

IMPLEMENTASI ALGORITMA K-MEANS DAN K-MEDOIDS DALAM PENGELOMPOKAN NILAI UJIAN NASIONAL TINGKAT SMK

Ninda Nurul Rhamadani^{1*)}, Ahmad Fauzi¹⁾, Euis Nurlaelasari¹⁾, Adi Rizky Pratama¹⁾

¹⁾ Teknik Informatika, Universitas Buana Perjuangan Karawang, Karawang

*Email Korespondensi : if16.nindarhamadani@mhs.ubpkarawang.ac.id

ABSTRAK

Pusat Penilaian Pendidikan melakukan evaluasi standar pendidikan setiap tahunnya dengan menyelenggarakan Ujian Nasional pada jenjang SD, SMP, SMA/K dan sederajat. Data hasil nilai UN dapat dilihat melalui website resmi Kementerian Pendidikan dan Kebudayaan (KEMDIKBUD) akan tetapi data yang disajikan tidak memiliki peringkat tertentu atau kelompok berdasarkan kategori nilai UN sehingga masyarakat tidak memiliki acuan saat mendaftarkan putra-putri mereka untuk menempuh pendidikan selanjutnya. Pada penelitian ini akan melakukan pengelompokan sekolah menjadi 3 kategori yaitu baik, sedang dan cukup untuk memudahkan peserta didik mendapatkan informasi kategori sekolah mereka dengan menerapkan algoritma K-Means dan K-Medoid. Hasil pengelompokan menggunakan algoritma k-means menghasilkan kluster baik sebanyak 14 anggota, kluster sedang 46 anggota dan kluster cukup 49 anggota. Lalu, pada algoritma k-medoids diperoleh hasil kluster baik 27 anggota, kluster sedang 43 anggota dan kluster cukup 39 anggota. Perbedaan jumlah kluster pada kinerja tiap algoritma memiliki pola perhitungan yang berbeda pada masing-masing iterasi tergantung pada dataset yang digunakan serta titik centroid yang dijadikan perhitungan pada algoritma.

Kata kunci: Dataset, K-Means, K-Medoids, Pengelompokan, Titik Centroid

ABSTRACT

The Education Assessment Center conducts an annual evaluation of educational standards by holding National Examinations (NE) from elementary until higher levels of education before college. Data on national examination scores can be seen through the official website of the Ministry of Education and Culture but the data presented did not have a certain rank or group based on the NE score category so that the community has no reference when registering their children for further education. In this study, schools grouped into 3 categories, namely good, moderate, and sufficient to make it easier for students to get information on their school category by applying the K-Means and K-Medoid algorithms. The results of grouping using the k-means algorithm produced good clusters of 14 members, medium clusters of 46 members, and sufficient clusters of 49 members. Then, in the k-medoids algorithm, the results of the cluster were 27 members, the cluster was 43 members, and the cluster was 39 members. The difference in the number of clusters in the performance of each algorithm had a different calculation pattern for each iteration depending on the dataset used and the centroid points that were used as calculations in the algorithm.

Keywords: Dataset, K-Means, K-Medoids, Cluster, Centroid

PENDAHULUAN

Ujian Nasional adalah ujian yang diselenggarakan secara Nasional pada tingkat akhir Sekolah Dasar, Sekolah Menengah Pertama dan Sekolah Menengah Atas [1]. Ujian Nasional diselenggarakan untuk mengukur pencapaian kompetensi lulusan peserta didik pada jenjang satuan pendidikan dasar dan pendidikan menengah sebagai hasil dari proses pembelajaran sesuai dengan Standar Kompetensi Lulusan (SKL). Selain itu, salah satu kegunaan hasil UN adalah untuk melakukan pemetaan tingkat pencapaian hasil belajar

siswa pada satuan pendidikan [2]. Ditemukan data hasil nilai rata-rata Ujian Nasional tingkat SMP, SMA dan SMK se-Indonesia yang diunggah oleh Kementerian Pendidikan dan Kebudayaan (KEMDIKBUD) [3] yang tersedia dalam bentuk format file Microsoft Excel (.xls).

Data pada website Pusat Penilaian Pendidikan tersedia sesuai urutan Provinsi dan Nomor Pokok Sekolah Nasional (NPSN). Namun, data yang disajikan tidak ada peringkat tertentu atau kelompok berdasarkan kategori sekolah dari hasil nilai Ujian Nasional sehingga data menjadi kurang jelas. Maka, dilakukanlah pengelompokan terhadap data nilai UN dengan menggunakan algoritma klustering, yaitu metode k-medoids dan k-means. Algoritma k-means merupakan algoritma dalam bidang data mining yang sederhana untuk diimplementasikan[4]. Terdapat pengembangan varian dari metode k-means klustering yaitu k-medoids. K-Medoids dikembangkan bertujuan untuk mengurangi sensitivitas dari partisi yang dihasilkan berdasarkan pengamatan mean yang dimiliki oleh setiap klaster[5].

Penelitian yang dilakukan oleh Pramesti et al.[6] menjelaskan perbedaan pada kedua algoritma terletak pada pemilihan titik centroid yang digunakan untuk perhitungan pada iterasi berikutnya. Pada algoritma k-medoids pemilihan titik centroid menggunakan objek sebagai perwakilan (medoid) sebanyak jumlah klaster yang akan dicari. Sedangkan, pada algoritma k-means pemilihan titik centroid menggunakan nilai rata-rata (mean) yang terbentuk dari hasil perhitungan terdekat dari iterasi pertama. Selanjutnya, penelitian sebelumnya tentang pengelompokan jumlah sumber daya manusia kesehatan puskesmas di provinsi Jawa Tengah oleh Dewi et al.[7]. Algoritma yang digunakan yaitu K-Means. Penelitian ini menghasilkan pengelompokan dengan klaster tinggi (kelebihan) dengan 4 kabupaten/kota, 24 kabupaten/kota termasuk ke dalam klaster sedang (cukup) dan 7 kabupaten/kota dengan nilai rendah (kurang). Lalu, penelitian yang dilakukan oleh Kamila et al.[5] terhadap pengelompokan data transaksi bongkar muat di Provinsi Riau menggunakan algoritma K-Means dan K-Medoids. Hasil yang diperoleh dengan menggunakan algoritma K-Means hanya membutuhkan waktu rata-rata 1 detik sedangkan pengolahan data pada K-Medoids membutuhkan rata-rata 1 menit 38 detik pada Rapidminer. Nilai Davies-Bouldin Index (DBI) pada K-Means lebih rendah dibandingkan K-Medoids yaitu masing-masing 0.112 dan 0.119.

Penelitian yang dilakukan oleh Silitonga et al. [8] tentang pengelompokan rumah tangga dalam perlakuan memilah sampah menurut provinsi dengan menerapkan algoritma K-Medoids. Hasil yang diperoleh dari penelitian tersebut yaitu 2 klaster yaitu klaster dengan tingkat memilah sampah rendah sebanyak 22 provinsi dan klaster dengan tingkat memilah sampah tinggi dengan 12 provinsi lainnya. Selanjutnya, penelitian yang dilakukan oleh Siska [9] tentang penerapan algoritma K-Means dalam menentukan dan menganalisa kubikasi air terjual. Dataset yang digunakan yaitu 20 sampel pelanggan berdasarkan pemakaian air setiap bulannya. Hasil dari penelitian tersebut menghasilkan klaster dengan pemakaian air sedang sebanyak 7 pelanggan, klaster dengan pemakaian air boros sebanyak 2 pelanggan, dan klaster dengan pemakaian air hemat sebanyak 11 pelanggan.

Berdasarkan beberapa penelitian yang telah dilakukan [5, 6, 7, 8, 9], terbukti kedua algoritma k-means dan k-medoids memiliki kemiripan dalam perhitungan untuk mengelompokan data sesuai dengan kategori yang dibutuhkan. Sehingga dari data hasil nilai Ujian Nasional yang didapatkan akan dilakukan penelitian menggunakan teknik data mining yaitu algoritma K-Means dan K-Medoids yang akan melakukan pengelompokan menjadi 3 kategori yaitu, baik, sedang dan cukup. Pengelompokan dilakukan untuk memudahkan siswa/i mendapatkan informasi mengenai kategori sekolah mereka, dan dapat dijadikan acuan kepada masyarakat saat mendaftarkan putra/i mereka untuk menempuh pendidikan selanjutnya.

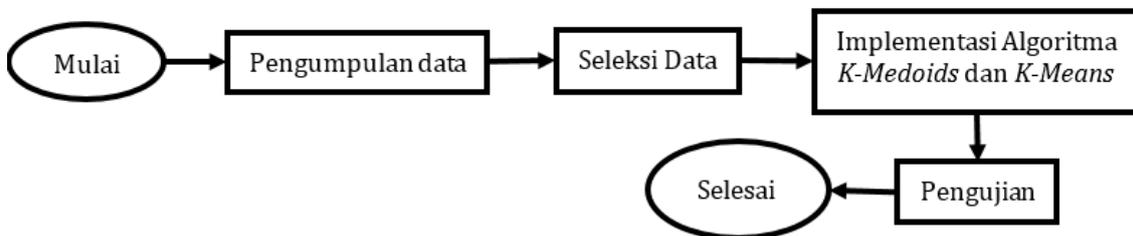
METODE PENELITIAN

Pengumpulan Data

Pengumpulan data dilakukan pada *website* resmi Pusat Penelitian Pendidikan, data yang digunakan dalam penelitian merupakan data nilai Ujian Nasional SMK tahun pelajaran 2018/2019 dengan jumlah satuan pendidikan 13.498 sekolah yang tersebar di 34 Provinsi dan 514 Kabupaten, dari data tersebut maka dilakukanlah penyeleksian data.

Gambaran Umum Penelitian

Penelitian dimulai dengan mengumpulkan data yang akan digunakan dalam penerapan algoritma. Data yang telah terkumpul lalu diseleksi untuk mengambil sampel yang bisa dijadikan perhitungan. Selanjutnya, data tersebut akan diolah menggunakan algoritma *K-Means* dan *K-Medoids* dan dari hasil implementasi lalu diuji menggunakan software *Rapidminer*. Adapun gambaran umum penelitian yang digunakan, seperti pada Gambar 1.



Gambar 1. Gambaran Umum Penelitian

Algoritma K-Medoids

K-Medoids merupakan pengelompokan partisi menggunakan sekumpulan objek yang dikelompokkan sebanyak jumlah *cluster*, dimana setiap objek mewakili *cluster* tersebut. Objek yang mewakilinya disebut dengan *medoids*. *Medoid* merupakan perwakilan *cluster* yang dijadikan sebagai pusat perhitungan jarak [10]. Titik acuan yang digunakan oleh *K-Medoids* tidak diambil dari perhitungan nilai *mean* dari objek dalam *cluster* seperti pada metode *K-Means*. Algoritma *K-Medoids* mengambil titik acuan berdasarkan objek representatif, dengan model partisi meminimalkan jumlah dari ketidaksamaan antara setiap objek dan titik acuan yang sesuai. Langkah-langkah algoritma *K-Medoids* sebagai berikut :

1. Tentukan *k* (jumlah *cluster*) yang diinginkan dari data yang diproses.
2. Pilih titik *centroid* secara acak atau berurutan dari data *medoid* awal sebanyak jumlah *k*.
3. Hitung jarak masing-masing data dengan menggunakan titik *centroid* sebanyak jumlah *k*, menggunakan rumus formula. Ada beberapa cara yang dapat digunakan untuk mengukur jarak data ke titik *centroid*. Pada penelitian ini pengukuran jarak akan dilakukan menggunakan rumus *Euclidean Distance* yang ditunjukkan pada persamaan (1).

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

Keterangan:

$D(x_2, x_1)$: Jarak variabel x_2 terhadap x_1
 j : variabel data ($j = 1, 2, 3, \dots, n$)
 p : dimensi data
 x_2 : variabel data yang diujikan (*testing*)
 x_1 : data sampel (*training*)

D merupakan jarak antara data x_2 dan x_1 , dan $| \cdot |$ bersifat mutlak.

4. Lakukan iterasi *medoid* dengan menghitung jarak menggunakan salah satu metode perhitungan jarak, terhadap seluruh data yang dijadikan data proses.
5. Hitung total simpangan (S) dengan hasil perhitungan jarak pada semua data. Dan lakukan perhitungan iterasi lagi dengan titik *centroid* yang baru, dengan menggunakan langkah awal kembali. Titik *centroid* yang baru diambil secara acak dari data *medoid*. Jika a adalah total perhitungan jarak terdekat dengan menggunakan *medoid* awal, dan b adalah total perhitungan jarak terdekat antara obyek ke *medoid* baru, maka total simpangan dapat dihitung menggunakan rumus persamaan (2).

$$S = b - a \quad (2)$$

Keterangan :

S : Total simpangan

b : Total perhitungan jarak terkecil antara obyek ke *medoid* baru

a : Total perhitungan jarak terkecil antara obyek ke *medoid* awal

Jika $S < 0$, maka hasil total perhitungan jarak a dan b dapat ditukar. Iterasi dihentikan apabila nilai total simpangan $b > a$, dengan anggota *cluster* terbentuk pada pola perhitungan jarak terkecil pada tiap *cluster* [10].

Algoritma K-Means

K-Means clustering merupakan metode pengelompokan data dalam bentuk satu atau lebih *cluster*. Data yang telah dihitung akan menghasilkan pola jarak terkecil sejumlah *cluster* yang dicari. Pola yang terbentuk pada perhitungan akan menghasilkan kelompok dengan karakteristik data yang berbeda pada setiap kelompoknya [11]. Adapun tahapan algoritma *K-Means* [12] adalah sebagai berikut :

1. Tentukan *k* sejumlah *cluster* yang akan dicari
2. Pilih titik *centroid* secara berurutan atau acak dari data awal sebanyak jumlah *k*.
3. Hitung jarak masing-masing data menggunakan titik *centroid* sebanyak jumlah *k*.
4. Setiap pusat *cluster* dihitung ulang berdasarkan dari nilai rata-rata dalam *cluster* yang didapatkan.
5. Langkah 3 dan 4 diulangi sampai kelompok jarak terkecil tidak mengalami perubahan pola *cluster*.

Seleksi Data

Penyeleksian data dilakukan untuk mengambil sampel dalam penerapan algoritma. Hasil seleksi data terhadap 13.498 sekolah SMK yang tersebar di 34 Provinsi dan 514 Kabupaten, difokuskan hanya terhadap Kabupaten Karawang Provinsi Jawa Barat yang tercatat memiliki 109 data sekolah. Dari 11 atribut yang terdapat pada data hasil nilai Ujian Nasional, diambil 6 atribut yang dapat digunakan dalam proses pengelompokan. Atribut yang dipilih untuk dijadikan acuan tersebut yaitu No, Satuan Pendidikan, Bahasa Indonesia, Bahasa Inggris, Matematika, dan Kompetensi. Berikut hasil dari seleksi data yang akan digunakan sebagai dataset terdapat pada Tabel 1.

Tabel 1. Data set Hasil Ujian Nasional di Kab. Karawang

No	Satuan Pendidikan	Bhs. Indonesia	Bhs. Inggris	Matematika	Kompetensi
1	SMK Negeri 1 Cikampek	70,98	44,72	37,1	41,3
2	SMK Negeri 1 Cilamaya	63,14	35,99	30,43	34,54
3	SMK Negeri 1 Jatisari	67,41	38,94	32,42	47,59

No	Satuan Pendidikan	Bhs. Indonesia	Bhs. Inggris	Matematika	Kompetensi
4	SMK Negeri 1 Karawang	71,91	46,73	35,39	42,05
5	SMK Negeri 1 Rengasdengklok	69,73	40,78	33,89	42,54
.					
108	SMK Al Inayah Kutamukti	59,81	34,22	32,87	39,21
109	SMK Mandalla Entrepreneur School	60,16	34,04	29,13	42,09

Pengujian

Pengujian dilakukan menggunakan *tools Rapidminer 5.3* yang bertujuan untuk mencocokkan kesesuaian hasil dari perhitungan implementasi algoritma.

HASIL DAN PEMBAHASAN

Hasil Algoritma K-Medoids

1. Jumlah *cluster* yang akan dicari adalah 3 *cluster*, yaitu nilai baik, sedang, dan cukup
2. Titik *centroid* yang akan dijadikan titik awal dipilih secara acak dari dataset, titik *centroid* yang dipilih sejumlah dengan *cluster* yang sudah ditentukan

Tabel 2. Titik *centroid* iterasi ke-1 Algoritma K-Medoids

Iterasi 1	Satuan Pendidikan	B.Indo	B.Ingggris	Matematika	Kompetensi
<i>Centroid 1</i>	SMK Yapinas	51,61	30,61	27,95	36,14
<i>Centroid 2</i>	SMK Rosma	63,41	40,01	31,95	40,43
<i>Centroid 3</i>	SMKN 1 Tirtamulya	62,79	37,95	31,4	39,89

Pada Tabel 2 ditentukan titik *centroid* yang akan dijadikan titik awal dalam perhitungan algoritma K-Medoids.

3. Pengukuran jarak dilakukan dengan menggunakan rumus *Euclidean Distance* yang mengacu pada persamaan 1 yang diterapkan pada perhitungan iterasi *medoid*.
4. Hasil iterasi *medoid* dilakukan dengan menghitung jarak terhadap seluruh data yang dijadikan data proses. Hasil dari perhitungan jarak iterasi ke-1 terhadap keseluruhan data proses dapat dilihat pada Tabel 3.

Tabel 3. Hasil Iterasi ke-1 Algoritma K-Medoids

No	D1	D2	D3	Jarak Terkecil	C1	C2	C3
1	26,1656	10,3256	12,1403	10,3256		OK	
2	13,0612	7,2967	5,7903	5,7903			OK
3	21,6821	8,2796	9,0915	8,2796		OK	
4	27,6084	11,4759	13,4480	11,4759		OK	
5	22,5390	6,9731	8,3304	6,9731		OK	
.							
.							
109	11,0301	7,5483	5,6743	5,6743			OK
Total Jarak Terkecil				798,2766			

Pada Tabel 3 hasil total perhitungan jarak terkecil pada perhitungan iterasi ke-1 yaitu 798,2766 dengan total perhitungan jarak terkecil diambil dari setiap hasil perhitungan titik *centroid* 1, 2 dan 3.

- Setelah mengetahui hasil iterasi ke-1 selanjutnya perhitungan jarak dilakukan kembali untuk mendapatkan hasil iterasi ke-2 sebagai perbandingan total simpangan diantara kedua iterasi dengan menggunakan titik *centroid* yang baru dapat dilihat pada Tabel 4.

Tabel 4. Titik *centroid* iterasi ke-2 Algoritma *K-Medoids*

Iterasi 2	Satuan Pendidikan	B.Indo	B.Ingggris	Matematika	Kompetensi
<i>Centroid</i> 1	SMK Kalimasyada	59,27	41,18	40,34	38,41
<i>Centroid</i> 2	SMK INOTEK Tempuran	51,7	34	28,71	32,5
<i>Centroid</i> 3	SMK Mandalla Entrepreneur School	60,16	34,04	29,13	42,09

Hasil perhitungan jarak titik *centroid* baru terhadap keseluruhan data proses seperti pada Tabel 5.

Tabel 5. Hasil Iterasi ke-2 Algoritma *K-Medoids*

No	D1	D2	D3	Jarak Terkecil	C1	C2	C3
1	12,9810	25,1887	17,1837	12,9810	OK		
2	12,4538	11,9144	8,4484	8,4484			OK
3	14,7742	22,6424	10,8465	10,8465			OK
4	15,1104	26,5767	18,3926	15,1104	OK		
5	12,9704	22,3312	12,6441	12,6441			OK
.							
.							
.							
109	13,8195	12,7952	0,0000	0,0000			OK
Total Jarak Terkecil				809,4800			

Tabel 5 hasil total perhitungan jarak terkecil pada perhitungan iterasi ke-2 yaitu 809,4800 dengan total perhitungan jarak terkecil diambil dari setiap hasil perhitungan titik *centroid* 1, 2 dan 3.

- Kemudian dilakukan perhitungan total simpangan (S) yang dihitung dari hasil perhitungan total jarak terkecil iterasi ke-1 dan ke-2. Dihitung dengan menggunakan perhitungan simpangan S dari jarak keduanya.

Total jarak iterasi ke-1 (a) = 798,2766

Total jarak iterasi ke-2 (b) = 809,4800

Total simpangan S = 809,4800 - 798,2766 = 11,2034

Setelah dilakukan perhitungan simpangan maka dapat disimpulkan bahwa $b > a$, sehingga iterasi diberhentikan dengan anggota *cluster* diambil dari iterasi terakhir yang terbentuk dari perhitungan jarak b. Lalu, menentukan *range* nilai Ujian Nasional maksimal dan minimal dari tiap *cluster*, yang bertujuan untuk mengelompokkan kategori *cluster*. Berikut hasil akhir dari pengelompokan berdasarkan kategori dapat dilihat pada Tabel 6.

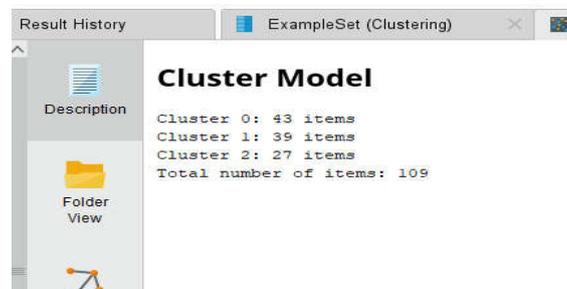
Tabel 6. Hasil Pengelompokan SMK di Kab. Karawang Algoritma *K-Medoids*

No	Satuan Pendidikan	Baik	Cukup	Sedang
1	SMK Negeri 1 Cikampek	OK		
2	SMK Negeri 1 Cilamaya			OK
3	SMK Negeri 1 Jatisari			OK
4	SMK Negeri 1 Karawang	OK		

No	Satuan Pendidikan	Baik	Cukup	Sedang
5	SMK Negeri 1 Rengasdengklok			OK
.				
107	SMK INOTEK Tempuran		OK	
108	SMK Al Inayah Kutamukti			OK
109	SMK Mandalla Entrepreneur School			OK
	Total	27	39	43

Pada Tabel 6 dijelaskan sekolah mana saja yang termasuk kategori baik, sedang dan cukup. Kategori baik memiliki 27 anggota satuan pendidikan, kategori cukup memiliki 39 anggota dan kategori sedang memiliki 43 anggota. Data hasil pengelompokan akan digunakan untuk pengujian menggunakan *tools*.

7. Pada pengujian menggunakan *Rapidminer* algoritma *K-Medoids* menjelaskan bahwa hasil pengujian sesuai dengan perhitungan seperti pada Gambar 2.



Gambar 2. Cluster Algoritma *K-Medoids*

Gambar 2 menjelaskan bahwa *cluster 0* memiliki 43 anggota satuan pendidikan SMK merupakan kategori sedang, *cluster 1* memiliki 39 anggota termasuk kedalam kategori cukup dan *cluster 2* memiliki 27 anggota satuan pendidikan SMK merupakan kategori baik dari hasil pengujian menggunakan algoritma *k-medoids*.

Hasil Algoritma K-Means

Tahap selanjutnya mengimplementasikan dataset menggunakan algoritma *k-means*. Langkah-langkah implementasi algoritma *K-Means* hampir sama dengan tahapan pada algoritma *K-Medoids* yang membedakan hanya pada titik *centroid* yang digunakan. *K-Medoids* melakukan pemilihan titik *centroid* baru secara acak atau berurutan pada perhitungan jarak ke-2, sedangkan *K-Means* mengambil nilai *mean* pada jarak terkecil untuk dijadikan titik *centroid* di perhitungan jarak iterasi selanjutnya.

1. Setelah melakukan perhitungan pada beberapa iterasi menggunakan rumus *Euclidean Distance* pada persamaan 1, perubahan pola *cluster* terhenti pada iterasi ke-10 yang berarti pola *cluster* yang terbentuk sudah tidak akan mengalami perubahan lagi. Berikut hasil dari perhitungan jarak iterasi ke-10 terhadap keseluruhan data proses menggunakan algoritma *k-means* :

Tabel 7. Hasil Iterasi ke-10 Algoritma *K-Means*

No	D1	D2	D3	C1	C2	C3
1	8,1348	22,1167	12,6238	OK		
2	19,4969	9,3194	7,4522			OK
3	10,2589	18,9622	9,5332			OK
4	8,5640	23,6139	14,1023	OK		
5	9,8698	18,9596	9,4005			OK
.						
109	18,4008	9,1323	6,0348			OK

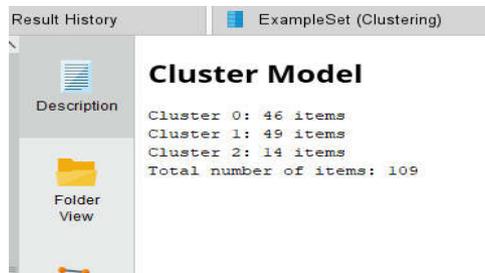
Pada Tabel 7 hasil iterasi algoritma *K-Means* jarak terkecil yang dihasilkan dari perhitungan titik *centroid* 1, 2, dan 3 sudah tidak mengalami perubahan lagi. Selanjutnya, menentukan *range* nilai Ujian Nasional maksimal dan minimal dari tiap *cluster*, yang bertujuan untuk mengelompokkan kategori *cluster*. Berikut hasil akhir dari pengelompokan berdasarkan kategori :

Tabel 8. Hasil Pengelompokan SMK di Kab. Karawang Algoritma *K-Means*

No	Satuan Pendidikan	Baik	Cukup	Sedang
1	SMK Negeri 1 Cikampek	OK		
2	SMK Negeri 1 Cilamaya			OK
3	SMK Negeri 1 Jatisari			OK
4	SMK Negeri 1 Karawang	OK		
5	SMK Negeri 1 Rengasdengklok			OK
...				
107	SMK INOTEK Tempuran		OK	
108	SMK Al Inayah Kutamukti			OK
109	SMK Mandalla Entrepreneur School			OK
	Total	14	49	46

Pada Tabel 8 dijelaskan sekolah mana saja yang termasuk kategori baik, sedang dan cukup. Kategori baik memiliki 14 anggota satuan pendidikan, kategori cukup memiliki 49 anggota dan kategori sedang memiliki 46 anggota. Data hasil pengelompokan akan digunakan untuk pengujian menggunakan *tools*.

2. Pada pengujian menggunakan *Rapidminer* algoritma *K-Means* menjelaskan bahwa hasil pengujian sesuai dengan perhitungan seperti pada gambar 3.



Gambar 3. Cluster Algoritma *K-Means*

Gambar 3 menjelaskan bahwa *cluster* 0 memiliki 46 anggota satuan pendidikan SMK merupakan kategori sedang, *cluster* 1 memiliki 49 anggota termasuk kedalam kategori cukup dan *cluster* 2 memiliki 14 anggota satuan pendidikan SMK merupakan kategori baik dari hasil pengujian menggunakan algoritma *k-means*.

Hasil Penelitian

Hasil penelitian terhadap data nilai Ujian Nasional SMK di Kabupaten Karawang menggunakan metode pengelompokan algoritma *K-Means* dan *K-Medoids* dapat dilihat pada Tabel 9.

Tabel 9. Hasil Penelitian

No	Algoritma <i>K-Means</i>	Algoritma <i>K-Medoids</i>
1.	Hasil perhitungan manual diperoleh hasil <i>cluster</i> baik sebanyak 14 anggota, <i>cluster</i> sedang 46 anggota, dan <i>cluster</i> cukup 49 anggota.	Hasil perhitungan manual diperoleh hasil <i>cluster</i> baik sebanyak 27 anggota, <i>cluster</i> sedang 43 anggota, dan <i>cluster</i> cukup 39 anggota.

No	Algoritma K-Means	Algoritma K-Medoids
2.	Hasil pengujian menggunakan <i>tools Rapidminer</i> memperoleh kesesuaian jumlah <i>cluster</i> yang sama dengan menggunakan perhitungan manual.	Hasil pengujian menggunakan <i>tools Rapidminer</i> memperoleh kesesuaian jumlah <i>cluster</i> yang sama dengan menggunakan perhitungan manual.
3.	Hasil diperoleh dengan menghitung sebanyak 10 iterasi.	Hasil didapatkan dengan menghitung sebanyak 2 iterasi.
4.	Pengujian menggunakan algoritma <i>k-means</i> pada <i>tools</i> menghasilkan jumlah <i>cluster</i> yang sama dengan menggunakan titik <i>centroid</i> berurutan atau secara acak	Pengujian algoritma <i>k-medoids</i> menggunakan <i>tools</i> tidak dapat memilih titik <i>centroid</i> , sehingga titik <i>centroid</i> yang digunakan pada perhitungan manual dicocokkan terlebih dahulu dengan yang digunakan pada pengujian.

Berdasarkan Tabel 9 dapat disimpulkan bahwa perhitungan manual menggunakan algoritma *K-Means* dan *K-Medoids* memperoleh jumlah anggota *cluster* yang sesuai saat dilakukan pengujian menggunakan *tools Rapidminer*.

KESIMPULAN

Hasil implementasi didapatkan 3 kluster yaitu baik, sedang, dan cukup terhadap data nilai Ujian Nasional. Hasil algoritma *k-means* diperoleh hasil kluster 1 dengan kategori baik memiliki 14 anggota, kluster 2 dengan kategori cukup memiliki 49 anggota, dan kluster 3 dengan kategori sedang memiliki 46 anggota. Hasil perhitungan manual algoritma *k-medoids* diperoleh hasil kluster 1 dengan kategori baik memiliki 27 anggota, kluster 2 dengan kategori cukup memiliki 48 anggota, dan kluster 3 kategori sedang memiliki 43 anggota. Pengujian pada masing-masing algoritma terhadap *tools* menghasilkan kesesuaian jumlah kluster yang sama antara hasil perhitungan manual algoritma *k-means* dengan *Rapidminer* dan algoritma *k-medoids* dengan *Rapidminer*. Pada penelitian ini disimpulkan bahwa pengelompokan data menggunakan algoritma *k-means* dan *k-medoids* memiliki pola perhitungan data yang berbeda. Sehingga keunggulan kinerja algoritma tergantung pada data yang ingin diproses.

UCAPAN TERIMA KASIH

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Ninda Nurul Rhamadani yang dibimbing oleh Ahmad Fauzi dan Euis Nurlaelasari.

REFERENSI

- [1] KBBI. (2019). Perpustakaan.
- [2] Pusat Penelitian Pendidikan Balitbang Kemdikbud. (2019). Ujian Nasional. https://puspendik.kemdikbud.go.id/ujian-nasional_un#Sejarah_Ujian_Nasional. (Diakses 10 Desember 2019).
- [3] Pusat Penelitian Pendidikan Kemdikbud. (2019). Laporan Hasil Ujian Nasional. <https://puspendik.kemdikbud.go.id/hasil-un/>. (Diakses 05 Desember 2019).
- [4] Balabantary, RC., *et al.* (2013). Document Clustering Using K-Means and K-Medoids. *Journal of Knowledge Based Computer System*.
- [5] Kamila, I., *et al.* (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. Vol. 5, No. 1, Hal. 119-125 e-ISSN 2502-8995 p-ISSN 2460-8181.

-
- [6] Pramesti, D. F., *et al.* (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data. 1(9), 723–732. <https://doi.org/10.1109/EUMC.2008.4751704>.
- [7] Dewi, S. C., *et al.* (2020). Pengelompokan Jumlah Sumber Daya Manusia Kesehatan Puskesmas untuk Menunjang Pemerataan pada Provinsi Jawa Tengah Menggunakan Algoritma K-Means. Vol. 1, No: 2, ISSN: 2715-2766.
- [8] Silitonga, D. A., *et al.* (2019). Penerapan Metode K-Medoids pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi. *Seminar Nasional Sains & Teknologi Informasi (SENSASI) SENSASI 2019*, ISBN: 978-602-52720-2-8.313-318.
- [9] Siska, ST. (2016). Analisa Dan Penerapan Data Mining Untuk Menentukan Kubikasi Air Terjual Berdasarkan Pengelompokan Pelanggan Menggunakan Algoritma K-Means Clustering. *Jurnal Teknologi Informasi & Pendidikan*, ISSN: 2086-4981.
- [10] Binty, Tri. (2019) K-Medoids/Partitioning Around Medoids (PAM)-Non *Hierarchical Clustering with R*. <https://medium.com/@tribinty/k-medoids-9d0af590bbc0>.
- [11] Agusta. (2007). Implementasi Algoritma Clustering dengan Singular Vector Decomposition Untuk Menunjang Keputusan Dalam Meningkatkan Produktivitas Tanaman Jagung. Bali.
- [12] Singh K, & Sharma, Naveen. (2011). Evolving Limitations in K-Means Algorithm in Data Mining and Their Removal. *IJCEM International Journal of Computational Engineering & Management*.