:

$$d_i = \sqrt{\sum_i^p (x_{2i} - x_{1i})^2}$$

Yang dimana:

d = jarak

p = dimensi data

X_{2i} = data latih

X_{1i} = data uji

Ketiga yaitu Menghitung kembali nilai titik pusat menggunakan mean dari semua data yang berada dalam cluster tertentu. Dan terakhir Mengulangi langkah 2 dan 3 sampai nilai dari titik pusat tidak berubah lagi. Jika sudah tidak berubah, artinya proses clustering telah selesai.

2.4. Elbow Method

K- Elbow Method merupakan metode yang sering digunakan untuk menentukan jumlah cluster terbaik dalam suatu dataset yang akan digunakan dalam proses clustering. Apabila terdapat nilai yang menunjukkan sudut pada grafik atau nilainya menunjukkan penurunan paling signifikan berarti jumlah cluster tersebut merupakan jumlah cluster terbaik [13].

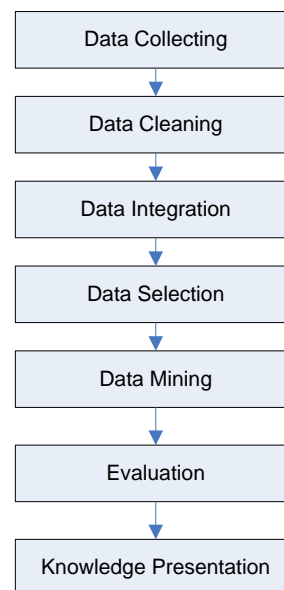
2.5. Davies-Bouldin Index

Davies Bouldin Index merupakan salah satu metode evaluasi pada *clustering* dengan menganalisa kuantitas dan kedekatan antar data hasil *cluster*. Dalam suatu pengelompokkan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat dari kluster yang diikuti. Sedangkan pemisahan didasarkan pada jarak antar titik pusat dari klasternya. Jumlah *cluster* terbaik ditunjukkan dengan nilai DB yang semakin kecil [14].

3. METODOLOGI PENELITIAN

3.1. Skema Alur Penelitian

Proses dan tahapan penelitian digambarkan pada Gambar 1. Tahapan penelitian yang digunakan yaitu menggunakan metodologi KDD atau *Knowledge Discovery on Database*.



Gambar 1. Diagram Alir

3.2. Pengumpulan Data

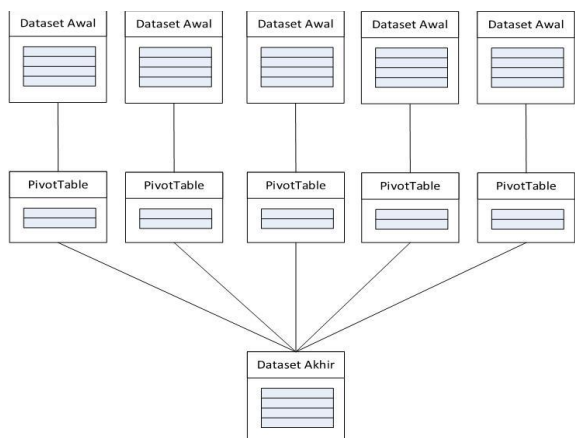
Proses pengumpulan data dari penelitian ini adalah mengambil dataset dari situs dapo.kemendikbud.go.id dan karawangkab.bps.go.id. Jumlah dataset yang digunakan yaitu 30 dataset yang berasal dari situs kemendikbud dan 1 dataset yang berasal dari BPS Kabupaten Karawang.

3.3. Data Cleaning

Pada tahapan *data cleaning*, 30 dataset yang telah diperoleh pada situs dapo.kemendikbud.go.id dan 1 dataset yang berasal dari situs karawangkab.bps.go.id kemudian akan dilakukan proses pembersihan dari *missing value* dan duplikasi data.

3.4. Data Integration

Pada tahapan *data integration*, Tahap awal data integration pada penelitian ini adalah adalah merubah dataset awal menjadi 26 atribut yang akan digunakan untuk tahap selanjutnya. Dataset akan dirubah menggunakan fitur PivotTable pada Microsoft Excel. Adapun tahapan *data integration* dapat dilihat pada gambar 3.



Gambar 2. Tahapan Data Integration

langkah selanjutnya yaitu menggabungkan 31 dataset menjadi satu dataset akhir yang akan digunakan untuk tahap selanjutnya.

3.5. Data Selection

Pada tahap ini merupakan tahapan pemilihan data yang relevan yang berada pada tabel, karena tidak semua data digunakan. Penghapusan atribut telah dilakukan pada tahap *data integration*. Atribut yang telah dihapus pada tahap *data integration* adalah NPSN, BP, Status, Last Sync, Jml Sync, PD, dan Rombel. Atribut-atribut tersebut tidak diperlukan dalam proses *data mining*. Adapun atribut yang digunakan pada penelitian ini antara lain.

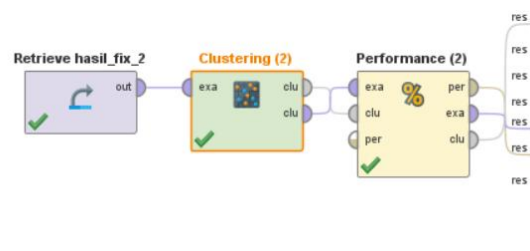
TABEL 1. ATRIBUT YANG DIGUNAKAN

No	Atribut
1	Jumlah SDN
2	Jumlah SMPN
3	Jumlah SMAN
4	Jumlah SMKN
5	Jumlah SLBN
6	Jumlah Guru SDN
7	Jumlah Guru SMPN
8	Jumlah Guru SMAN
9	Jumlah Guru SMKN
10	Jumlah Guru SLBN
11	Jumlah Kelas SDN
12	Jumlah Kelas SMPN
13	Jumlah Kelas SMAN
14	Jumlah Kelas SMKN

15	Jumlah Kelas SLBN
16	Jumlah Laboratorium SDN
17	Jumlah Laboratorium SMPN
18	Jumlah Laboratorium SMAN
19	Jumlah Laboratorium SMKN
20	Jumlah Laboratorium SLBN
21	Jumlah Perpustakaan SDN
22	Jumlah Perpustakaan SMPN
23	Jumlah Perpustakaan SMAN
24	Jumlah Perpustakaan SMKN
25	Jumlah Perpustakaan SLBN
26	Jumlah Penduduk

3.6. Data Mining

Pada tahap ini akan dilakukan proses pengolahan data menggunakan algoritma *K-Means*. Dalam menentukan jumlah *cluster* yang akan digunakan yaitu menggunakan *Elbow Method*. Kemudian akan dilakukan pengolahan data menggunakan hasil *Elbow Method* dan algoritma *K-Means* pada Rapidminer.



Gambar 3. Pemodelan Menggunakan K-Means

3.7. Evaluation

Pada tahap ini akan dilakukan proses Tahap evaluasi *clustering* menggunakan *Davies-Bouldin Index*.

3.8. Knowledge Presentation

Presentasi pengetahuan merupakan penyajian pengetahuan atau visualisasi hasil dari penelitian dalam bentuk yang mudah dimengerti. Visualisasi dapat membantu mengkomunikasikan hasil penelitian data mining kepada pihak lain dengan lebih mudah dipahami.

4. HASIL DAN PEMBAHASAN

Pada tahap pertama penelitian ini merupakan tahap *data cleaning* dengan melakukan pemeriksaan *missing value*, duplikasi data, maupun *outlier* pada 31 *dataset* awal. Pada tahap

ini tidak ditemukan adanya *missing value* maupun duplikasi data.

```
data.isnull().values.any()
False

data.duplicated().values.any()
False
```

Gambar 4. Data Cleaning

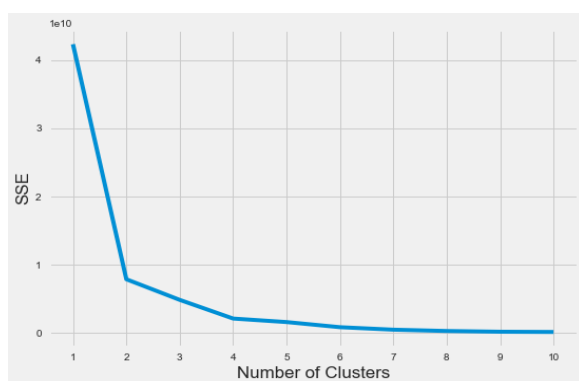
Pada gambar 4.1 kedua *output* yang dihasilkan yaitu False artinya tidak terdapat duplikasi data dan *missing value* pada *dataset*.

Tahap *data integration* merupakan tahap penggabungan seluruh *dataset* yang digunakan menjadi satu *dataset*. Tahap *data integration* pada penelitian ini terdiri dari 2 tahapan yaitu mengubah *dataset* awal menggunakan PivotTable dan penggabungan menjadi satu dataset akhir (digunakan dalam proses *data mining*).

TABEL 2. DATASET PENGGABUNGAN

No	Nama Kecamatan	Jumlah SDN	Jumlah penduduk
1	Banyusari	21	56780
...
...
30	Tirtamulya	25	50620

Dalam menentukan jumlah *cluster* yang akan digunakan yaitu menggunakan *Elbow Method*.



Gambar 5. Grafik Elbow Method

berdasarkan grafik *elbow method* tersebut dapat disimpulkan jumlah *cluster* terbaik yaitu berada di angka 2. Selanjutnya akan dilakukan proses clustering menggunakan algoritma *K-Means*. Dari proses clustering diperoleh hasil sebagai berikut.

TABEL 3. HASIL CLUSTERING

Cluster	Anggota
Cluster 1	Kec. Cikampek, Kec. Karawang Barat, Kec. Karawang Timur, Kec. Klari, Kec. Kotabaru, Kec. Rengasdengklok, Kec. Teluk Jambe Timur
Cluster 2	Kec. Banyusari, Kec. Batujaya, Kec. Ciampel, Kec. Cibuya, Kec. Cilamaya Kulon, Kec. Cilamaya Wetan, Kec. Cilebar, Kec. Jatisari, Kec. Jayakarta, Kec. Kutawaluya, Kec. Lemahabang, Kec. Majalaya, Kec. Pakisjaya, Kec. Pangkalan, Kec. Pedes, Kec. Purwasari, Kec. Rawamerta, Kec. Tegalwaru, Kec. Telagasari, Kec. Teluk Jambe Barat, Kec. Tempuran, Kec. Tirtajaya, Kec. Tirtamulya

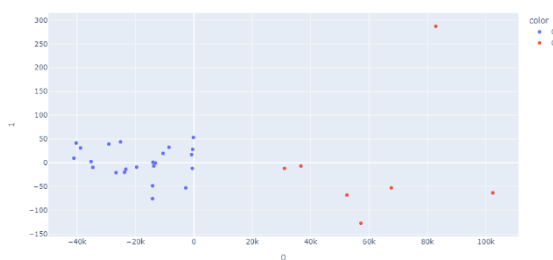
Pada tahap *evaluation* akan dilakukan evaluasi menggunakan *Davies-Bouldin Index*.

```
PerformanceVector
PerformanceVector:
Davies Bouldin: 0.408
```

Gambar 6. Hasil Evaluasi

Pada Gambar 6 menunjukkan hasil evaluasi *Davies-Bouldin Index* sebesar 0,408. Hasil tersebut menunjukkan bahwa struktur data penelitian ini kuat.

Pada tahapan *knowledge presentation* akan disajikan hasil *data mining* yang telah dilakukan berupa scatter plot seperti pada Gambar dibawah ini.



Gambar 7. Visualisasi Clustering

Berikut karakteristik *Cluster 1* yang disajikan bentuk tabel.

TABEL 4. KARAKTERISTIK CLUSTER 1

Karakteristik	Min	Max	Rata-rata	Jumlah
Jumlah SDN	24	41	31,42	220
Jumlah SMPN	2	8	4	28
Jumlah SMAN	0	4	1,85	13
Jumlah SMKN	0	2	1	7
Jumlah SLBN	0	1	0,14	1
Jumlah Guru SDN	318	540	408,28	2858
Jumlah Guru SMPN	91	329	158,85	1084
Jumlah Guru SMAN	0	181	86,71	607
Jumlah Guru SMKN	0	217	71,28	499
Jumlah Guru SLBN	0	1	0,14	1
Jumlah Kelas SDN	218	410	281	1967
Jumlah Kelas SMPN	63	254	116,14	813
Jumlah Kelas SMAN	0	113	50,42	353
Jumlah Kelas SMKN	0	132	39,14	274
Jumlah Kelas SLBN	0	4	0,57	4
Jumlah Laboratorium SDN	2	6	3,57	25
Jumlah Laboratorium SMPN	5	23	10,14	71
Jumlah	0	25	10	70

Laboratorium SMAN				
Jumlah Laboratorium SMKN	0	10	3,28	23
Jumlah Laboratorium SLBN	0	0	0	0
Jumlah Perpustakaan SDN	12	25	16,28	114
Jumlah Perpustakaan SMPN	2	7	3,57	25
Jumlah Perpustakaan SMAN	0	4	2	14
Jumlah Perpustakaan SMKN	0	2	1	7
Jumlah Perpustakaan SLBN	0	2	1	7
Jumlah Penduduk	109718	181111	140161,14	981128

Selanjutnya yaitu karakteristik *Cluster 2* adalah sebagai berikut.

TABEL 5. KARAKTERISTIK CLUSTER 2

Karakteristik	Min	Max	Rata-rata	Jumlah
Jumlah SDN	16	39	27,47	632
Jumlah SMPN	1	4	2,65	61
Jumlah SMAN	0	1	0,73	17
Jumlah SMKN	0	1	0,47	11
Jumlah SLBN	0	1	0,04	1
Jumlah Guru SDN	149	315	214,73	4939
Jumlah Guru SMPN	14	98	67,34	1549
Jumlah Guru SMAN	0	56	28,56	657
Jumlah Guru	0	48	15,86	365

SMKN				
Jumlah Guru SLBN	0	27	1,17	27
Jumlah Kelas SDN	118	249	187,3	4308
Jumlah Kelas SMPN	9	82	50,52	1162
Jumlah Kelas SMAN	0	40	17,95	413
Jumlah Kelas SMKN	0	33	7,47	172
Jumlah Kelas SLBN	0	12	0,52	12
Jumlah Laboratorium SDN	0	7	1,17	27
Jumlah Laboratorium SMPN	1	12	4,73	109
Jumlah Laboratorium SMAN	0	8	2,82	65
Jumlah Laboratorium SMKN	0	3	0,43	10
Jumlah Laboratorium SLBN	0	0	0	0
Jumlah Perpustakaan SDN	7	19	11,69	269
Jumlah Perpustakaan SMPN	0	4	2,34	54
Jumlah Perpustakaan SMAN	0	1	0,73	17
Jumlah Perpustakaan SMKN	0	2	0,30	7
Jumlah Perpustakaan SLBN	0	2	0,08	2
Jumlah Penduduk	37626	78565	59995,26	1379891

Dari Kedua tabel karakteristik tersebut dapat disimpulkan bahwa *Cluster 1* memiliki jumlah rata-rata pada atribut sekolah, guru, kelas, laboratorium, perpustakaan di semua jenjang pendidikan (SD, SMP, SMA, SMK, dan SLB) berstatus negeri dan juga penduduk yang lebih tinggi dibandingkan dengan *Cluster 2*. Sehingga dapat disimpulkan bahwa *Cluster 1* merupakan kategori tinggi dan *Cluster 2* merupakan kategori rendah.

5. Kesimpulan dan Saran

Penelitian ini menggunakan metode *elbow* dalam menentukan jumlah *cluster* optimal menghasilkan jumlah *cluster* optimal yaitu 2. Berdasarkan hasil evaluasi *clustering* yang dilakukan diperoleh nilai *Davies-Bouldin Index* (DBI) sebesar 0.408. *Cluster 1* terdiri dari 7 kecamatan yaitu Kecamatan Cikampek, Karawang Barat, Karawang Timur, Klari, Kotabaru, Rengasdengklok, Teluk Jambe Timur. *Cluster 2* terdiri dari 23 anggota kecamatan selain anggota dari *Cluster 1*. Berdasarkan analisis karakteristik hasil *clustering*, *Cluster 1* memiliki jumlah rata-rata pada atribut sekolah, guru, kelas, laboratorium, perpustakaan di semua jenjang pendidikan (SD, SMP, SMA, SMK, dan SLB) berstatus negeri dan juga penduduk yang lebih tinggi dibandingkan dengan *Cluster 2*. Sehingga dapat disimpulkan bahwa *Cluster 1* merupakan kategori tinggi dan *Cluster 2* merupakan kategori rendah. *Cluster 1* memiliki faktor penunjang yang cukup sedangkan *Cluster 2* membutuhkan perhatian lebih dari pemerintah khususnya kementerian pendidikan dalam pembangunan pendidikan di kecamatan tersebut sehingga dalam rangka terciptanya kualitas pendidikan yang merata. Saran untuk penelitian selanjutnya yaitu memvisualisasikan hasil *clustering* berupa peta dan sebagainya supaya hasil *clustering* menjadi lebih informatif dan menarik.

6. UCAPAN TERIMA KASIH

Terimakasih kepada semua yang telah mendukung proses penulisan jurnal ini khususnya bapak dan ibu dosen yang selalu membimbing penulis pada proses penulisan jurnal ini.

Daftar Pustaka:

- [1] E. Y. T. P. Dewi and I. Kamila, "PENGELOMPOKAN WILAYAH BERDASARKAN FAKTOR PENDUKUNG PENDIDIKAN DENGAN JUMLAH SEKOLAH

- DAN JUMLAH GURU MENGGUNAKAN ALGORITMA K-MEANS," *INTERVAL J. Ilm. Mat.*, vol. 2, no. 1, pp. 1–12, 2022.
- [2] S. Kartika, H. Husni, and S. Millah, "Pengaruh Kualitas Sarana Dan Prasarana Terhadap Minat Belajar Siswa Dalam Pembelajaran Pendidikan Agama Islam," *J. Penelit. Pendidik. Islam.*, vol. 7, no. 1, pp. 113–126, 2019.
- [3] A. Yoga Pratama, Y. Umaidah, and A. Voutama, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 897–910, 2021.
- [4] L. Mutawalli, M. T. A. Zaen, and W. Bagye, "KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto)," *J. Inform. dan Rekayasa Elektron.*, vol. 2, no. 2, p. 43, 2019, doi: 10.36595/jire.v2i2.117.
- [5] M. Hariyanto and R. T. Shita, "Clustering pada Data Mining untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance," *Sist. Komput. dan Tek. Inform.*, vol. 1, no. 1, pp. 117–122, 2018.
- [6] W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia," *Ilk. J. Ilm.*, vol. 13, no. 1, pp. 31–35, 2021, doi: 10.33096/ilkom.v13i1.763.31-35.
- [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)," *Indones. J. Comput. Model.*, vol. 1, no. 1, pp. 18–27, 2018, doi: 10.24246/j.icm.2018.v1.i1.p18-27.
- [8] F. Nur, M. Zarlis, and B. B. Nasution, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolah menengah Kejuruan Untuk Clustering Jurusan," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 2, pp. 100–105, 2017, doi: 10.30743/infotekjar.v1i2.70.
- [9] Euis Saraswati, Yuyun Umaidah, and Apriade Voutama, "Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19," *Gener. J.*, vol. 5, no. 2, pp. 109–118, 2021, doi: 10.29407/gj.v5i2.16125.
- [10] D. L. Rianti, Y. Umaidah, and A. Voutama, "Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 6, no. 1, p. 98, 2021, doi: 10.30998/string.v6i1.9993.
- [11] M. Wahyudi and S. Pujiastuti, Lise, "Penerapan Data Mining Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan," *Pros. Semin. Nas. Ris. dan Inform. Sci.*, vol. 2, pp. 432–440, 2020.
- [12] M. Riyyan and H. Firdaus, "PERBANDINGAN ALGORITME NAÏVE BAYES DAN KNN TERHADAP DATA PENERIMAAN BEASISWA (Studi Kasus Lembaga Beasiswa Baznas Jabar)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 1, pp. 1–10, 2022.
- [13] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [14] A. F. Khairati, A. . Adlina, G. . Hertono, and B. . Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," *Prism. Pros. Semin. Nas. Mat.*, vol. 2, pp. 161–170, 2019, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/28906>.