

## PENGELOMPOKAN WILAYAH BERDASARKAN FAKTOR PENDUKUNG PENDIDIKAN DENGAN JUMLAH SEKOLAH DAN JUMLAH GURU MENGGUNAKAN ALGORITMA K-MEANS

Elgi Yanty Tri Putri Dewi\*, Isti Kamila  
Program Studi Matematika, Universitas Pakuan  
e-mail: elgiyanty31@gmail.com

Diterima: 31 Januari 2022, disetujui: 1 Maret 2022, dipublikasi: 22 April 2022

**Abstract:** Education is one field that has a role in creating human resources and national development. The national education system states that the government can ensure equal distribution of education, quality improvement, the relevance and efficiency of education management. The implementation of education in regions of Indonesia has obstacles, including in remote areas have not been evenly distributed in accessing education services. The provision of schools and teaching staff is a supporting factors for teaching and learning activities. All regions, including sub-districts in Cianjur Regency. One of the efforts for equity is to use the k-means method to classify sub-districts in Cianjur Regency in the provision of schools and teachers by the Government. The data used number of schools and teachers at the elementary school, junior high school, high school and vocation high school level in every sub-district in Cianjur Regerency in 2020. The results that have been researched show two clusters. Cluster 1 or an area that has several schools and the number of teachers lacked in Cianjur Regency consisting of 21 sub-districts and cluster 2 or an area had a better number of schools and the number of teachers than cluster 1 in Cianjur Regency consisting of 11 sub-districts.

**Keywords:** Classification, Education, K-means Clustering

**Abstrak:** Pendidikan merupakan salah satu bidang yang memiliki peran dalam menciptakan sumber daya manusia serta pembangunan suatu bangsa. Sistem pendidikan nasional menyatakan bahwa pemerintah harus mampu menjamin pemerataan pendidikan, peningkatan mutu serta relevansi dan efisiensi manajemen pendidikan. Pelaksanaan pendidikan di daerah-daerah Indonesia memiliki berbagai kendala, diantaranya pada daerah yang jauh dari perkotaan belum terdistribusi secara merata dalam mengakses layanan pendidikan. Penyediaan sekolah dan tenaga pendidik merupakan salah satu faktor pendukung kegiatan belajar mengajar, terutama pada daerah terpencil sehingga pemerataan pendidikan dapat dirasakan oleh semua daerah termasuk kecamatan-kecamatan yang ada di Kabupaten Cianjur. Salah satu upaya untuk adanya pemerataan yaitu dengan menggunakan metode k-means untuk mengelompokkan kecamatan di Kabupaten Cianjur dalam penyediaan sekolah dan guru oleh Pemerintah. Data yang digunakan yaitu jumlah sekolah dan jumlah guru pada tingkat sekolah dasar, sekolah menengah pertama, sekolah menengah atas dan sekolah menengah kejuruan setiap kecamatan di Kabupaten Cianjur pada Tahun 2020. Hasil penelitian menunjukkan bahwa terdapat dua cluster yaitu cluster 1 atau wilayah yang dapat diprioritaskan dalam penyediaan sekolah dan guru terdiri dari 21 kecamatan dan cluster 2 atau wilayah yang memiliki jumlah sekolah dan jumlah guru lebih baik dari cluster 1 di Kabupaten Cianjur terdiri dari 11 kecamatan.

**Kata Kunci:** Klasifikasi, Pendidikan, K-means Clustering

## PENDAHULUAN

Pendidikan merupakan salah satu aktivitas untuk meningkatkan kesejahteraan manusia. Pendidikan dipandang sebagai bagian dari usaha manusia untuk meningkatkan kesejahteraan atau sebagai bagian dari pembangunan nasional. Penguatan dari pandangan tersebut yaitu dengan adanya pendapat bahwa pendidikan mempunyai peranan dalam pembangunan nasional dan pembangunan ekonomi khususnya [1].

Pendidikan memiliki peran penting dalam menciptakan sumber daya manusia pada suatu bangsa. Manusia mendapatkan berbagai pengetahuan, ketrampilan, keahlian dan macam-macam tatanan hidup baik yang berupa norma-norma, aturan-aturan positif, dan sebagainya. Semakin tinggi pendidikan maka orang tersebut memiliki kesempatan untuk mendapatkan pekerjaan yang lebih baik dengan pendapatan yang lebih besar [2].

Pendidikan nasional memiliki fungsi dalam mengembangkan kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa dan pendidikan nasional juga memiliki tujuan mengembangkan potensi peserta didik agar menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, Berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri, dan menjadi warga negara yang demokratis serta bertanggung jawab [3]

Undang-undang nomor 20 tahun 2003 mengenai sistem pendidikan nasional menyatakan bahwa pemerintah harus mampu menjamin pemerataan kesempatan pendidikan, peningkatan mutu serta relevansi dan efisiensi manajemen pendidikan untuk menghadapi tantangan sesuai dengan tuntutan perubahan kehidupan lokal, nasional dan global sehingga perlu dilakukan pembaharuan pendidikan secara terencana, terarah dan berkesinambungan.

Pemerataan pendidikan yang dilaksanakan di berbagai daerah Indonesia mempunyai bermacam-macam kendala dalam pelaksanaannya. Permasalahan pendidikan diantaranya disebabkan oleh daerah pedesaan yang terpencil dan jauh dari perkotaan dalam mengakses layanan pendidikan yang masih belum terdistribusi secara merata [4]. Dalam penyelenggaraan pendidikan, penyediaan sekolah dan tenaga pendidik merupakan salah satu faktor pendukung dalam kegiatan belajar mengajar yang efektif dan efisien [5]. Faktor pendukung harus dapat dirasakan oleh semua daerah terutama pada daerah terpencil sehingga adanya pemerataan pendidikan termasuk pada wilayah-wilayah di Kabupaten Cianjur. Pengelompokan wilayah Cianjur perlu dilakukan agar pemerintah Kabupaten Cianjur dapat memperhatikan *cluster* wilayah mana yang membutuhkan perbaikan Pendidikan.

Salah satu cara untuk menemukan pola adalah data mining. Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan di masa depan [6]. Data mining ini juga dikenal dengan istilah *pattern recognition* [7]. *Pattern recognition* atau disebut dengan pengenalan pola merupakan suatu proses pencarian pola atau kesamaan data ke dalam beberapa kategori.

Salah satu metode dalam pengenalan pola yaitu analisis klaster. Analisis klaster merupakan penggabungan objek atau data yang berdasarkan kesamaan karakteristiknya. Analisis klaster terdiri dari *hierarchical clustering* dan *non hierarchical clustering*. *Hierarchical clustering* adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk semacam pohon dimana ada hierarki

(tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster* [8]. Berbeda dengan metode *hierarchical clustering*, metode *non hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan (dua *cluster*, tiga *cluster*, atau lain sebagainya). Setelah jumlah *cluster* diketahui, baru proses *cluster* dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *k-means clustering* atau algoritma *k-means* [9].

Penerapan algoritma *k-means* telah diterapkan dalam berbagai bidang. *K-means* juga dapat diterapkan pada bidang pendidikan seperti penelitian yang bertujuan untuk pengelompokan tingkat putus sekolah pada jenjang SMP di Indonesia [10]. Penelitian lain pada penerapan *k-means* yang memiliki manfaat pada bidang pendidikan yaitu pengelompokan data guru untuk menganalisis persebaran guru di Indonesia [11].

*K-means* merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi sehingga walaupun mengelompokkan data yang cukup besar, metode ini menghasilkan kelompok dengan waktu yang relatif cepat dan efisien [12]. Algoritma ini merupakan implementasi dari algoritma *clustering* partisional yang sederhana sehingga populer dan sering digunakan [13]. *K-means* berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain [14]. *K-means* dapat digunakan untuk mengelompokkan wilayah agar pemerintah dapat memperhatikan wilayah-wilayah sesuai dengan *cluster* untuk penyediaan sekolah dan guru di wilayah-wilayah Kabupaten Cianjur.

## METODOLOGI PENELITIAN

### Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder dari Badan Pusat Statistik Kabupaten Cianjur meliputi data jumlah sekolah dan jumlah guru pada tingkat sekolah dasar, sekolah menengah pertama, sekolah menengah atas dan sekolah menengah kejuruan setiap kecamatan di Kabupaten Cianjur pada Tahun 2020 [15]. Metode yang digunakan yaitu algoritma *k-means*.

### Tahapan Algoritma *K-means*

Algoritma *K-means* sangat terkenal karena kemudahan dan kemampuannya untuk mengelompokkan data besar dan *outlier* dengan sangat cepat. *K-means* ini merupakan salah satu *non hierarchical clustering* dengan tujuan untuk mengelompokkan objek ke dalam suatu *cluster* yang berdasarkan pada karakteristiknya, atau objek memiliki kesamaan karakteristik yang maksimum dengan objek dalam *cluster* dan objek memiliki kesamaan karakteristik yang minimum dengan objek di *cluster* lain [16]. Dalam metode *k-means* setiap data harus termasuk ke *cluster* tertentu pada suatu tahapan proses, pada tahapan proses berikutnya dapat berpindah ke *cluster* yang lain.

Berikut ini adalah langkah-langkah dalam algoritma *k-means*.

1. Melakukan standarisasi data. Rumus standarisasi data sebagai berikut.

$$Z = \frac{x - \bar{x}}{s} \quad (1)$$

Keterangan:

$x$  = Nilai dari data

$\bar{x}$  = Rata-rata nilai dari data

$S$  = Simpangan baku

2. Tentukan  $k$  sebagai jumlah *cluster* yang di bentuk. Untuk menentukan banyaknya *cluster*  $k$  dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak *cluster*.
3. Tentukan  $k$  *centroid* (titik pusat *cluster*) awal secara random. Penentuan *centroid* awal dilakukan secara random/acak dari objek-objek yang tersedia sebanyak  $k$  *cluster*, kemudian *centroid cluster* ke- $i$  berikutnya dihitung dengan menggunakan rumus sebagai berikut.

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}; i, j, k = 1, 2, 3, \dots n \quad (2)$$

Keterangan:

$V_{ij}$  = *Centroid* rata-rata *cluster* ke- $i$  untuk variabel ke- $j$

$N_t$  = Jumlah anggota *cluster* ke- $i$

$i, k$  = Indeks dari *cluster*

$j$  = Indeks dari variabel

$X_{kj}$  = Nilai data ke- $k$  variabel ke- $j$  dalam *cluster* tersebut

4. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *cluster*. Untuk menghitung jarak antara objek dengan *centroid* penulis menggunakan *Euclidian Distance*.

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (3)$$

Keterangan:

$De$  = *Euclidian Distance*

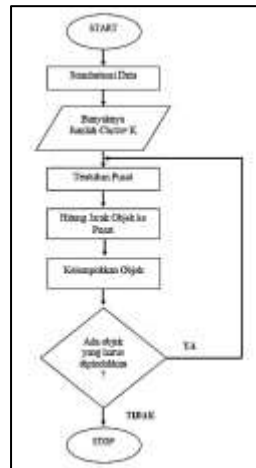
$I$  = Banyaknya objek

$(x, y)$  = Koordinat objek

$(s, t)$  = Koordinat *centroid*

5. Alokasikan masing-masing objek ke dalam masing-masing
6. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan rumus yang ada di langkah 2.
7. Ulangi langkah 3 jika posisi *centroid* baru tidak sama. Pengecekan konvergensi dilakukan dengan membandingkan matriks *group assignment* pada iterasi sebelumnya dengan matriks *group assignment* pada iterasi yang sedang berjalan.

Berikut ini merupakan diagram *flowchart* dari algoritma *k-means* dan menggambarkan langkah-langkah dalam algoritma *k-means* pada Gambar 1 berikut ini:



Gambar 1. *Flowchart* Algoritma *K-Means*

## HASIL DAN PEMBAHASAN

Pengelompokan wilayah-wilayah di Kabupaten Cianjur dengan analisis algoritma *k-means* sebagai berikut:

1. Menentukan jumlah *cluster*

Dalam penelitian ini data-data yang ada akan dikelompokkan menjadi 2 *cluster*. *Cluster 1* yaitu wilayah yang diprioritaskan dalam penyediaan sekolah dan guru untuk pemerataan pendidikan, sedangkan *cluster 2* yaitu wilayah yang memiliki jumlah sekolah dan guru yang lebih baik dari *cluster 1* dan dapat diprioritaskan setelah *cluster 1* dalam penyediaan sekolah dan guru.

2. Melakukan standarisasi data dengan rumus (1), sehingga diperoleh hasil sebagai pada Tabel 1.

Tabel 1. Hasil Standarisasi Data

No	Kecamatan	Jumlah Sekolah SD (Negeri)	Jumlah Sekolah SMP (Negeri)	Jumlah Sekolah SMA (Negeri)	Jumlah Sekolah SMK (Negeri)	Jumlah Guru SD (Negeri)	Jumlah Guru SMP (Negeri)	Jumlah Guru SMA (Negeri)	Jumlah Guru SMK (Negeri)
1	Agrabinta	-0.6054	-0.29827	-0.90884	0.18942	-0.90096	-1.16668	-0.69369	-0.16362
2	Leles	-0.87323	-0.29827	-0.90884	1.70477	-0.81603	-0.95032	-0.69369	0.35736
3	Sindangbarang	-0.2483	1.29252	0.70687	0.18942	-0.45017	0.17955	0.30773	-0.45016
4	Cidaun	0.28736	2.08791	-0.90884	0.18942	-0.11698	0.92478	-0.69369	0.04477
5	Naringgul	-0.15902	2.08791	-0.90884	0.18942	-0.5351	0.41994	-0.69369	-0.31992
6	Cibinong	0.82301	2.48561	2.32259	0.18942	0.14434	0.68438	0.34111	-0.45016
7	Cikadu	-0.51613	0.49712	-0.90884	0.18942	-0.90096	-0.61376	-0.69369	-0.37202
8	Tanggeung	-0.51613	-0.29827	-0.90884	0.18942	-0.78336	-0.83012	-0.69369	0.40946
9	Pasirkuda	-1.23034	0.09942	0.70687	0.18942	-1.10349	-0.92628	-0.29312	-0.29387
10	Kadupandak	0.19808	0.49712	0.70687	-1.32593	-0.40444	-0.39741	0.20758	-0.91905
11	Cijati	-0.69468	-0.69597	-0.90884	0.18942	-0.85523	-1.33495	-0.69369	-0.16362
12	Takokak	-0.42685	0.89482	-0.90884	0.18942	-0.84216	0.22763	-0.69369	-0.21572
13	Sukanagara	-0.6054	0.89482	0.70687	0.18942	-0.6919	0.01127	0.67491	-0.39807
14	Pagelaran	0.91229	0.49712	0.70687	0.18942	-0.02552	0.08339	-0.02608	0.38341
15	Campaka	0.46591	0.49712	-0.90884	0.18942	-0.24111	0.01127	-0.69369	-0.05943
16	Campakamulya	-1.76599	-1.09367	-0.90884	-1.32593	-1.58041	-1.45515	-0.69369	-0.91905
17	Cibeber	2.16216	0.49712	0.70687	-1.32593	1.67963	1.71809	0.77505	-0.91905
18	Warungkondang	-0.87323	-0.69597	-0.90884	-1.32593	-0.11698	0.22763	-0.69369	-0.91905
19	Gekbrong	-1.23034	-1.49137	0.70687	-1.32593	-0.8095	-1.09456	0.24096	-0.91905
20	Cilaku	1.44795	-0.69597	0.70687	3.22013	1.39217	0.05935	1.34252	4.18662
21	Sukaluyu	-0.51613	-1.09367	-0.90884	0.18942	0.05941	-0.758	-0.69369	0.14897
22	Bojongpicung	0.55519	-0.69597	0.70687	0.18942	0.92178	0.29974	0.04068	0.9565
23	Haurwangi	-1.40889	-0.69597	-0.90884	0.18942	-0.58084	-0.06085	-0.69369	-0.08547
24	Ciranjang	-0.51613	-1.09367	0.70687	-1.32593	0.19007	-0.758	1.14223	-0.91905
25	Mande	0.55519	-0.69597	0.70687	-1.32593	0.26847	0.32378	0.70829	-0.91905
26	Karangtengah	1.53722	-0.29827	0.70687	0.18942	1.79069	1.7902	-0.29312	0.20107
27	Cianjur	1.53722	0.09942	2.32259	0.18942	3.07772	3.25662	3.9128	1.37329
28	Cugenang	1.2694	-0.69597	-0.90884	0.18942	0.81725	0.41994	-0.69369	-0.29387
29	Pacet	-1.05179	-1.09367	0.70687	1.70477	-0.41751	-1.16668	0.37449	1.13884
30	Cipanas	-0.06975	-0.69597	-0.90884	0.18942	0.75846	0.34782	-0.69369	0.8002
31	Sukaresmi	0.10881	-0.29827	0.70687	-1.32593	0.13781	0.22763	1.64294	-0.91905
31	Cikalongkulon	1.44795	0.49712	-0.90884	0.18942	0.93485	0.29974	-0.69369	0.61786

Dari Tabel 1 terjadi transformasi data yang memiliki satuan berbeda menjadi satuan yang sama untuk menghindari perhitungan yang bias atau tidak valid.

- Menentukan titik pusat awal (*centroid*) dari setiap *cluster*. Dalam penelitian ini titik pusat awal ditentukan secara acak dan didapat titik pusat dari setiap *cluster* dapat dilihat pada Tabel 2.

Tabel 2. Titik Pusat Awal setiap *Cluster*

Titik Pusat Cluster	Kecamatan	Jumlah Sekolah SD (Negeri)	Jumlah Sekolah SMP (Negeri)	Jumlah Sekolah SMA (Negeri)	Jumlah Sekolah SMK (Negeri)	Jumlah Guru SD (Negeri)	Jumlah Guru SMP (Negeri)	Jumlah Guru SMA (Negeri)	Jumlah Guru SMK (Negeri)
1	Warungkondang	-0.87323	-0.69597	-0.90884	-1.32593	-0.11698	0.22763	-0.69369	-0.91905
2	Bojongpicung	0.55519	-0.69597	0.70687	0.18942	0.92178	0.29974	0.04068	0.9565

- Menghitung jarak setiap data yang ada terhadap setiap *centroid*. Jarak setiap data yang ada terhadap *centroid* dapat dihitung menggunakan rumus (3) yaitu

### Euclidean Distance.

Jarak data wilayah pertama ke *centroid* pertama:

$$D(1,1) = \sqrt{(-0,6054 - (-0,87323))^2 + (-0,29827 - (-0,69597))^2 + (-0,90884 - (-0,90884))^2 + (0,18942 - (-1,32593))^2 + (-0,90096 - (-0,11698))^2 + (-1,16668 - (0,22763))^2 + (-0,69369 - (-0,69369))^2 + (-0,16362 - (-0,91905))^2}$$

$$= 2,3781$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data wilayah pertama dengan *centroid* pertama adalah 2,3781.

Jarak data wilayah pertama ke *centroid* kedua:

$$D(1,2) = \sqrt{(-0,6054 - (0,55519))^2 + (-0,29827 - (-0,69597))^2 + (-0,90884 - (0,70687))^2 + (0,18942 - (0,18942))^2 + (-0,90096 - (0,92178))^2 + (-1,16668 - (0,29974))^2 + (-0,69369 - (0,04068))^2 + (-0,16362 - (-0,87323))^2}$$

$$= 3,3738$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data wilayah pertama dengan *centroid* kedua adalah 3,3738.

Berdasarkan hasil perhitungan di atas dapat disimpulkan bahwa data pertama memiliki jarak paling dekat dengan *centroid* kesatu, sehingga data pertama dimasukkan ke dalam *cluster* 1.

Jarak data wilayah kedua ke *centroid* pertama:

$$D(2,1) = \sqrt{(-0,87323 - (-0,87323))^2 + (-0,29827 - (-0,69597))^2 + (-0,90884 - (-0,90884))^2 + (1,70477 - (-1,32593))^2 + (-0,81603 - (-0,11698))^2 + (-0,95032 - (0,22763))^2 + (-0,69369 - (-0,69369))^2 + (0,35736 - (-0,91905))^2}$$

$$= 3,5845$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data wilayah kedua dengan *centroid* pertama adalah 3,5845.

Jarak data wilayah kedua ke *centroid* kedua:

$$D(2,2) = \sqrt{(-0,87323 - (0,55519))^2 + (-0,29827 - (-0,69597))^2 + (-0,90884 - (0,70687))^2 + (0,170477 - (0,18942))^2 + (-0,81603 - (0,92178))^2 + (-0,95032 - (0,29974))^2 + (-0,69369 - (0,04068))^2 + (0,35736 - (-0,87323))^2}$$

$$= 3,35477$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data wilayah kedua dengan *centroid* pertama adalah 3,35477.

Berdasarkan hasil perhitungan di atas dapat disimpulkan bahwa data kedua memiliki jarak paling dekat dengan *centroid* kedua, sehingga data kedua



dimasukkan ke dalam *cluster* 2. Hasil perhitungan sampai diperoleh jarak data ke-32.

5. Menempatkan data ke masing-masing *centroid* terdekat. Setelah menghitung jarak setiap data ke *centorid* 1 dan 2, maka data dikelompokkan pada masing-masing *centroid* dan diperoleh hasil seperti pada Tabel 3.

Tabel 3. Hasil Perhitungan Jarak Masing-Masing Data ke Setiap *Centroid* pada Iterasi Pertama.

Data ke-i	Jarak ke		Cluster
	Centroid 1	Centroid 2	
1	2,378147	3,37378	1
2	3,584518	3,547711	2
3	3,253838	2,923421	2
4	3,578856	3,643143	1
5	3,335736	3,895651	1
6	5,225484	3,952979	2
7	2,338605	3,41216	1
8	2,43101	2,989824	1
9	2,919417	3,331734	1
10	2,543085	3,104464	1
11	2,426007	3,434623	1
12	2,459006	3,337007	1
13	3,167172	2,966794	2
14	3,419009	1,681683	2
15	2,51281	2,654975	1
16	2,434864	4,88369	1
17	4,566624	3,597936	2
18	0	3,476837	1
19	2,544022	3,823123	1
20	7,822079	4,73178	2
21	2,173757	2,640135	1
22	3,476837	0	2
23	1,891153	3,237441	1
24	2,707869	3,160299	1
25	2,602742	2,585937	2
26	4,286163	2,186577	2
27	8,06454	5,724144	2
28	2,861428	2,291007	2
29	4,404095	3,02086	2
30	2,584285	1,895728	2
31	3,042671	3,059157	1
32	3,547008	2,34204	2



6. Melakukan iterasi kembali dengan memperbaharui nilai *centroid*. Nilai *centroid* baru dapat dihitung menggunakan rumus (2) dan diperoleh hasil seperti pada Tabel 4.

Tabel 4. Titik Pusat Baru setiap *Cluster*

Titik Pusat Cluster baru	Jumlah Sekolah SD (Negeri)	Jumlah Sekolah SMP (Negeri)	Jumlah Sekolah SMA (Negeri)	Jumlah Sekolah SMK (Negeri)	Jumlah Guru SD (Negeri)	Jumlah Guru SMP (Negeri)	Jumlah Guru SMA (Negeri)	Jumlah Guru SMK (Negeri)
1	-0,55289	-0,06433	-0,43363	-0,34541	-0,55201	-0,43276	-0,31669	-0,38734
2	0,62661	0,07291	0,49144	0,39147	0,62562	0,49046	0,35891	0,43899

7. Selanjutnya, hitung jarak setiap data yang ada terhadap setiap pusat *cluster* (*centroid*) baru pada iterasi kedua dengan rumus (3) yaitu *Euclidean Distance*. Adapun hasil *clustering* dalam bentuk matriks *group assignment* dengan iterasi sebanyak 5 kali dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Clustering* dalam Bentuk Matriks *Group Assignment*.

Data ke-i	Sebelum Iterasi		Iterasi ke-1		Iterasi ke-2		Iterasi ke-3		Iterasi ke-4		Iterasi ke-5	
	Centroid 1	Centroid 2	Centroid 1	Centroid 2	Centroid 1	Centroid 2	Centroid 1	Centroid 2	Centroid 1	Centroid 2	Centroid 1	Centroid 2
1	0	0	1	0	1	0	1	0	1	0	1	0
2	0	0	0	1	1	0	1	0	1	0	1	0
3	0	0	0	1	1	0	1	0	1	0	1	0
4	0	0	1	0	0	1	0	1	1	0	1	0
5	0	0	1	0	1	0	1	0	1	0	1	0
6	0	0	0	1	0	1	0	1	0	1	0	1
7	0	0	1	0	1	0	1	0	1	0	1	0
8	0	0	1	0	1	0	1	0	1	0	1	0
9	0	0	1	0	1	0	1	0	1	0	1	0
10	0	0	1	0	1	0	1	0	1	0	1	0
11	0	0	1	0	1	0	1	0	1	0	1	0
12	0	0	1	0	1	0	1	0	1	0	1	0
13	0	0	0	1	1	0	1	0	1	0	1	0
14	0	0	0	1	0	1	0	1	0	1	0	1
15	0	0	1	0	1	0	1	0	1	0	1	0
16	0	0	1	0	1	0	1	0	1	0	1	0
17	0	0	0	1	0	1	0	1	0	1	0	1
18	0	0	1	0	1	0	1	0	1	0	1	0
19	0	0	1	0	1	0	1	0	1	0	1	0
20	0	0	0	1	0	1	0	1	0	1	0	1
21	0	0	1	0	1	0	1	0	1	0	1	0
22	0	0	0	1	0	1	0	1	0	1	0	1
23	0	0	1	0	1	0	1	0	1	0	1	0
24	0	0	1	0	1	0	1	0	1	0	1	0
25	0	0	0	1	0	1	0	1	0	1	0	1
26	0	0	0	1	0	1	0	1	0	1	0	1
27	0	0	0	1	0	1	0	1	0	1	0	1
28	0	0	0	1	0	1	0	1	0	1	0	1
29	0	0	0	1	1	0	1	0	1	0	1	0
30	0	0	0	1	0	1	1	0	1	0	1	0
31	0	0	1	0	0	1	0	1	0	1	0	1
32	0	0	0	1	0	1	0	1	0	1	0	1

Pada analisis algoritma *k-means* ini iterasi *clustering* sebanyak 5 kali atau pada iterasi ke 5 memiliki posisi *centroid* sama dengan iterasi sebelumnya sehingga iterasi dihentikan.

#### 8. Interpretasi Hasil *Cluster*

Anggota *cluster* yang terbentuk dari algoritma *k-means* terdiri dari:

- Cluster* 1 terdiri 21 kecamatan di Kabupaten Cianjur diantaranya Kecamatan Agrabinta, Leles, Sindangbarang, Cidaun, Naringgul, Cikadu, Tanggeung, Pasirkuda, Kadupandak, Cijati, Takokak, Sukanagara,

Campaka, Campakamulya, Warungkondang, Gekbrong, Sukaluyu, Haurwangi, Ciranjang, Pacet dan Cipanas merupakan wilayah yang memiliki jumlah sekolah dan jumlah guru yang masih kurang.

- b. *Cluster 2* terdiri dari 11 kecamatan di Kabupaten Cianjur diantaranya Kecamatan Cibinong, Pagelaran, Cibeber, Cilaku, Bojongpicung, Mande, Karangtengah, Cianjur, Cugenang, Sukaresmi dan Cikalongkulon. Wilayah yang ada pada *cluster 2* merupakan wilayah yang memiliki jumlah sekolah dan jumlah guru lebih baik dari *cluster 1* dan dapat diprioritaskan setelah *cluster 1*.

## KESIMPULAN

Berdasarkan hasil analisis algoritma *k-means* terhadap data jumlah sekolah dan jumlah guru di Kabupaten Cianjur yaitu beberapa kecamatan yang dapat diprioritaskan dalam penyediaan sekolah dan guru untuk pemerataan pendidikan terlebih dahulu yaitu pada *cluster 1* yang terdiri Kecamatan Agrabinta, Leles, Sindangbarang, Cidaun, Naringgul, Cikadu, Tanggeung, Pasirkuda, Kadupandak, Cijati, Takokak, Sukanagara, Campaka, Campakamulya, Warungkondang, Gekbrong, Sukaluyu, Haurwangi, Ciranjang, Pacet dan Cipanas. Pemerintah Kabupaten Cianjur, diharapkan lebih memperhatikan kembali 21 kecamatan yang terdapat pada *cluster 1*. Sehingga masyarakat mendapatkan pemerataan dalam bidang pendidikan melalui penyediaan sekolah dan guru. Penelitian dapat dikembangkan kembali dengan menambahkan data pendukung lainnya.

## DAFTAR PUSTAKA

- [1] Hakim, L. (2016). Pemerataan Akses Pendidikan Bagi Rakyat Sesuai dengan Amanat Undang-Undang Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional. *Jurnal EduTech*. **2** (1): 63.
- [2] Widyastuti, A. (2012). Analisis Hubungan Antara Produktivitas Pekerja dan Tingkat Pendidikan Pekerja Terhadap Kesejahteraan Keluarga di Jawa Tengah Tahun 2009. *Economics Development Analysis Journal*. **1** (2): 2.
- [3] Sujana, I. (2019). Fungsi dan Tujuan Pendidikan Indonesia. *Jurnal Pendidikan Dasar*. **4** (1): 30.
- [4] Aristo, T. (2019). Analisis Permasalahan Pemerataan Pendidikan Di Kabupaten Sintang. *Jurnal Akuntabilitas Manajemen Pendidikan*. **7** (1): 26
- [5] Indah, P. (2019). Peran Pendidikan Dalam Implementasi Media Pembelajaran Terhadap Peserta Didik Generasi 4.0. Prosiding Seminar Nasional Pendidikan FKIP Universitas Sultan Ageng Tirtayasa. **2** (1): 511.
- [6] Handoko, S., Fauziah, A.T.E. Handayani. (2020). Implementasi Data Mining untuk menentukan Tingkat Penjualan Paket Data Telkomsel menggunakan Metode K-means Clustering. *Jurnal Ilmiah Teknologi dan Rekayasa*. **25**(1): 76-88.
- [7] Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.

- [8] Santoso, S. (2010). *Statistik Multivariat*. Jakarta: Elex Media Komputindo.
- [9] Sibuea, F.L., A. Sapta. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*. **4** (1): 85-92. <https://doi.org/10.33330/jurteks.v4i1.28>.
- [10] Wuni, S., A.M. Siregar, D.S. Kusumaningrum. (2020). K-Means Clustering untuk mengelompokan Tingkat Putus Sekolah Jenjang SMP di Indonesia. *Scientific Student Journal for Information*. **1**(1): 47-54.
- [11] Idris, F., F. Azmi, P.D. Kusuma. (2019). Pengelompokan Data Guru di Indonesia menggunakan K-means Clustering. *E-Proceeding of Engineering*. **6**(2): 5648-5653.
- [12] Putra, I.M.A.W., I. Gede., K.Y.E. Aryanto. Sistem Rekomendasi berdasarkan Data Transaksi Perpustakaan Daerah Tabanan dengan menggunakan K-means Clustering. *Jurnal Ilmu Komputer Indonesia*. **3**(1): 18-22.
- [13] Sonang, S., A.T. Purba, F.O.I. Pardede. (2019). Pengelompokan Jumlah Penduduk berdasarkan Kategori Usia dengan Metode K-Means. *Jurnal Teknik dan Informasi*. **2**(2): 166- 172.
- [14] Sari, A.I., A.S. Tambunan, W. Saputra, I. Sudahri. (2020). Implementasi Algoritma K-means dalam Pengelompokan Penyandang Disabilitas menurut Kecamatan Kabupaten Simalungun. *Prosiding Seminar Nasional Riset dan Indormasi Science*. **2**(1): 54-61.
- [15] Badan Pusat Statistik Kabupaten Cianjur. (2021). Tentang BPS. <https://cianjurkab.bps.go.id/menu/1/tentang-bps.html> [diakses 2 Februari 2021]
- [16] Asroni, H. Fitri, E. Prasetyo. (2018). Penerapan Metode Clustering dengan Algoritma K-means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta. *Semesta Teknika*. **21**(1): 60-64.