



Implementasi Algoritma *K-Means* untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA

Miftahul Hasanah^{1,2}, Sarjon Defit², Gunadi Widi Nurcahyo³

^{1,2,3}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK Padang
mhasanah45@gmail.com

Abstract

The abundance of students causes student data in the system to also be abundant. Schools often find it difficult to manage large amounts of data manually, especially in selecting National Science Olympiad participants and decisions made are less effective. So this research was conducted with the aim of helping the school in selecting OSN participants appropriately and effectively. The method used is Clustering with *K-Means* algorithm on the report card grades of students majoring in Natural Sciences at SMA Negeri 5 Sijunjung. The results in this study get 3 clusters of students on the selection of OSN participants, namely students who are Very Competent, Competent and Less Competent. This research can be used as a benchmark used by schools in making decisions on the selection of OSN participants.

Keywords: *Data Mining*, *K-Means Algorithm*, *Clustering*, *Centroid*, OSN

Abstrak

Melimpahnya jumlah siswa menyebabkan data-data siswa dalam sistem juga melimpah. Pihak sekolah sering kesulitan dalam mengelola data yang banyak secara manual khususnya dalam pemilihan peserta Olimpiade Sains Nasional (OSN) dan keputusan yang diambil kurang efektif. Maka dilakukan penelitian ini dengan tujuan membantu pihak sekolah dalam pemilihan peserta OSN secara tepat dan efektif. Metode yang digunakan adalah *Clustering* dengan algoritma *K-Means* terhadap data nilai rapor siswa jurusan IPA di SMA Negeri 5 Sijunjung. Hasil dalam penelitian ini mendapatkan 3 cluster siswa terhadap pemilihan peserta OSN, yaitu siswa yang Sangat Berkompeten, Berkompeten dan Kurang Berkompeten. Penelitian ini dapat dijadikan sebagai tolak ukur yang digunakan sekolah dalam pengambilan keputusan terhadap pemilihan peserta OSN.

Kata kunci : *K-Means Clustering*, *Centroid*, Olimpiade Sains, Kompetensi, Siswa Terbaik

© 2019 JSisfotek

1. Pendahuluan

Olimpiade Sains Nasional (OSN) merupakan suatu ajang kompetisi dalam bidang sains yang dilakukan sekali dalam setahun. Jumlah siswa dalam institusi pendidikan khususnya Sekolah Menengah Atas (SMA) semakin meningkat yang menyebabkan data-data siswa dalam sistem juga semakin meningkat. Pengelolaan data siswa yang banyak secara manual sering menyulitkan pihak sekolah dalam pengambilan keputusan. Pemilihan peserta OSN membutuhkan proses seleksi akurat dari pihak sekolah agar hasil yang didapat juga memuaskan. Data yang digunakan dalam penyeleksian peserta OSN adalah data nilai rapor akademik siswa pada beberapa semester sebelumnya.

Data Mining adalah suatu proses pengolahan data yang bertujuan untuk mendapatkan, menggali dan menemukan pengetahuan yang tersembunyi dari sebuah *dataset* atau sekumpulan data yang berukuran sangat besar [1]. Hasil dari pengolahan data berupa suatu informasi berharga yang dapat digunakan untuk

mengambil keputusan di masa depan [2]. Menurut Maulida (2018) *Data Mining* merupakan suatu proses pencarian informasi dalam sebuah kelompok data dengan menggunakan teknik atau metode tertentu, dimana metode dalam *Data Mining* bervariasi, tergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) [3].

Salah satu metode yang dapat digunakan dalam *Data Mining* adalah metode *Clustering* dengan algoritma *K-Means* [4]. Algoritma *K-Means Clustering* adalah teknik dalam *Data Mining* yang mempartisi data yang ada ke dalam beberapa *cluster* sehingga data yang memiliki karakteristik yang sama akan dikelompokkan ke dalam satu *cluster* sedangkan data dengan karakteristik yang berbeda akan dikelompokkan ke dalam *cluster* lain [5]. Algoritma *K-Means* melakukan dua proses yaitu proses penentuan pusat *cluster* (*centroid*) dan proses pencarian anggota dari tiap-tiap *cluster* [6].

Algoritma *K-Means* membutuhkan parameter *input* sebanyak *k* atau jumlah *cluster*, kemudian akan

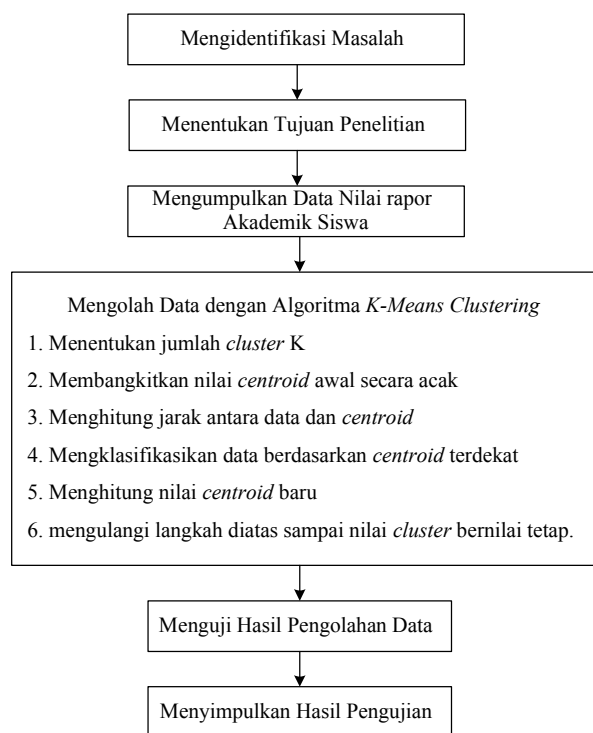
membagi sekumpulan objek n ke dalam cluster, sehingga akan didapatkan tingkat kemiripan data dengan pusat cluster, sedangkan kemiripan dengan data pada cluster lain sangat rendah [7]. Penerapan *Data Mining* dalam mengolah data pada sebuah kasus sangat berguna dalam meningkatkan daya tarik digital dan juga dapat menghasilkan informasi yang lebih bermakna [8]. Algoritma *K-Means Clustering* dapat diimplementasikan dengan Algoritma Genetika [9].

Penelitian terdahulu menggunakan Algoritma *K-Means Clustering* dilakukan oleh Yunita (2018) yang meneliti tentang pengelompokan penerimaan mahasiswa baru yang menghasilkan 3 kelompok mahasiswa baru yaitu C1 195 orang, C2 271 orang dan C3 50 orang, dimana hasil cluster digunakan sebagai dasar strategi promosi program studi yang ada [10]. Penelitian lain dilakukan oleh Lubis (2016) yang meneliti tentang segmentasi pelanggan di Kantor Pos Medan, yang menghasilkan model aturan untuk pelanggan yang berpotensi dan jenis layanan terbanyak yang digunakan pelanggan [11]. Selain itu Purba, dkk. (2018) melakukan penelitian tentang kelompok siswa yang berpotensi *drop-out*, yang menghasilkan terdapat 2 *cluster* siswa, yaitu C1 merupakan siswa yang berpotensi *drop-out* dan C2 siswa yang tidak berpotensi *drop-out* dari sekolah [12]. *K-means Clustering* dapat mendeteksi tingkat kerusakan pada buah [13]. Pemilihan *centroid* awal berpengaruh terhadap keefektifan hasil dari *Clustering* [14]. Ismi dalam penelitiannya menemukan bahwa *K-Means Clustering* dapat dijadikan landasan pemilihan data potensial dari data yang berdimensi tinggi [15].

Penelitian ini dilakukan untuk mencari kelompok nilai akademik siswa dengan cara menghitung jarak terdekat antara data dengan titik tengah (*centroid*) suatu *cluster*. Sehingga pengelompokan tersebut akan menghasilkan beberapa *cluster* siswa yaitu siswa yang sangat berkompeten, berkompeten, dan kurang berkompeten untuk mengikuti Olimpiade Sains Nasional (OSN). Melalui penelitian ini dengan menggunakan algoritma *K-Means Clustering* diharapkan dapat membantu pihak sekolah dan guru dalam mengelompokkan siswa berdasarkan nilai akademik dengan hasil yang efektif dan efisien.

2. Metodologi Penelitian

Agar penelitian berjalan dengan baik, maka diperlukan kerangka kerja dari penelitian. Kerangka penelitian merupakan tahapan-tahapan yang dilakukan dalam menyelesaikan penelitian. Adapun kerangka kerja penelitian dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian pada Gambar 1. Maka dapat diuraikan sebagai berikut:

2.1. Mengidentifikasi Masalah

Pada tahap ini dilakukan proses mendeskripsikan masalah, menganalisa masalah yang ada pada sistem berjalan serta memahami masalah tersebut agar didapatkan solusi dan penyelesaian dari masalah.

2.2. Tujuan penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah menentukan *cluster* siswa yang berhak untuk mengikuti OSN berdasarkan nilai rapor akademik pada semester sebelumnya sebanyak 3 *cluster*, yaitu (C1) Sangat Berkompeten, (C1) Berkompeten dan (C1) Kurang Berkompeten. Selain itu penelitian ini bertujuan untuk membantu pihak sekolah dalam pemilihan peserta OSN secara efektif.

2.3. Mengumpulkan Data

Data yang digunakan dalam penelitian ini adalah data nilai rapor akademik siswa jurusan IPA di SMA Negeri 5 Sijunjung. Data didapatkan dari hasil penelitian lapangan yang dilakukan di SMA Negeri 5 Sijunjung melalui pengambilan file dan melakukan wawancara dengan pihak sekolah.

2.4. Mengolah Data dengan Algoritma *K-Means Clustering*

Langkah-langkah mengolah data dengan Algoritma *K-means Clustering* adalah:

1. Menentukan jumlah *cluster* K . Jumlah *cluster* tergantung pada kebutuhan sistem

2. Menentukan nilai titik tengah (*centroid*) secara acak sebanyak jumlah *cluster* K.

3. Alokasikan masing-masing data ke pusat *cluster* terdekat, dengan rumus:

$$D(x,y) = \sqrt{(x1 - yi)^2 + (xi - yi)^2} \quad (1)$$

Di mana :

D(x,y) = Jarak data ke pusat *cluster*

x = Data *record*

y = Data *centroid*

4. Hitung ulang pusat *cluster* (*centroid*) baru, dengan rumus:

$$Cl = \frac{x1 + x2 + x3 + \dots + xn}{\sum x} \quad (2)$$

Di mana :

Cl = Centroid baru

X1 = Nilai data *record* ke-1

Xn = Nilai data *record* ke-n

$\sum x$ = Jumlah data *record*

5. Alokasikan lagi data ke pusat *cluster* yang baru didapatkan. Jika pusat *cluster* masih mengalami perubahan, maka proses diulangi dari langkah 3, tapi jika tidak mengalami perubahan, maka proses pencarian dihentikan.

2.5. Menguji Hasil Pengolahan Data

Menguji hasil bertujuan untuk mendapatkan keakuratan dari sistem yang dikembangkan sehinggal dapat berjalan dengan baik dan sesuai dengan tujuan penelitian, dapat dilakukan dengan cara sebagai berikut:

- Menyediakan data yang digunakan dalam pengujian perangkat lunak
- Menentukan hasil penelitian, yaitu *cluster* peserta Olimpiade.
- Menjalankan *software RapidMiner* menggunakan data yang sudah didapatkan
- Membandingkan hasil yang didapatkan dari perhitungan manual dengan perhitungan *RapidMiner*.

2.6. Menyimpulkan Hasil Pengujian

Kesimpulan diambil berdasarkan hasil perhitungan Algoritma K-Means Clustering dan pengujian menggunakan *software RapidMiner*, dimana hasilnya akan dijadikan rekomendasi untuk sekolah dan sebagai bahan pertimbangan dalam pengambilan keputusan.

3. Hasil dan Pembahasan

Menentukan pengelompokan siswa berdasarkan nilai rapor akademik untuk mengikuti OSN memiliki tahapan yang akan diimplementasikan sebagai berikut :

- Menentukan Data Nilai Siswa
Jumlah sampel data yang diolah dalam penelitian ini berjumlah 23 data yang disajikan pada Tabel 1.

Tabel 1. Sampel Data Nilai Siswa

Id Siswa	MTK	FSK	KMA	BLG	TIK
1	89,33	88,67	90	94	94
2	82	81,33	76,67	86,67	82,33
3	82,67	78	76,67	87	75
4	86,33	90,67	82,67	95	87,33
5	77,67	83,33	77,33	91	84,33
6	75,33	80,33	73,67	84,66	83,67
7	84,67	86	82	90,33	92,67
8	76	73,67	63	83	76
9	75,33	71,33	57,67	62	73,67
10	76	80	77	91	84
11	75,33	79	72	83,33	82,67
12	87,33	83	79	83,67	79,33
13	82,33	82,33	80,33	93	88,67
14	87	89	88,33	93,67	91,33
15	79,67	79	76	82,67	80,67
16	85,67	80,67	79,33	89,67	86,33
17	86,67	90,33	84	92	90,33
18	79,67	77,67	78,33	81,67	84,33
19	84	77,33	77,33	83,33	79,33
20	84,33	79	83,67	84	87
21	80	84,33	78,67	93,33	87,33
22	78,67	78	77	88,33	85
23	76,67	75,67	73,33	79,67	82,33

b. Menentukan Jumlah Cluster

Jumlah *cluster* ditentukan berdasarkan kebutuhan yang akan menjadi acuan untuk mendapatkan hasil. Dalam penelitian ini jumlah *cluster* dibagi menjadi 3 *cluster*, yaitu (C1) Sangat Berkompeten, (C2) Berkompeten dan (C3) Kurang Berkompeten.

c. Menentukan Titik Centroid Secara Acak

Nilai *centroid* awal dilakukan secara acak dengan mengambil nilai siswa urutan atas, tengah dan bawah. Pusat *cluster* akan dikelompokkan ke dalam 3 *cluster* (C1, C2, C3). Data nilai siswa yang dijadikan *centroid* awal dapat dilihat pada Tabel 2.

Tabel 2. Nilai Centroid Awal

Id Siswa	Centroid	MTK	FSK	KMA	BLG	TIK
4	C1	86,33	90,67	82,67	95	87,33
12	C2	87,33	83	79	83,67	79,33
23	C3	76,67	75,67	73,33	79,67	82,33

d. Menghitung Jarak Terdekat Data dengan Centroid Iterasi 1 (Pertama)

Untuk menghitung pusat *cluster* 1 dihitung dari jarak tiap data ke titik *centroid* menggunakan rumus (1):

$$D_1 = \sqrt{(89,33 - 86,33)^2 + (88,67 - 90,67)^2 + (90 - 82,67)^2 + (94 - 95)^2 + (94 - 87,33)^2} = 10,59329$$

$$D_2 = \sqrt{(82 - 87,33)^2 + (81,33 - 83)^2 + (76,67 - 79)^2 + (86,67 - 83,67)^2 + (82,33 - 79,33)^2} = 7,39031$$

$$D_1 = \sqrt{(89,33 - 87,33)^2 + (88,67 - 83)^2 + (90 - 79)^2 + (94 - 83,67)^2 + (94 - 79,33)^2}$$

$$= 21,88823$$

$$D_2 = \sqrt{\frac{(82 - 87,33)^2 + (81,33 - 83)^2 + (76,67 - 79)^2}{+(86,67 - 83,67)^2 + (82,33 - 79,33)^2}}$$

$$= 7,39031$$

Cluster 3

$$D_1 = \sqrt{\frac{(89,33 - 76,67)^2 + (88,67 - 75,67)^2 + (90 - 73,33)^2}{+(94 - 79,67)^2 + (94 - 82,33)^2}}$$

$$= 30,80241$$

$$D_2 = \sqrt{\frac{(82 - 76,67)^2 + (81,33 - 75,67)^2 + (76,67 - 73,33)^2}{+(86,67 - 79,67)^2 + (82,33 - 82,33)^2}}$$

$$= 10,98490$$

Untuk nilai D_n seterusnya akan ditampilkan pada Tabel 3.

Tabel 3. Pengelompokan Data ke Cluster Terdekat Iterasi I

Id	Jarak ke C1	Jarak ke C2	Jarak ke C3	C1	C2	C3
1	10,59329	21,88823	30,80241	✓		
2	15,37242	7,39031	10,98490		✓	
3	20,63680	9,05551	12,65307		✓	
4	0	16,29996	25,80023	✓		
5	13,50181	13,22831	14,42929		✓	
6	20,71130	14,11940	7,10481			✓
7	8,67651	15,73108	21,60136	✓		
8	32,48056	21,97743	12,74110			✓
9	48,94282	35,16507	25,55256			✓
10	16,72518	14,72925	12,80307			✓
11	22,99570	14,84140	5,31305			✓
12	16,29996	0	14,98524		✓	
13	9,83498	14,19698	18,53202	✓		
14	7,28288	19,16136	28,04421	✓		
15	20,52856	9,29899	6,24447			✓
16	11,87569	9,65957	16,06238		✓	
17	4,47169	16,42094	25,39770	✓		
18	20,46851	10,79655	6,78233			✓
19	20,29431	6,79001	9,74392		✓	
20	16,19345	10,28342	14,74055		✓	
21	9,94823	14,59610	18,06830	✓		
22	17,35382	12,56647	10,25379			✓
23	25,80023	14,98524	0			✓
JUMLAH				7	7	9

Setelah melakukan perhitungan pada iterasi pertama maka didapatkan pengelompokan dari anggota C1, C2 dan C3 seperti di bawah ini:

- Anggota cluster 1 (C1) terdiri dari 7 siswa dengan nomor urut: (1, 4, 7, 13, 14, 17, 21)
- Anggota cluster 2 (C2) terdiri dari 7 siswa dengan nomor urut: (2, 3, 5, 12, 16, 19, 20)
- Anggota cluster 3 (C3) terdiri dari 9 siswa dengan nomor urut: (6, 8, 9, 10, 11, 15, 18, 22, 23)

Iterasi VI (Keenam)

Iterasi keenam merupakan iterasi terakhir dalam pengolahan data nilai siswa pemilihan peserta OSN.

Untuk melakukan iterasi keenam, maka nilai centroid harus ditentukan dengan cara menghitung rata-rata data pada centroid yang sama dari hasil pengelompokan iterasi kelima, dengan anggota C1, C2 dan C3 seperti dibawah ini:

- Anggota cluster 1 (C1) terdiri dari 7 siswa dengan nomor urut: (1, 4, 7, 13, 14, 17, 21)
- Anggota cluster 2 (C2) terdiri dari 14 siswa dengan nomor urut: (2, 3, 5, 6, 10, 11, 12, 15, 16, 18, 19, 20, 22, 23)
- Anggota cluster 3 (C3) terdiri dari 2 siswa dengan nomor urut: (8, 9)

Menghitung nilai centroid baru untuk iterasi keenam berdasarkan hasil iterasi kelima dapat menggunakan rumus (2).

$$C1(x1) = \frac{89,33 + 86,33 + 84,67 + 82,33 + 87 + 87,67 + 80}{7}$$

$$= 85,19$$

$$C1(x2) = \frac{88,67 + 90,67 + 86 + 82,33 + 89 + 90,33 + 84,33}{7}$$

$$= 87,33$$

$$C1(x3) = \frac{90 + 82,67 + 82 + 80,33 + 88,33 + 84 + 78,67}{7}$$

$$= 83,71$$

$$C1(x4) = \frac{94 + 95 + 90,33 + 93 + 93,67 + 92 + 93,33}{7}$$

$$= 85,19$$

$$C1(x5) = \frac{94 + 87,33 + 92,67 + 88,67 + 91,33 + 90,33 + 87,33}{7}$$

$$= 90,24$$

$$C2(x1) = \frac{82 + 82,67 + 77,67 + 75,33 + 76 + 75,33 + 87,33 + 79,67 + 85,67 + 79,67 + 84 + 84,33 + 78,67 + 76,67}{14}$$

$$= 80,36$$

$$C2(x2) = \frac{81,33 + 78 + 83,33 + 80,33 + 80 + 79 + 83 + 79 + 80,67 + 77,67 + 77,33 + 79 + 78 + 75,67}{14}$$

$$= 79,45$$

$$C2(x3) = \frac{76,67 + 76,67 + 77,33 + 73,67 + 77 + 72 + 79 + 76 + 79,33 + 78,33 + 77,33 + 83,67 + 77 + 73 + 73,33}{14}$$

$$= 76,95$$

$$C2(x4) = \frac{86,67 + 87 + 91 + 84,67 + 91 + 83,33 + 83,67 + 82,67 + 89,67 + 81,67 + 83,33 + 84 + 88,33 + 79,67}{14}$$

$$= 85,48$$

$$C2(x5) = \frac{82,33 + 75 + 84,33 + 83,67 + 84 + 82,67 + 79,33 + 80,67 + 86,33 + 84,33 + 79,33 + 87 + 85 + 82,33}{14}$$

$$= 82,59$$

$$C3(x1) = \frac{76 + 75,33}{2}$$

$$= 75,67$$

$$C3(x2) = \frac{73,67 + 71,33}{2}$$

$$= 72,50$$

$$C3(X3) = \frac{63 + 57,67}{2} = 60,34$$

$$C3(X4) = \frac{83 + 62}{2} = 72,50$$

$$C3(X5) = \frac{76 + 73,67}{2} = 74,84$$

Berdasarkan perhitungan di atas, maka diperoleh nilai *centroid* baru untuk iterasi keenam seperti pada Tabel 4.

Tabel 4. Nilai Centroid Baru untuk Iterasi VI

Centroid	MTK	FSK	KMA	BLG	TIK
C1	85,19	87,33	83,71	93,05	90,24
C2	80,36	79,45	76,95	85,48	82,59
C3	75,67	72,50	60,34	72,50	74,84

Untuk menghitung pusat *cluster* 1 dihitung dari jarak tiap data ke titik *centroid* menggunakan rumus (1):

Cluster 1

$$D_1 = \sqrt{(89,33 - 85,19)^2 + (88,67 - 87,33)^2 + (90 - 83,71)^2 + (94 - 93,05)^2 + (94 - 90,24)^2} = 8,57291$$

$$D_2 = \sqrt{(82 - 85,19)^2 + (81,33 - 87,33)^2 + (76,67 - 83,71)^2 + (86,67 - 93,05)^2 + (82,33 - 90,24)^2} = 14,10802$$

$$D_3 = \sqrt{(82,67 - 85,19)^2 + (78 - 87,33)^2 + (76,67 - 83,71)^2 + (87 - 93,05)^2 + (75 - 90,24)^2} = 20,2935$$

Cluster 2

$$D_1 = \sqrt{(89,33 - 80,36)^2 + (88,67 - 79,45)^2 + (90 - 76,95)^2 + (94 - 85,48)^2 + (94 - 82,59)^2} = 23,20344$$

$$D_2 = \sqrt{(82 - 80,36)^2 + (81,33 - 79,45)^2 + (76,67 - 76,95)^2 + (86,67 - 85,48)^2 + (82,33 - 82,59)^2} = 2,79297$$

$$D_3 = \sqrt{(82,67 - 80,36)^2 + (78 - 79,45)^2 + (76,67 - 76,95)^2 + (87 - 85,48)^2 + (75 - 82,59)^2} = 8,21765$$

Cluster 3

$$D_1 = \sqrt{(89,33 - 75,67)^2 + (88,67 - 72,5)^2 + (90 - 60,34)^2 + (94 - 72,5)^2 + (94 - 74,84)^2} = 46,45066$$

$$D_2 = \sqrt{(82 - 75,67)^2 + (81,33 - 72,5)^2 + (76,67 - 60,34)^2 + (86,67 - 72,5)^2 + (82,33 - 74,84)^2} = 25,33506$$

$$D_3 = \sqrt{(82,67 - 75,67)^2 + (78 - 72,5)^2 + (76,67 - 60,34)^2 + (87 - 72,5)^2 + (75 - 74,84)^2} = 23,58876$$

Untuk nilai D atau jarak masing-masing data ke *cluster* terdekat dapat dilihat pada Tabel 5.

Tabel 5. Pengelompokan Data ke *Cluster* Terdekat Iterasi I

Id	Jarak ke			C1	C2	C3
	C1	C2	C3			
1	8,572912	23,20344	46,45066	✓		
2	14,10802	2,792974	25,33506		✓	
3	20,2935	8,217654	23,58876		✓	
4	5,076144	17,52874	40,06257	✓		
5	12,34499	7,480031	29,02773		✓	
6	18,98612	6,216064	21,57275		✓	
7	4,274871	14,5621	36,9975	✓		
8	31,68525	17,22615	10,96283			✓
9	47,64405	33,07239	10,96283			✓
10	15,0469	7,195872	27,57411		✓	
11	21,34297	7,389662	18,89909		✓	
12	15,89158	8,906141	27,1958		✓	
13	6,86251	10,8221	33,94948	✓		
14	5,379301	20,20227	43,64148	✓		
15	18,93966	3,630332	21,00232		✓	
16	9,517655	8,177168	30,89419		✓	
17	3,518003	17,61545	40,24305	✓		
18	17,81319	4,804425	23,25396		✓	
19	18,85523	5,757507	22,78638		✓	
20	12,74845	9,094058	30,69351		✓	
21	8,361554	10,53825	32,94121	✓		
22	14,97855	4,348422	25,90393		✓	
23	23,6239	8,652559	16,95438		✓	
JUMLAH				7	14	2

Setelah melakukan perhitungan pada iterasi keenam maka didapatkan hasil pengelompokan yang sama dengan iterasi kelima atau tidak ada perubahan data pada kedua iterasi, anggota C1, C2 dan C3 dari *cluster* keenam yaitu:

- Anggota *cluster* 1 (C1) terdiri dari 7 siswa dengan nomor urut: (1, 4, 7, 13, 14, 17, 21)
- Anggota *cluster* 2 (C2) terdiri dari 14 siswa dengan nomor urut: (2, 3, 5, 6, 10, 11, 12, 15, 16, 18, 19, 20, 22, 23)
- Anggota *cluster* 3 (C3) terdiri dari 2 siswa dengan nomor urut: (8, 9)

e. Mengelompokkan Data ke Masing-masing *Cluster*
Setelah mendapatkan hasil pengelompokan pada iterasi keenam dan hasil yang didapatkan sudah sama dengan hasil di iterasi kelima, maka proses pencarian dihentikan. Hasil *cluster* yang didapatkan adalah pada *cluster* 1 (C1) terdapat 7 siswa yang Sangat Berkompeten, pada *cluster* 2 (C2) terdapat 14 siswa yang Berkompeten dan *cluster* 3 (C3) terdapat 2 siswa yang Kurang Berkompeten.

Kelompok data yang termasuk ke dalam *cluster* 1 (Sangat Berkompeten) dapat dilihat pada Tabel 6.

Tabel 6. Pengelompokan Data *Cluster* 1 (Sangat Berkompeten)

Id Siswa	Jarak ke C1	Jarak ke C2	Jarak ke C3	Jarak Terdekat C1
1	8,572912	23,20344	46,45066	✓
4	5,076144	17,52874	40,06257	✓
7	4,274871	14,5621	36,9975	✓
13	6,86251	10,8221	33,94948	✓
14	5,379301	20,20227	43,64148	✓
17	3,518003	17,61545	40,24305	✓
21	8,361554	10,53825	32,94121	✓
JUMLAH				7

Kelompok data yang termasuk ke dalam *cluster* 2 (Berkompeten) dapat dilihat pada Tabel 7.

Tabel 7. Pengelompokan Data ke *Cluster* 2 (Berkompeten)

Id Siswa	Jarak ke C1	Jarak ke C2	Jarak ke C3	Jarak Terdekat C2
2	14,10802	2,792974	25,33506	✓
3	20,2935	8,217654	23,58876	✓
5	12,34499	7,480031	29,02773	✓
6	18,98612	6,216064	21,57275	✓
10	15,0469	7,195872	27,57411	✓
11	21,34297	7,389662	18,89909	✓
12	15,89158	8,906141	27,1958	✓
15	18,93966	3,630332	21,00232	✓
16	9,517655	8,177168	30,89419	✓
18	17,81319	4,804425	23,25396	✓
19	18,85523	5,757507	22,78638	✓
20	12,74845	9,094058	30,69351	✓
22	14,97855	4,348422	25,90393	✓
23	23,6239	8,652559	16,95438	✓
JUMLAH				14

Kelompok data yang termasuk ke dalam *cluster* 3 (Kurang Berkompeten) dapat dilihat pada Tabel 8.

Tabel 8. Pengelompokan Data *Cluster* 3 (Kurang Berkompeten)

Id Siswa	Jarak ke C1	Jarak ke C2	Jarak ke C3	Jarak Terdekat C3
8	31,68525	17,22615	10,96283	✓
9	47,64405	33,07239	10,96283	✓
JUMLAH				2

4. Kesimpulan

Berdasarkan hasil *cluster* yang didapat, maka dapat disimpulkan bahwa pada C1 terdapat 12 siswa yang Sangat Berkompeten untuk mengikuti OSN, pada C2 terdapat 14 siswa yang Berkompeten namun belum bisa mengikuti OSN dan pada C3 2 siswa yang Kurang Berkompeten untuk mengikuti OSN.

Daftar Rujukan

- [1] Setiawan R. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta). *J. Lentera Ict*, 3(1), 76–92.
- [2] Heni S., & Irahm Gufroni, A. (2017). Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 03(02), 299–305. <https://doi.org/10.25077/TEKNOSI.v3i2.2017.299-305>

- [3] Maulida L. (2018). Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167. <https://doi.org/10.14421/jiska.2018.23-06>
- [4] Wicaksono A. E. (2016). Implementasi Data Mining Dalam Pengelompokan Peserta Didik di Sekolah untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus SMA N 6 Bekasi). *Jurusan Teknik Informatika, Universitas Gunadarma*, 21(3), 206–216. <http://dx.doi.org/10.35760/tr.2016.v21i3.1599>
- [5] Sindhu R., Nandal R., Dhamija P., & Sehwat H. K. (2017). A Review On K-Mean Algorithm And It's Different Distance Matrices. *International Journal of Engineering and Technology*, 9(2), 1423–1430. <https://doi.org/10.21817/ijet/2017/v9i2/170902227>
- [6] Jaroji J., Danuri D., & Putra Fajri P. (2016). K-Means Untuk Menentukan Calon Penerima Beasiswa Bidik Misi Di Polbeng. *INOVTEK Polbeng - Seri Informatika*, 1(1), 87–94. <https://doi.org/10.35314/ISL.V1I1.129>
- [7] Raval U. R., & Jani C. (2016). Implementing & Improvisation of K-means Clustering Algorithm. *International Journal of Computer Science and Mobile Computing*, 55(5), 191–203. Retrieved from <http://www.ijcsmc.com/docs/papers/May2016/V5I5201647.pdf>
- [8] Yunita (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri). *Jurnal Sistemasi*, 7(September), 238–249. <https://doi.org/10.32520/stmsi.v7i3.388>
- [9] Abdullah H. A. (2018). Improve the Performance of K-Means by Using Genetic Algorithm for Classification Heart Attack. *International Journal of Electrical and Computer Engineering (IJECE)*, 2088–8708, <https://doi.org/10.11591/ijece.v8i2.pp1256-1261>
- [10] Lubis H. A. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Berbasis Customer Relationship Management. *Jurnal Sistemasi*, 7(September), 238–249.
- [11] Purba W., Tamba S., & Saragih J. (2018). The effect of mining data k-means clustering toward students profile model drop out potential. *Journal of Physics: Conference Series*, 1007(1). <https://doi.org/10.1088/1742-6596/1007/1/012049>
- [12] Khade S., Pandhare P., Navale S., Patil K., & Gaikwad V. (2016). Fruit Quality Evaluation Using K-Means Clustering Approach. *International Journal of Advances in Science Engineering and Technology*, 4(2). 2321-9009.
- [13] Haribabu T., & Raju, I. (2019). An Optimal Approach of Initial Centroid Selection for Effective Clustering. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(5), 2278-3075
- [14] Ismi D. P., Panchoo, & S., Mrinto. (2016). K-Means Clustering Based Filter Feature Selection on High Dimensional Data. *International Journal of Advances in Intelligent Informatics*. 38-45, <http://dx.doi.org/10.12928/ijain.v2i1.54>
- [15] Metisen B. M., & Sari H. L. (2016). Analisis Clustering Menggunakan Metode K-Means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2). 1858-2680