



Artikel Penelitian

Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional

Ninik Tri Hartanti

Sistem Informasi, Universitas Amikom, Yogyakarta, 55288, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 17 Februari 2020

Revisi Akhir: 02 Juli 2020

Diterbitkan Online: 30 Agustus 2020

KATA KUNCI

K-Means,
Elbow,
Clustering,
Ujian Nasional

KORESPONDENSI

E-mail: ninik.t@amikom.ac.id

A B S T R A C T

Keberhasilan siswa dalam menempuh ujian nasional (UN) dapat terlihat dari perolehan nilai mata pelajaran yang diujikan, tiga diantaranya adalah nilai matematika, Bahasa Indonesia, dan Bahasa Inggris. Mengukur kesiapan siswa dalam menghadapi Ujian Nasional sangat diperlukan demi keberhasilan siswa. Untuk mengukur tingkat kesiapan siswa algoritma yang digunakan adalah K-Means. Penerapan metode Elbow untuk menentukan optimasi banyaknya *cluster* yang selanjutnya akan digunakan dalam perhitungan *clustering* dengan algoritma K-Means. Proses perhitungan K-Means diawali dengan penentuan jumlah *cluster*, dengan menggunakan metode Elbow. Penerapan algoritma K-Means adalah untuk membentuk kelompok-kelompok siswa berdasarkan nilai rata-rata ketiga mata pelajaran UN (matematika, Bahasa Indonesia, dan Bahasa Inggris). Penentuan jumlah optimal *cluster* menggunakan metode *Elbow*, yang menunjukkan banyaknya *cluster* optimal adalah 3, kemudian dihasilkan 3 *cluster* yang terdiri dari *cluster* kategori “Siap”, “Cukup Siap” dan “Tidak Siap”. Dengan masing-masing kategori terdiri dari 7 siswa, 30 siswa, dan 29 siswa. Tujuan dari penelitian ini adalah mengelompokkan skor nilai siswa SMK Syubbanul Wathon terhadap tiga mata pelajaran UN (matematika, bahasa Indonesia, bahasa Inggris), sehingga mempermudah pihak sekolah mengambil tindakan lebih lanjut demi keberhasilan siswa dan predikat sekolah.

1. PENDAHULUAN

Pendidikan di Indonesia khususnya pendidikan formal, terdiri atas pendidikan dasar, pendidikan menengah dan pendidikan tinggi, pada sistem penilaiannya yang mengacu pada skor nilai ujian masih diberlakukan di Indonesia. Hasil penilaian tersebut akan berpengaruh terhadap predikat sekolah dan minat masyarakat sebagai calon siswa. Salah satu proses penilaian ini adalah dengan adanya ujian nasional setiap tahunnya.

Dalam rangka untuk mempersiapkan siswanya menghadapi ujian nasional, SMK Syubbanul Wathon Magelang selama ini menerapkan sistem yang mengklasifikasikan nilai rapor siswanya dengan metode manual. Sistem tersebut dilakukan dengan cara wali kelas mengumpulkan data nilai siswa baik nilai rapor dan nilai harian, kemudian merekap nilai-nilai tersebut, sehingga terbentuk sebuah data nilai perkelas. Nilai perkelas tersebut yang akan menjadi panduan untuk menginformasikan keadaan siswa dalam kesiapan untuk mengikuti ujian nasional.

Sangat dibutuhkan sebuah sistem atau panduan yang memudahkan pihak sekolah dan berfungsi untuk membantu dalam mengambil keputusan tentang penilaian kesiapan siswa dalam mengikuti ujian nasional mendatang. Penelitian ini bertujuan untuk memberikan panduan dalam mengambil keputusan tentang penilaian kesiapan siswa menghadapi ujian nasional. Sehingga akan membantu pihak sekolah agar dapat memberikan tindakan kelas sehubungan dengan keputusan yang dihasilkan.

Teknik *clustering* adalah suatu metode dalam analisa data untuk menyelesaikan persoalan tentang pengelompokan data. Salah satu metode *clustering* adalah K-Means, yang mempunyai kemampuan dalam klasifikasi data yang cukup besar jumlahnya relatif cepat dan efisien. Kekurangannya, hasil *clustering* dengan menerapkan algoritma K-Means bergantung pada penentuan awal pusat *cluster*, sehingga hasil perhitungan *clustering* dengan metode K-Means akan baik jika penentuan pusat *cluster* tepat [1].

Clustering merupakan salah satu metode penugasan suatu data sebuah objek ke suatu himpunan bagian, sehingga objek yang berada di himpunan (*cluster*) yang sama akan serupa dalam beberapa hal. Selain itu *Clustering* merupakan metode pembelajaran tanpa pengawasan yang digunakan dalam berbagai

bidang termasuk penambangan data (data mining), *machine learning*, pengenalan pola, analisis gambar dan bioinformatika [2]. Metode *clustering* K-Means, merupakan algoritma yang cukup sederhana untuk mengklasifikasikan atau mengelompokkan banyak data dengan atribut tertentu ke dalam suatu kelompok (*cluster*) sebanyak K [3].

Penelitian dengan menerapkan metode Elbow terdapat dalam kasus Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means *Clustering*, studi kasus data yang diambil adalah data di Sekolah Tinggi Agama Hindu Negeri Tampung Penyang Palangka Raya. Pada penelitian tersebut, jumlah *cluster* terbaik adalah 3, dengan menguji nilai K=2 sampai dengan nilai K=8. Pengujian nilai K dilakukan sebanyak 4 kali, ujicoba1 dengan 54 data, ujicoba2 dengan 139 data, ujicoba3 dengan 261 data dan ujicoba4 dengan 54 data. Berdasarkan hasil keempat uji coba tersebut, hasil *Sum of Square Error* yang mengalami penurunan paling besar adalah pada K=3 dengan jumlah data yang berbeda.[4].

Metode Elbow dalam penelitian kali ini dilakukan sebanyak 5 kali ujicoba. Pengujian nilai K berawal dari K= 2 sampai dengan K= 5. Ujicoba1 dengan 25 data, Ujicoba2 dengan 30 data, dan Ujicoba3 dengan 35 data, Ujicoba4 dengan 45 data, dan Ujicoba5 dengan 50 data. Berdasarkan kelima ujicoba tersebut, dihasilkan *Sum of Square Error* yang mengalami penurunan paling besar adalah pada K=3 sehingga jumlah cluster terbaik adalah 3.

Selain itu, terdapat penelitian tentang penerapan algoritma K-Means yang digunakan untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. Pada penelitian tersebut, K-Means digunakan untuk menentukan *cluster* yang terbaik untuk pemilihan mahasiswa terbaik yang akan mengikuti lomba. Data pengujian yang digunakan memiliki 5 atribut yaitu nim mahasiswa, nilai mata kuliah algoritma dan pemrograman 1, nilai mata kuliah fisika dasar, nilai kalkulus 1, dan IPK mahasiswa, serta jumlah mahasiswa adalah 124. Berdasarkan perhitungan dengan Weka dihasilkan 4 kelompok dengan hasil bahwa cluster 0, dengan IPK = 0.51 sebanyak 9 mahasiswa (7%), cluster 1 dengan IPK = 3.41 sebanyak 28 mahasiswa (23%), cluster 2 mahasiswa dengan IPK = 3.30 sebanyak 40 mahasiswa (32%), dan untuk cluster 3 dengan IPK = 3.89, sebanyak 47 mahasiswa dari 124 Mahasiswa (38%). Sehingga diambil kesimpulan bahwa cluster 1 dengan IPK tertinggi yang bisa digunakan untuk memilih 5 Mahasiswa untuk bisa mengikuti lomba [5].

Penggunaan K-Means dalam penelitian kali ini, bertujuan untuk menentukan *cluster* kesiapan siswa dalam mengikuti Ujian nasional. Dengan memanfaatkan Microsoft Excel, pengujian dilakukan dengan menggunakan 3 atribut yaitu nilai Matematika, Bahasa Indonesia, dan Bahasa Inggris, dengan jumlah siswa adalah 66. Berdasarkan perhitungan *clustering* dengan menggunakan K-Means dihasilkan 3 *cluster* yaitu kategori “Siap” terdiri dari 7 siswa (14%), kategori “Cukup Siap” terdiri dari 30 siswa (60%) dan kategori “Tidak Siap” sebanyak 29 siswa (58%). Berdasarkan hasil

tersebut, maka prosentasi jumlah siswa yang siap untuk mengikuti ujian nasional adalah paling rendah dari kategori lainnya.

2. METODE

Penelitian ini menggunakan data nilai siswa Tahun Ajaran 2017-2018 dari kelas X (semester 1, 2), kelas XI (semester 3, 4) dan nilai siswa kelas XII (semester 5). Daftar nilai rapor yang digunakan adalah daftar nilai dari tiga mata pelajaran ujian nasional, yaitu Bahasa Indonesia, Bahasa Inggris, dan Matematika. Gambaran tentang metode penelitian ditampilkan pada gambar 1 di bawah ini.



Gambar 1. Alur Penelitian

2.1. Identifikasi Masalah

Identifikasi masalah dalam penelitian ini adalah bagaimana menentukan tingkat kesiapan siswa dalam menghadapi ujian nasional, sehingga pihak sekolah dapat memberlakukan langkah nyata berupa kegiatan yang dapat menambah tingkat kesiapan siswa.

2.2. Studi Literature

Penelitian yang diusulkan adalah mengenai panduan pengambilan keputusan dalam kesiapan sekolah, dengan menerapkan algoritma K-Means dengan disertai perhitungan optimasi *cluster* yang digunakan yaitu metode Elbow, sehingga peneliti melakukan studi *literature* dengan membaca tutorial yang membahas tentang penerapan data mining, paper di jurnal dan penelitian yang berhubungan dengan *clustering* sehingga dapat memperbanyak sumber-sumber penelitian yang berkaitan.

2.3. Pengumpulan Data

Proses pengumpulan data dari penelitian yang diusulkan ini dengan mengadakan tanya jawab langsung ke pihak yang berwenang sehubungan dengan kurikulum dan persiapan UN, yaitu dengan WaKaSek bagian kurikulum atau Kaprodi. Data selanjutnya adalah data nilai rapor siswa dari tiga mata pelajaran UN (Bahasa Indonesia, Bahasa Inggris, dan Matematika) dari kelas X (semester 1, 2), kelas XI (semester 3, 4) dan kelas XII (semester 5).

2.4. Proses Perhitungan Metode Elbow

Tujuan dari metode elbow adalah untuk memilih nilai k yang kecil dan masih memiliki nilai *withinss* yang rendah [6]. Penentuan optimal jumlah *cluster* pada penelitian ini menggunakan salah satu

metode analisis *cluster* yaitu metode *Elbow*, dengan memperhatikan nilai perbandingan (dari perhitungan SSE untuk setiap nilai *cluster*) antara jumlah *cluster* yang akan membentuk siku pada suatu titik, sehingga semakin besar jumlah *cluster* k maka nilai SSE akan semakin kecil. Rumus SSE sebagai berikut [7].

$$SSE = \sum_{k=1}^k \sum_{x_1 \in S_k} \|x_1 - C_k\|_2^2 \quad (1)$$

Berdasarkan persamaan (1) di atas, jumlah *cluster* adalah 3, kemudian mencari nilai jarak kedua objek terdekat dengan menggunakan persamaan (2). Adapun hasil perhitungan dengan menggunakan metode *Elbow* terlampir dalam tabel 1 di bawah ini.

Tabel 1. Detail Hasil Data Ujicoba

Uji Coba	jml data	Cluster			
		2	3	4	5
1	25	74.489108	86.521803	85.488156	70.921717
2	35	190.15696	140.6364	104.89381	143.40227
3	45	340.69025	273.27798	295.50029	260.18237
4	50	390.12757	549.89352	345.83671	249.01385
5	60	419.43084	861.28683	337.19983	357.142
6	66	604.90172	610.18096	13341.701	25505.241

Pada tabel 1 di atas, dijelaskan bahwa metode *Elbow* dilakukan ujicoba sebanyak 6 kali dengan input nilai $k=2$ sampai nilai $k=5$.

2.5. Proses Perhitungan K-Means

Berdasarkan detail hasil perhitungan *Elbow* di atas, langkah selanjutnya adalah melakukan perhitungan *clustering* dengan algoritma K-Means. Langkah kerja K-Means diantaranya adalah memposisikan setiap data objek ke *cluster* terdekat, yang ditentukan berdasarkan jarak kedua objek tersebut.

1. Menentukan kembali pusat *cluster* dengan anggota *cluster* yang baru.
2. Ulangi lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* sudah tidak berubah lagi, maka proses selesai.
3. Kembali ke langkah 3 sampai pusat *cluster* tidak mengalami perubahan lagi.

Berikut adalah rumus untuk perhitungan jarak terdekat:

$$d_{(x,y)} = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

2.6. Kesimpulan

Merunut langkah-langkah dalam proses perhitungan di atas, dapat diambil kesimpulan diantaranya adalah banyaknya *cluster* yang ditentukan, berapa banyak jumlah iterasi proses perhitungan K-Means, dan jumlah siswa dalam setiap *cluster* nya. Beberapa point kesimpulan di atas, diharapkan dapat membantu pihak sekolah untuk memberikan tindakan lebih lanjut sehubungan dengan prestasi sekolah yang akan diperoleh.

3. HASIL DAN PEMBAHASAN

Berdasarkan metode penelitian di atas, hasil dan pembahasan dalam penelitian ini diawali dari proses perhitungan dengan metode *Elbow* untuk menentukan jumlah *cluster* terbaik. Merujuk tabel 1 di atas, dapat diketahui hasilnya, bahwa ujicoba1 dengan data sebanyak 25 data siswa akan dihasilkan nilai SSE (*Sum of Square Error*) untuk *cluster* ke2 adalah 74.489108, *cluster* ke3 = 86.521803, *cluster* ke4 = 85.488156 dan *cluster* ke5 adalah 70.921717.

Adapun selisih dari perhitungan SSE untuk ujicoba1 adalah sebagai berikut:

Tabel 2. Selisih nilai SSE dari Ujicoba1

Cluster	SSE	Selisih
2	74.489108	74.489108
3	86.521803	12.032694
4	85.488156	1.033647
5	70.921717	14.566439
6	30.150.9650	30,080.0432

Tabel 2 menjelaskan bahwa selisih terbesar ada di *cluster* 3, sehingga hasil ujicoba1 menghasilkan *cluster* terbaik adalah 3. Pada ujicoba2 menggunakan 35 data dengan menghasilkan nilai SSE dan selisihnya seperti terlampir di tabel 3 berikut.

Tabel 3. Selisih nilai SSE dari Ujicoba2

Cluster	SSE	Selisih
2	190.156959	190.156959
3	140.636398	49.520561
4	104.893808	35.742591
5	143.402272	38.508464
6	118.289510	25.112762

Tabel 3 menjelaskan bahwa selisih terbesar juga terjadi di *cluster* 3, sehingga hasil ujicoba2 menghasilkan *cluster* terbaik adalah 3. Sedangkan selisih nilai SSE untuk ujicoba3 dengan menggunakan 45 data, terlampir dalam tabel 4 berikut.

Tabel 4. Selisih nilai SSE dari Ujicoba3

Cluster	SSE	Selisih
2	340.69	340.69
3	273.28	67.41
4	295.50	22.22
5	260.18	35.32

Tabel 4 menjelaskan bahwa selisih nilai SSE terbesar terjadi di *cluster* 3, sehingga hasil ujicoba3 dengan menggunakan 45 data menghasilkan *cluster* terbaik adalah 3. Sedangkan hasil perhitungan selisai nilai SSE untuk ujicoba4 dan ujicoba5 terlampir dalam tabel

5 dan tabel 6. Untuk ujicoba4 menggunakan 50 data dan ujicoba5 menggunakan 60 data.

Tabel 5. Selisih nilai SSE dari Ujicoba4

Cluster	SSE	Selisih
2	390.13	390.13
3	549.89	159.77
4	345.84	204.06
5	249.01	96.82

Tabel 6. Selisih nilai SSE dari Ujicoba5

Cluster	SSE	Selisih
2	419.43	419.43
3	861.29	441.86
4	337.20	524.09
5	357.14	19.94

Terlihat bahwa selisih nilai SSE terbesar dalam setiap ujicoba terdapat di *cluster* 3, sehingga optimasi *cluster* terbaik adalah 3.

Tabel 7. Daftar Nilai Siswa (66 siswa)

No	NIS	NAMA SISWA	RATA-RATA		
			IND	ING	MAT
1	02631	ACHSAN FUADI	79	81	81
2	05172	ADAM ACHMADSYAH	78	79	78
3	05178	AFIF NUR FAHRUDIN	81	79	83
...					
20	02893	M. SUBHAN ABIYU	76	76	76
21	05341	M. SYAFNI ALIF	78	76	77
22	05342	M. TANFIDZUL UMAM	79	79	81
...					
38	05232	DANIEL ADI P	77	79	77
39	05241	DIMAS RAGIL A DZIKY	80	81	82
40	02735	HIDAYATULLAH	80	86	88
41	05253	FAJAR SUBHAN	80	81	78
...					
50	02871	M. GALANG ROMADHONI	76	80	76
51	05348	M. IRVAN KHABIBI	79	82	78
...					
61	02996	SYAFA KHOIRUR R	80	82	84
62	05411	TAUFIQ NUR H	76	78	74
...					
66	02635	ADRIAN SAEFUL A	79	82	85

3.1. Proses Perhitungan K-Means

Langkah awal dalam perhitungan K-Means adalah menentukan jumlah *cluster* (k), dimana dalam penelitian ini jumlah *cluster* ditentukan sebanyak 3, berdasarkan proses perhitungan optimasi jumlah *cluster* dengan menggunakan metode Elbow di atas. Namun sebelumnya, berikut dilampirkan data nilai siswa akan digunakan dalam proses *clustering*.

Pada tabel 7 di atas merupakan data nilai siswa yang akan digunakan dalam proses *clustering* dengan menggunakan K-Means. Langkah awal dalam *clustering* adalah mencari nilai secara random sebagai nilai *centroid* awal, sehingga diperoleh data seperti dalam tabel 8 di bawah ini.

Tabel 8. Nilai untuk *Centroid* kesatu (awal)

	No	NIS	RATA-RATA		
			IND	ING	MAT
centroid 1	1	02631	78.68	81.23	81.20
centroid 2	22	05342	79.30	79.38	80.52
centroid 3	62	05411	75.78	77.66	73.80

Tabel 8 di atas merupakan data nilai untuk *centroid* awal. Dengan nilai *centroid* awal tersebut, langkah selanjutnya adalah menghitung jarak terdekat dalam setiap *cluster*. Perhitungan untuk langkah ini menggunakan persamaan (2), sehingga hasil perhitungan dengan menggunakan nilai *centroid* awal di atas terlampir dalam tabel 9 di bawah ini.

Tabel 9. Hasil Perhitungan Jarak Terpendek - Iterasi 1

No	NIS	Jarak Terpendek		
		dc1	dc2	dc3
1	02631	0.00	2.07	8.71
2	05172	4.55	3.23	4.65
3	05178	3.36	2.73	10.40
...				
44	02776	8.85	7.55	1.80
45	05277	8.56	7.03	3.05
...				
52	05347	2.10	2.83	10.44
53	05358	7.78	5.98	4.78
...				
62	05411	8.71	7.78	0.00
66	02635	4.23	5.36	12.73

Tabel 9 di atas merupakan hasil perhitungan K-Means dengan menerapkan persamaan (1), sehingga perhitungannya dapat dilihat adalah seperti berikut ini:

Untuk siswa dengan NIS 02631

$$\begin{aligned} &\text{Perhitungan jarak terdekat centroid awal untuk cluster 1 (dc1)} \\ &= \text{SQRT}((78.68-78.68)^2+(81.23-81.23)^2+(81.20-81.20)^2) \\ &= \text{SQRT}((0)^2+(0)^2+(0)^2) \\ &= 0.00 \end{aligned}$$

$$\begin{aligned} &\text{Perhitungan jarak terdekat centroid awal untuk cluster 2 (dc2)} \\ &= \text{SQRT}((78.68-79.30)^2+(81.23-79.38)^2+(81.20-80.52)^2) \\ &= \text{SQRT}((0.384)^2+(3.429)^2+(0.454)^2) \\ &= 2.07 \end{aligned}$$

$$\begin{aligned} &\text{Perhitungan jarak terdekat centroid awal untuk cluster 3 (dc3)} \\ &= \text{SQRT}((78.68-75.78)^2+(81.23-77.66)^2+(81.20-73.80)^2) \\ &= \text{SQRT}((8.448)^2+(12.78)^2+(54.71)^2) \\ &= 8.71 \end{aligned}$$

Untuk siswa dengan NIS 05172

$$\begin{aligned} &\text{Perhitungan jarak terdekat centroid awal untuk cluster 1 (dc1)} \\ &= \text{SQRT}((78.38-78.68)^2+(78.53-81.23)^2+(77.55-81.20)^2) \\ &= \text{SQRT}((0.091)^2+(7.322)^2+(13.32)^2) \\ &= 4.55 \end{aligned}$$

$$\begin{aligned} &\text{Perhitungan jarak terdekat centroid awal untuk cluster 2 (dc2)} \\ &= \text{SQRT}((78.38-79.30)^2+(78.53-79.38)^2+(77.55-80.52)^2) \\ &= \text{SQRT}((0.848)^2+(0.73)^2+(8.852)^2) \\ &= 3.23 \end{aligned}$$

$$\begin{aligned} &\text{Perhitungan jarak terdekat centroid awal untuk cluster 3 (dc3)} \\ &= \text{SQRT}((78.38-75.78)^2+(78.53-77.66)^2+(77.55-73.80)^2) \\ &= \text{SQRT}((6.787)^2+(0.755)^2+(14.04)^2) \\ &= 4.65 \end{aligned}$$

Setelah diketahui hasil perhitungan jarak terdekat seperti di atas, maka langkah selanjutnya adalah membandingkan nilai jarak tersebut mana yang terdekat diantara ketiganya. Dengan menggunakan Ms. Excel, penentuan jarak terdekat diantara dc1, dc2, dan dc3 adalah sebagai berikut.

Untuk isian data IND pada *cluster1*-Iterasi1

$$= \text{IF}(\text{dc1}<\text{dc2}, \text{IF}(\text{dc1}<\text{dc3}, 78.68, ""), "")$$

Sedangkan untuk isian data ING *cluster1*-Iterasi1

$$= \text{IF}(\text{dc1}<\text{dc2}, \text{IF}(\text{dc1}<\text{dc3}, 81.23, ""), "")$$

Dan untuk isian data Mat *cluster1*-Iterasi1

$$= \text{IF}(\text{dc1}<\text{dc2}, \text{IF}(\text{dc1}<\text{dc3}, 81.20, ""), "")$$

Perhitungan di atas akan terus digunakan untuk menentukan nilai terdekat dari dc1, dc2, dan dc3 dalam setiap *cluster*. Sehingga akan dihasilkan data seperti terlampir dalam tabel 10. Tabel 10 merupakan hasil perhitungan *cluster1* berdasarkan nilai *centroid* awal. Dapat diketahui hasilnya adalah terdapat 20 data dalam *cluster1* dengan jumlah nilai mata pelajaran Bahasa Indonesia (IND) adalah 1,573.50, Bahasa Inggris (ING) adalah 1,651.40, dan Matematika (MAT) adalah 1,652.62.

Tabel 10. Hasil Perhitungan *Cluster1*- Iterasi1

No.	NIS	CI		
		IND	ING	MAT
1	02631	78.68	81.23	81.20
7	05187	78.68	81.28	79.66
13	05280	76.95	81.79	79.03
26	05380	78.47	80.89	81.80
31	05165	82.77	82.52	85.21
32	05169	80.08	85.67	88.09
34	05192	64.51	84.20	83.23
35	05206	76.30	81.50	77.39
36	05222	83.60	85.42	91.38
39	05241	80.18	80.80	82.25
40	02735	79.61	85.89	87.77
42	02753	78.24	81.91	80.86
43	02771	80.22	82.60	78.90
51	05348	78.70	81.78	78.21
52	05347	79.67	80.73	82.98
55	05376	79.66	82.73	79.99
59	05389	78.35	81.74	81.56
60	02980	79.27	84.81	83.76
61	02996	80.27	82.24	83.98
66	02635	79.27	81.67	85.37
		78.67	82.57	82.63

Sehingga dapat ditentukan nilai *centroid* kedua berdasarkan hasil tersebut dengan perhitungan sebagai berikut:

$$= \text{jumlah nilai perkolom} / \text{banyaknya data}$$

Untuk kolom IND, nilainya adalah:

$$= 1,573.50 / 20$$

$$= 78.67$$

Untuk kolom ING, nilainya adalah:

$$= 1,651.40 / 20$$

$$= 82.57$$

Untuk kolom MAT, nilainya adalah:

$$= 1,652.62 / 20$$

$$= 82.63$$

Kemudian dilanjutkan dengan menentukan *Cluster2*- Iterasi1. Proses perhitungannya sama dengan langkah menentukan *cluster1* di atas. Sehingga akan didapatkan hasil perhitungan *cluster2* untuk iterasi1 dalam tabel 11 berikut.

Tabel 11. Hasil Perhitungan *Cluster2*- Iterasi1

No.	NIS	C1		
		IND	ING	MAT
2	05172	78.38	78.53	77.55
3	05178	80.60	79.07	82.90
8	05188	80.85	80.34	80.88
12	05226	77.04	79.77	78.37
16	05319	81.87	79.97	79.26
17	05322	78.83	80.29	80.00
19	05339	76.69	78.39	79.84
22	05342	79.30	79.38	80.52
23	05345	78.70	78.15	77.64
25	05372	77.88	79.30	78.32
27	05386	78.19	78.52	79.25
28	05401	80.28	80.48	79.68
30	03018	78.43	80.52	78.10
33	02632	78.58	80.79	78.36
41	05253	79.59	80.50	77.78
54	05363	79.81	81.05	78.45
56	05377	80.82	81.23	78.98
64	03016	78.01	79.61	77.27
65	05426	76.53	79.85	78.07
		78.97	79.78	79.01

Untuk *cluster2*-iterasi1 diperoleh data sebanyak 19 data, dan nilai rata-rata per mata pelajaran IND, ING, dan MAT adalah 78.97, 79.78 dan 79.01.

Terakhir, dihitung nilai rata-rata untuk *cluster3*-Iterasi1, dengan menggunakan langkah-langkah perhitungan di atas, maka diperoleh hasil sebagai berikut.

Tabel 12. Hasil Perhitungan *Cluster3*- Iterasi1

No.	NIS	C3		
		IND	ING	MAT
4	05182	75.43	78.92	75.87
5	05183	76.95	78.57	76.16
6	05186	76.66	74.53	75.40
9	05189	77.44	78.51	75.73
10	05196	74.94	75.11	76.28
11	05219	78.42	79.31	75.42
14	05311	75.66	75.64	74.10
15	05315	76.84	75.69	75.38
18	05331	75.42	79.29	76.85

20	02893	75.85	75.77	75.93
21	05341	77.84	75.53	77.14
24	05354	75.28	78.60	74.96
29	05407	76.35	76.04	77.10
37	05231	79.03	79.24	73.89
38	05232	76.52	79.08	76.82
44	02776	77.14	76.48	73.90
45	05277	77.23	75.35	75.15
46	05283	78.05	80.25	74.06
47	05294	76.73	77.21	74.93
48	02846	76.92	75.31	73.84
49	05329	77.49	76.36	73.82
50	02871	75.66	80.18	75.98
53	05358	79.59	75.64	75.87
57	05385	77.67	80.07	74.38
58	02958	77.59	79.56	76.43
62	05411	75.78	77.66	73.80
63	05417	77.09	79.27	76.11
		76.87	77.53	75.38

Untuk *cluster3*-Iterasi1 diperoleh data sebanyak 27 data, dan nilai rata-rata per mata pelajaran IND, ING, dan MAT adalah 76.87, 77.53 dan 75.38.

Setelah hasil perhitungan jarak dan jumlah siswa dalam per *cluster* diperoleh, langkah selanjutnya adalah menentukan *centroid* lagi berdasarkan nilai rata-rata pada iterasi 1 tersebut di atas. Penentuan nilai *centroid* yang kedua ini berdasarkan nilai-nilai yang berada dalam setiap *cluster* dibagi dengan banyaknya elemen dalam setiap *cluster* tersebut. Adapun nilai *centroid* yang kedua adalah sebagai berikut:

Tabel 13. Nilai *centroid* kedua

centroid 1	78,67	82,57	82,63
centroid 2	78,97	79,78	79,01
centroid 3	75,43	78,92	75,87

Berdasarkan perhitungan nilai *centroid* kedua tersebut di atas, perhitungan selanjutnya adalah menentukan jarak terdekat setiap objeknya dengan menggunakan persamaan (1), sehingga menghasilkan data sebagai berikut, terdapat 15 siswa dalam *cluster1*-iterasi2, 23 siswa dalam *cluster2*-iterasi2, dan 28 siswa dalam *cluster3*-iterasi2. Masih terdapat perubahan jumlah siswa dalam setiap *clusternya*, sehingga harus diproses kembali dengan menentukan nilai *centroid* ketiga. Adapun nilai *centroid* ketiga adalah sebagai berikut:

Tabel 14. Nilai *centroid* ketiga

centroid 1	79,92	82,49	83,94
centroid 2	78,87	79,99	78,70
centroid 3	76,31	77,97	75,72

Mengacu data dalam tabel 14 di atas, nilai dengan *centroid* ketiga menghasilkan data sebagai berikut, terdapat 12 siswa dalam *cluster1*-Iterasi3, 26 siswa dalam *cluster2*-Iterasi3, dan 28 siswa dalam *cluster3*-Iterasi3. Berdasarkan hasil tersebut, masih terdapat perbedaan dalam jumlah siswa dalam setiap *cluster*, sehingga harus diproses kembali untuk mendapatkan hasil jumlah siswa dalam setiap *cluster* adalah sama. Nilai *centroid* untuk yang keempat adalah sebagai berikut, sesuai dalam tabel 15.

Tabel 15. Nilai *centroid* keempat

centroid 1	80,18	82,62	84,75
centroid 2	78,74	80,44	78,98
centroid 3	76,43	77,76	75,66

Berdasarkan nilai *centroid* keempat di atas, dihasilkan data sebagai berikut terdapat 10 siswa dalam *cluster1*-Iterasi4, 29 siswa dalam *cluster2*-Iterasi4, dan 27 siswa dalam *cluster3*-Iterasi4. Sejauh ini sudah dihasilkan 4 iterasi atau 4 *centroid*, tetapi jumlah elemen dalam setiap *cluster* masih terdapat jumlah yang berbeda, sehingga perlu dilakukan perhitungan lagi seperti langkah-langkah di atas.

Setelah dilakukan perhitungan dengan langkah-langkah di atas, proses perhitungan K-Means dalam penelitian ini menghasilkan sebanyak 9 iterasi atau dengan kata lain diperlukan 9 nilai *centroid*, seperti yang terlampir dalam tabel 16 berikut.

Tabel 16. Perolehan Jumlah Elemen Dalam Setiap Cluster

Iterasi	Cluster		
	C1	C2	C3
1	20	19	27
2	15	23	28
3	12	26	28
4	10	29	27
5	9	29	28
6	8	29	29
7	7	30	29
8	7	30	29
9	7	30	29

Hasil perhitungan setiap iterasi proses K-Means seperti dalam tabel 16 di atas, diperoleh hasil bahwa pada proses iterasi ke-7 sampai dengan ke-9 tidak terdapat perubahan dalam jumlah elemennya untuk setiap *cluster*. Sehingga diperoleh data bahwa terdapat sejumlah 7 siswa dalam *cluster1*, kemudian 30 siswa dalam *cluster2*, dan sejumlah 29 siswa dalam *cluster3*. Hasil ini mengindikasikan bahwa proses perhitungan K-Means selesai,

karena sudah tidak ada lagi perubahan dalam jumlah elemen setiap *clusternya*.

4. KESIMPULAN

Metode *Elbow* dapat digunakan untuk mencari optimasi dalam penentuan jumlah cluster yang akan diterapkan dalam algoritma *clustering*, seperti K-Means.

Pada proses perhitungan yang telah dilakukan tentang penerapan *clustering* dengan menggunakan algoritma K-Means dan metode optimasi *cluster* yaitu metode *Elbow*, dapat diambil kesimpulan bahwa penerapan metode *Elbow* sesuai dengan penelitian ini menghasilkan jumlah *cluster* terbaik adalah 3, selanjutnya dengan menerapkan metode K-Means dihasilkan data bahwa jumlah siswa sebanyak 66 siswa, dengan proses perhitungan sebanyak 9 iterasi dihasilkan ada 3 *cluster* yaitu kategori “Siap”, “Cukup Siap”, dan “Belum Siap”, dengan hasil *cluster* “Siap” sebanyak 7 siswa, *cluster* “Cukup Siap” sebanyak 30 siswa, dan *cluster* “Belum Siap” sebanyak 29 siswa.

Hasil tersebut dapat menjadikan pedoman atau pengukuran kekuatan dalam hal jumlah siswa dalam setiap kategorinya untuk pihak sekolah, dengan memberikan tindakan kelas yang sesuai agar jumlah siswa dalam kategori “Siap” menjadi bertambah, dan dalam kategori “Tidak Siap” menjadi berkurang.

DAFTAR PUSTAKA

- [1] L. Rahmawati, S. Widya Sihwi, and E. Suryani, “Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret),” *J. Teknol. Inf. ITS* smart, vol. 3, no. 2, p. 66, 2016, doi: 10.20961/its.v3i2.654.
- [2] S. Leela, “Enhancing K-Means Clustering Algorithm,” *IJCST*, vol. 8491, no. 4, pp. 73–77, 2011.
- [3] E. Rivani, “Aplikasi K-Means Cluster untuk Pengelompokan Provinsi Berdasarkan Produksi Padi, Jagung, Kedelai, dan Kacang Hijau Tahun 2019,” *J. Mat Stat*, vol. 10, no. 2, pp. 122–134, 2010.
- [4] N. Putu, E. Merliana, and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means,” *Pros. Semin. Nas. MULTI DISIPLIN ILMU CALL Pap. UNISBANK*, pp. 978–979, 2015.
- [5] A. R. Asroni, “Penerapan Metode K-Means untuk Clustering Mahasiswa Berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus pada Jurusan Teknik Informatika UMM Magelang,” *J. Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015, doi: 10.1038/hdy.2009.180.
- [6] N. Ahmad, A. Mukharil, T. Informatika, F. Teknik, and U. K. Indonesia, “Jurnal Sistem Informasi (Journal of Information Systems). 2 / 12 (2016), 82-89 DOI : <http://dx.doi.org/10.21609/jsi.v12i2.481>,” *J. Sist. Inf. (Journal Inf. Syst.)*, vol. 12, pp. 82–89, 2016.
- [7] E. Muningsih and S. Kiswati, “Sistem Aplikasi Berbasis Optimasi Metode *Elbow* Untuk Penentuan Clustering Pelanggan,” *Joutica*, vol. 3, no. 1, p. 117, 2018, doi:

- 10.30736/jti.v3i1.196.
- [8] A. T. Rahman, Wiranto, and A. Rini, "Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)," *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 1, pp. 24–31, 2017, doi: 10.20961/ITS.V6I1.11296.
- [9] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no. no. juni, pp. 10–20, 2013.
- [10] K. H. Muchammad Ridho A., "Implementasi Algoritma K-Means Untuk Mengukur Tingkat Kepuasan Siswa Terhadap Proses Pembelajaran," *FTI Univ. Dian Nuswantoro*, vol. 2011, no. 1, p. 24, 2011.
- [11] Widiarina, "Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial," *J. Intell. Syst.*, vol. 1, no. 1, pp. 33–36, 2015.
- [12] R. Nainggolan, "OPTIMASI PERFORMA CLUSTER K-MEANS MENGGUNAKAN SSE," *METHOMIKA J. Manaj. Inform. Komputerisasi Akunt.*, vol. 2, no. 2, pp. 103–108, 2018.
- [13] N. T. Hartanti, K. Kusriani, and A. Amborowati, "Sistem Pendukung Keputusan Untuk Menentukan Program Keahlian di SMK Syubbanul Wathon Magelang," *Konf. Nas. Sist. Inform. 2015*, pp. 419–424, 2015.
- [14] B. M. Metisen and H. L. Sari, "Analisis Clustering Menggunakan Metode K-Means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.

BIODATA PENULIS



Ninik Tri Hartanti
 Staf pengajar Program Studi Sistem Informasi Universitas Amikom Yogyakarta.
 Penulis menempuh pendidikan S1 dan S2 di Universitas Amikom Yogyakarta.
 Tertarik pada penelitian yang bertema: *decision support system*, *education data mining*, dan *artificial intelligence*.

NOMENKLATUR

d = titik dokumen, x = data record, y = data pusat
 k = jml klaster, SSE = *Sum of Square Error*