



Metode *Elbow K-Means* dalam Implementasi Data Mining pada Pemetaan Penyebaran Guru SMK

Ninik Tri Hartanti*¹, Erni Seniwati², Rina Pramitasari³

*^{1,2}Program Studi Sistem Informasi, Universitas Amikom, Yogyakarta, Indonesia

³Program Studi Teknik Komputer, Universitas Amikom, Yogyakarta, Indonesia

*Email Penulis Korespondensi: ninik.t@amikom.ac.id

Abstrak

Kualitas dari sistem pendidikan akan selalu dievaluasi sehingga mutu pendidikan akan semakin baik. Salah satu langkah untuk mendukung mutu pendidikan adalah adanya faktor sarana prasarana yang memadai dan adanya peran guru dalam proses pembelajaran yang berkualitas. Pola penyebaran guru menjadi penting guna mendukung pemerataan dalam penempatan guru sesuai dengan kebutuhan untuk setiap daerah. Penelitian ini membahas tentang bagaimana pola penyebaran guru SMK di Magelang, berdasarkan jumlah sekolah, jumlah guru dan siswa SMK di Magelang. Metode data mining yang digunakan adalah K-means Clustering, dengan dikombinasikan dengan metode penentuan jumlah kluster yaitu metode Elbow. Melalui metode Elbow akan dihasilkan nilai Sum of Squared Error (SSE) yang berfungsi untuk menentukan jumlah kluster. Hasil penelitian adalah terdapat 3 kluster dalam penyebaran guru SMK di Magelang. Kluster terbanyak adalah kluster ke2 dengan 11 kecamatan yaitu kecamatan Grabag, Sawangan, Dukun, Tempuran, Srumbung, Candimulyo, Kajoran, Pakis, Ngluwar, Kaliangkrik, dan Ngablak. Kemudian kluster ke3 dengan kategori cukup yang terdapat 8 kecamatan, yaitu Mertoyudan, Salaman, Mungkid, Secang, Borobudur, Bandongan, Tegalrejo dan Windusari. Sedangkan kluster terkecil adalah kluster ke1 dengan 2 kecamatan, yaitu kecamatan Muntilan dan Salam.

Kata kunci—K-Means Clustering, Sum of Squared Error, metode Elbow

Abstract

The quality of the education system will always be evaluated so that the quality of education will get better. One step to support the quality of education is the existence of adequate infrastructure and the role of teachers in a quality learning process. The pattern of teacher distribution is important to support equality in teacher placement according to the needs of each region. This research discusses the pattern of distribution of vocational school teachers in Magelang, based on the number of schools, number of vocational school teachers and students in Magelang. The data mining method used is K-means Clustering, combined with the method for determining the number of clusters, namely the Elbow method. Using the Elbow method, a Sum of Squared Error (SSE) value will be generated which functions to determine the number of clusters. The results of the research are that there are 3 clusters in the distribution of vocational school teachers in Magelang. The largest cluster is the second cluster with 11 sub-districts,

namely Grabag, Sawangan, Dukun, Tempuran, Srumbung, Candimulyo, Kajoran, Pakis, Ngluwar, Kaliangkrik, and Ngablak sub-districts. Then the third cluster with the sufficient category contains 8 sub-districts, namely Mertoyudan, Salaman, Mungkid, Secang, Borobudur, Bandongan, Tegalrejo and Windusari. Meanwhile, the smallest cluster is the 1st cluster with 2 sub-districts, namely Muntilan and Salam sub-districts

Keywords— *K-Means Clustering, Sum of Squared Error, metode Elbow*

1. PENDAHULUAN

Sistem pendidikan dasar di Indonesia berjalan dari tingkat SD hingga SMP pada usia sekitar 7 hingga 13 tahun, kemudian dapat diteruskan ke tingkat menengah setelah lulus SMP, yaitu SMA. Evaluasi terus-menerus dilakukan terhadap kualitas sistem pendidikan untuk memastikan peningkatan mutunya. Untuk meningkatkan kualitas pendidikan, salah satu langkah yang dapat diambil adalah dengan memperhatikan sarana dan prasarana yang sesuai serta peran guru yang berkualitas dalam proses pembelajaran. Kualitas lulusan dari sistem pendidikan sangat tergantung pada kualitas pendidikan yang diberikan. Ketika mutu pendidikan rendah, Kemampuan untuk menciptakan tenaga kerja berkualitas juga rendah. Oleh karena itu, pemerintah terus mengkaji kebijakan-kebijakan pendidikan serta faktor-faktor yang dapat meningkatkan kualitas pendidikan[1]. Guru memiliki peran yang sangat krusial dalam menentukan kualitas pendidikan, selain dari sarana dan prasarana yang memadai. Inti masalah dalam dunia pendidikan sebagian besar terletak pada pendekatan konvensional dalam proses belajar mengajar, yang menitikberatkan pada interaksi tatap muka antara guru dan siswa. Meskipun menjadi sasaran yang mudah dan umum untuk ditingkatkan, sistem ini dianggap kurang efektif karena kurang sesuai dengan perkembangan zaman. Pertukaran informasi yang semakin cepat tidak sejalan dengan kecepatan yang ditawarkan oleh pendekatan konvensional ini. Lembaga pendidikan yang masih mengandalkan model tradisional ini cenderung lambat dalam memberikan informasi dan tertinggal dalam mengadopsi perkembangan teknologi informasi[2]. Elemen kunci dalam meningkatkan mutu pendidikan adalah kualitas dari profesi guru. Tanpa guru yang kompeten dan dikelola dengan efisien untuk menjadikan kinerja mereka profesional, maka mutu pendidikan tidak akan pernah tercapai [3]

Elemen berikutnya dalam peningkatan mutu pendidikan adalah sarana dan prasarana, dimana Pengelolaan sarana dan prasarana pendidikan bertujuan untuk menyediakan fasilitas dan pelayanan di sekolah, sehingga proses pendidikan di sekolah dapat terealisasi secara efektif dan efisien [4]. Disamping elemen-elemen di atas terdapat elemen lain yaitu siswa, dimana siswa untuk setiap kelulusannya diharapkan akan siap untuk melanjutkan pendidikan dan bagi siswa SMK khususnya diharapkan telah mempunyai *skill* atau keterampilan, kemampuan yang akan diterapkan di dunia kerja dan dunia industri. Terdapat komponen utama dalam proses belajar mengajar, yaitu guru, bahan pelajaran, dan siswa, saling berkaitan erat. Hubungan ini menjadikan ketiga komponen tersebut sebagai syarat penting dalam keberhasilan proses belajar mengajar.

Beberapa penelitian yang terkait adalah penelitian dengan data asli dikelompokkan terlebih dahulu menggunakan metode *clustering K-Means* sebelum diterapkan algoritma genetika. Adapun algoritma genetika yang digunakan adalah memanfaatkan operator *Ordered Crossover* (OX) dan operator mutasi *Partial Shuffle Mutation* (PSM). Hasil penelitian menunjukkan bahwa melakukan *clustering K-Means* sebelum proses optimasi akan menghasilkan data yang lebih baik jika dibandingkan dengan tanpa menggunakan *clustering* terlebih dahulu[5].

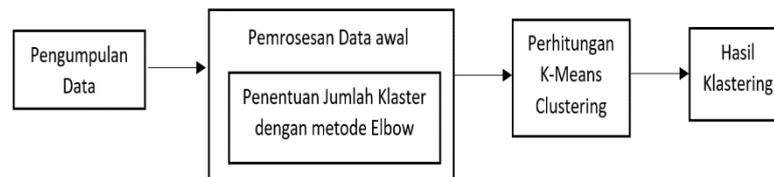
Masalah pendidikan di Indonesia masih terkait dengan isu-isu mendasar seperti kekurangan jumlah guru. Karena kurangnya upaya dalam meningkatkan mutu pendidikan, terjadi ketidakseimbangan dalam rasio guru dan murid yang sangat signifikan di beberapa provinsi. Dalam hal mengembangkan perangkat lunak untuk mengelompokkan data guru dengan menggunakan metode *K-Means Clustering*, dengan tujuan dapat untuk menganalisis data guru

dan distribusinya di Indonesia. Penelitian tersebut menunjukkan bahwa kesenjangan dalam kualitas dan kuantitas guru di Indonesia masih terjadi[6]. Merujuk hasil dari proses penentuan untuk jumlah kluster dengan menerapkan metode *Elbow* pada kasus untuk identifikasi kluster profil pelanggan terbaik mendapatkan jumlah kluster yang sama dalam data yang berbeda, sehingga jumlah kluster terbaik melalui metode *Elbow* menjadi data baku untuk karakteristik pada studi kasus. Proses perhitungan nilai *K-Means* menghasilkan kluster terbaik berdasarkan nilai *SSE* pada 500 kluster pengunjung batik. Hasilnya menunjukkan kluster mengalami penurunan tajam berada pada $K = 3$, sehingga K sebagai *cut-off point* merupakan kluster terbaik [7].

Salah satu kelebihan metode *Elbow*, dapat mudah untuk menentukan jumlah kluster, yaitu dengan melihat hasil kluster yang membentuk siku di titik tertentu untuk setiap klasternya [8]. Pemanfaatan *K-means clustering* dengan menggunakan metode *Elbow* untuk mengidentifikasi jumlah kluster yang optimal, menghasilkan jumlah kluster optimal yang ditemukan adalah tiga ($k = 3$). Kluster terbesar adalah Kluster 2, memiliki 16 guru, mayoritas anggota berlokasi di wilayah Jakarta Timur dan Barat. Kluster 0 terdiri dari 386 anggota, mayoritas berasal dari Jakarta Timur, dan mencakup 86 sekolah. Sekolah-sekolah dalam Kluster 0 mendapat manfaat dari dukungan badan siswa, sedangkan Kluster 2 memiliki jumlah guru yang paling banyak. Namun, analisis menunjukkan bahwa sekolah-sekolah dalam Kluster 1 memiliki pengalaman pendidikan yang lebih komprehensif karena memiliki rasio siswa-guru yang lebih baik[9].

2. METODE PENELITIAN

Merujuk judul penelitian ini, metode penelitian diawali dari pengumpulan, proses perhitungan dan diakhiri dengan penentuan jumlah kluster, sehingga pengelompokan data bisa diperoleh dapat dilihat pada Gambar 1.



Gambar 1 Alur Penelitian

2.1 Pengumpulan Data

Dalam penelitian ini data bersumber dari <https://dapo.kemdikbud.go.id/> yang meliputi data guru, data siswa, dan data sekolah yang berada di kabupaten Magelang. Pada tahap ini, data ditentukan sesuai dengan kebutuhan. Adapun data yang digunakan adalah data tahun ajaran 2022-2023 meliputi data siswa SMK di kabupaten Magelang, data guru SMK di kabupaten Magelang dan data sekolah SMK di kabupaten Magelang.

2.2 Pemrosesan Data Awal

Tahap pemrosesan data awal yaitu penentuan jumlah kluster dengan metode *Elbow*, yang akan memudahkan dalam proses *K-Means Clustering*. Adapun persamaan yang digunakan dalam metode *Elbow* seperti pada persamaan 1, dengan mencari nilai *Sum of Square Error* (*SSE*).

$$SSE = \sum_k^k = 1 \sum x \in S_k \|x_1 - C_k\|_2^2 \quad (1)$$

Dimana k adalah kluster ke $-c$, x_i merupakan jarak data objek ke- i dan C_k adalah pusat kluster ke- i . Berdasarkan proses perhitungan dengan menggunakan metode *Elbow* tersebut di atas, maka akan diperoleh jumlah kluster optimal untuk proses *K-Means Clustering*. Proses metode *Elbow* akan dihasilkan diagram yang berbentuk menyerupai lengan, sehingga titik koordinat yang membentuk menyerupai lengan tersebut yang akan digunakan sebagai jumlah klasternya [10].

2.3 Perhitungan K-Means Clustering

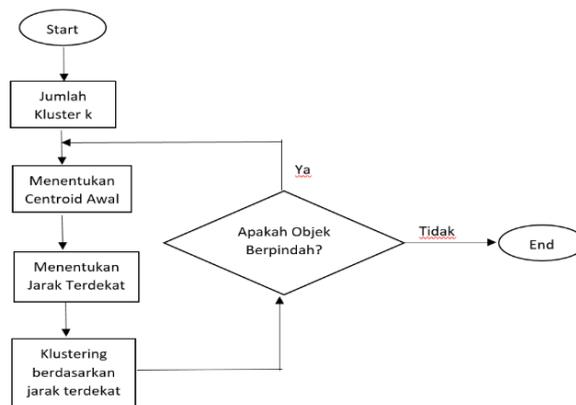
Berdasarkan hasil perhitungan *Elbow* sebelumnya, langkah berikutnya adalah melakukan perhitungan klastering menggunakan *K-Means*. Adapun tahapan dalam perhitungan *K-Means* adalah memposisikan setiap objek data ke dalam klaster terdekat, yang ditentukan berdasarkan jarak antara kedua objek tersebut.

- 1) Mulai dengan menentukan kembali pusat-pusat klaster dengan anggota-anggota klaster yang baru
- 2) Lakukan iterasi kembali untuk setiap objek dengan menggunakan pusat-pusat klaster yang baru. Proses ini diulang sampai pusat-pusat klaster tidak berubah lagi, menandakan konvergensi.
- 3) Ulangi proses tersebut di atas sampai tidak ada lagi perubahan pada pusat-pusat klaster.

Adapun persamaan *Euclidean Distance* untuk menghitung jarak terdekat adalah seperti pada formula (2).

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

Dengan persamaan 2 akan diperoleh hasil klastering sehingga data telah menjadi beberapa klaster [11]. Berikut *flowchart* dalam proses perhitungan dengan *K-Means* [12].



Gambar 2 *Flowchart* Proses K-Means

Berdasarkan Gambar 2, proses perhitungan dengan *K-Means* diawali dengan penentuan jumlah klaster kemudian menentukan titik pusat (*centroid*) awal dari data perhitungan *K-Means*. Penentuan titik pusat tersebut dilakukan secara random, sebagai titik awal untuk proses pengklasteran untuk iterasi awal. Selanjutnya proses perhitungan dilakukan sampai tidak ada data anggota kelompok yang berbeda untuk setiap klasternya.

2.4 Hasil Klastering

Klastering merupakan cara pengelompokan data pada *database* berdasarkan kesamaan atau keterkaitan antara data dengan menerapkan teknik observasi. Setiap algoritma mempunyai kekurangan dan kelebihan masing-masing, namun algoritma klastering berfungsi untuk mengelompokan data sesuai dengan karakteristik dan mengukur kemiripan jarak antar data dalam suatu kelompok [13].

Teknik kriteria siku atau metode *Elbow* merupakan metode untuk menentukan jumlah klaster data, digunakan untuk mendapatkan jumlah klaster yang optimal untuk sekumpulan titik data secara sederhana dan mudah diterapkan. Dengan menerapkan algoritma *K-Means clustering*, metode *Elbow* dapat menempatkan variasi data yang dijelaskan melalui jumlah klaster dan memilih kurva siku untuk mendapatkan jumlah klaster [14]. Di tahap ini, telah diperoleh hasil klastering sesuai proses perhitungan dengan metode *Elbow* dan *K-Means*.

3. HASIL DAN PEMBAHASAN

Tahap ini merupakan uraian hasil dan pembahasan dalam penelitian, dengan beberapa langkah.

3.1 Penentuan Data

Seperti telah disebutkan di atas, bahwa data yang akan digunakan adalah data publik yang terdapat pada link <https://dapo.kemdikbud.go.id/>. Tabel 1 memperlihatkan data tentang jumlah guru SMK berjenis kelamin laki –laki dan wanita di Magelang, di tahun ajaran 2022-2023.

Tabel 1 Data guru SMK

No.	Kecamatan	Jml	L	P
1.	Kec. Muntilan	172	86	86
2.	Kec. Mertoyudan	101	53	48
3.	Kec. Salaman	91	31	60
4.	Kec. Mungkid	74	50	24
5.	Kec. Grabag	20	12	8
6.	Kec. Salam	197	95	102
7.	Kec. Secang	74	44	30
8.	Kec. Borobudur	61	26	35
9.	Kec. Bandongan	43	19	24
10.	Kec. Sawangan	9	2	7
11.	Kec. Tegalrejo	53	30	23
12.	Kec. Dukun	19	5	14
13.	Kec. Tempuran	23	17	6
14.	Kec. Srumbung	0	0	0
15.	Kec. Candimulyo	9	2	7
16.	Kec. Kajoran	14	6	8
17.	Kec. Pakis	0	0	0
18.	Kec. Windusari	86	44	42
19.	Kec. Ngluwar	15	2	13
20.	Kec. Kaliangkrik	8	5	3
21.	Kec. Ngablak	45	17	28

Tabel 2 memperlihatkan data tentang jumlah siswa SMK berjenis kelamin laki –laki dan wanita di Magelang, di tahun ajaran 2022-2023.

Tabel 2 Data Siswa SMK

No.	Kecamatan	Jml	L	P
1.	Kec. Muntilan	3.203	1.927	1.276
2.	Kec. Mertoyudan	1.726	1.324	402
3.	Kec. Salaman	2.074	920	1.154
4.	Kec. Mungkid	1.500	1.311	189
5.	Kec. Grabag	525	427	98
6.	Kec. Salam	3.205	2.146	1.059
7.	Kec. Secang	1.348	930	418
8.	Kec. Borobudur	1.021	510	511
9.	Kec. Bandongan	1.249	775	474
10.	Kec. Sawangan	206	66	140

11.	Kec. Tegalrejo	1.352	772	580
12.	Kec. Dukun	253	117	136
13.	Kec. Tempuran	523	500	23
14.	Kec. Srumbung	0	0	0
15.	Kec. Candimulyo	92	67	25
16.	Kec. Kajoran	319	157	162
17.	Kec. Pakis	0	0	0
18.	Kec. Windusari	1.269	889	380
19.	Kec. Ngluwar	166	0	166
20.	Kec. Kaliangkrik	98	96	2
21.	Kec. Ngablak	827	465	362

Tabel 3 memperlihatkan jumlah sekolah SMK di Magelang yang berstatus SMK Negeri maupun SMK Swasta, di tahun ajaran 2022-2023. Ketiga tabel tersebut adalah data awal yang akan diolah, sehingga data awal akan menjadi seperti Tabel 4.

Tabel 3 Data Sekolah SMK

No.	Kecamatan	Jml	N	S
1.	Kec. Muntilan	9	0	9
2.	Kec. Mertoyudan	3	0	3
3.	Kec. Salaman	4	0	4
4.	Kec. Mungkid	2	0	2
5.	Kec. Grabag	1	0	1
6.	Kec. Salam	5	1	4
7.	Kec. Secang	4	0	4
8.	Kec. Borobudur	3	0	3
9.	Kec. Bandongan	2	0	2
10.	Kec. Sawangan	1	0	1
11.	Kec. Tegalrejo	2	0	2
12.	Kec. Dukun	1	0	1
13.	Kec. Tempuran	1	0	1
14.	Kec. Srumbung	0	0	0
15.	Kec. Candimulyo	1	0	1
16.	Kec. Kajoran	1	0	1
17.	Kec. Pakis	0	0	0
18.	Kec. Windusari	3	1	2
19.	Kec. Ngluwar	1	0	1
20.	Kec. Kaliangkrik	1	0	1
21.	Kec. Ngablak	1	1	0

Tabel 4 Data Awal Perhitungan

No.	Kecamatan	Guru	Siswa	Sekolah
1.	Kec. Muntilan	172	3203	9
2.	Kec. Mertoyudan	101	1726	3
3.	Kec. Salaman	91	2074	4
4.	Kec. Mungkid	74	1500	2

5.	Kec. Grabag	20	525	1
6.	Kec. Salam	197	3205	5
7.	Kec. Secang	74	1348	4
8.	Kec. Borobudur	61	1021	3
9.	Kec. Bandongan	43	1249	2
10.	Kec. Sawangan	9	206	1
11.	Kec. Tegalrejo	53	1352	2
12.	Kec. Dukun	19	253	1
13.	Kec. Tempuran	23	523	1
14.	Kec. Srumbung	0	0	0
15.	Kec. Candimulyo	9	92	1
16.	Kec. Kajoran	14	319	1
17.	Kec. Pakis	0	0	0
18.	Kec. Windusari	86	1269	3
19.	Kec. Ngluwar	15	166	1
20.	Kec. Kaliangkrik	8	98	1
21.	Kec. Ngablak	45	827	1

Merujuk Tabel 4, merupakan data awal untuk perhitungan selanjutnya yaitu proses penentuan jumlah kluster optimal dengan menerapkan metode *Elbow*, adapun persamaan yang digunakan adalah persamaan (1).

3.2 Penentuan Kluster dengan Metode Elbow

Dalam perhitungan metode Elbow ditentukan beberapa point untuk prosesnya, seperti jumlah kluster, banyaknya data uji, dan hasil setiap uji data tersebut, seperti pada gambar dan Tabel 5 dan Tabel 6.

Tabel 5 Data Uji1-Klaster1

No.	Kecamatan	Guru	Siswa	Sekolah
1.	Kec. Muntilan	172	3203	9
2.	Kec. Mertoyudan	101	1726	3
3.	Kec. Salaman	91	2074	4
...
11.	Kec. Tegalrejo	53	1352	2

Tabel 6 Data Uji1-Klaster2

No.	Kecamatan	Guru	Siswa	Sekolah
12.	Kec. Dukun	19	253	1
13.	Kec. Tempuran	23	523	1
14.	Kec. Srumbung	0	0	0
...
21.	Kec. Ngablak	45	827	1

Proses perhitungan dengan metode Elbow, diawali dengan data uji 1 yang terdiri dari 2 kluster, seperti pada Tabel 5 dan Tabel 6. Kluster pertama terdiri dari 11 data yaitu data ke1 sampai dengan data ke11, sedangkan kluster kedua terdiri dari 10 data terdiri dari data ke12 sampai data ke21. Dengan menggunakan persamaan (1) maka diperoleh data nilai SSE untuk kluster1= 279.664,7 dan nilai SSE untuk kluster2 = 1.009.355,1, sehingga jumlah nilai SSE = 1.289.019,8. Nilai SSE diperoleh dengan cara menentukan nilai rata-rata setiap kluster untuk setiap data ujinya, kemudian menentukan nilai minimal dalam setiap kluster untuk identifikasi kluster,

Tabel 7 Hasil Nilai Rata-rata Data Uji1-Klaster1

Data ke	Wilayah	Rata-rata
1.	Kec. Muntilan	1128,0
2.	Kec. Mertoyudan	610,0
...		...
11.	Kec. Tegalgrejo	469,0

Tabel 8 Hasil Nilai Rata-rata Data Uji1-Klaster2

Data ke	Wilayah	Rata-rata
12.	Kec. Dukun	91,0
13.	Kec. Tempuran	182,3
...		...
21.	Kec. Ngablak	291,0

Pada tabel 7 dan 8 merupakan hasil rata-rata dari 2 kluster untuk data uji1. Selanjutnya berdasarkan kedua tabel tersebut, akan diperoleh nilai maksimal dan nilai minimal. Untuk data uji1 kluster1 nilai maksimal = 1.135,7 nilai minimal = 72,0 sehingga nilai rata-rata data uji1 kluster 1 adalah 603,8. Selanjutnya untuk data uji1 kluster2, diperoleh nilai maksimal = 452,7 nilai minimal = 0,0 sehingga nilai rata-rata = 226,3. Langkah berikutnya, untuk menentukan nilai SSE adalah menghitung nilai selisih dari rata-rata dan kuadratnya sesuai kluster, sehingga diperoleh hasil pada Tabel 10 dan Tabel 11. Perhitungan berdasarkan tabel 8 data ke1 sampai dengan data ke11 adalah berikut ini:

Data ke 1 – rata-rata = $1.128,0 - 603,8 = 524,2$

Data ke 2 – rata-rata = $610,0 - 603,8 = 6,2$

Data ke 3 – rata-rata = $723,0 - 603,8 = 119,2$

Data ke 4 sampai dengan data ke 11, sedangkan untuk tabel 9 proses perhitungan dimulai dari data ke 12 sampai dengan data ke 21, seperti berikut:

Data ke 12 – rata-rata = $91,0 - 226,3 = -135,3$

Data ke 13 – rata-rata = $182,3 - 226,3 = -44,0$

Data ke 14 sampai dengan data ke 21.

Tabel 9 Nilai Selisih dan Kuadrat Data Uji1-Klaster1

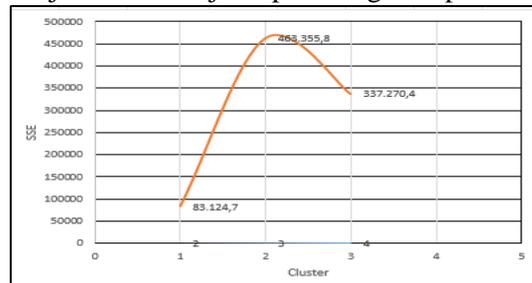
Data ke	Selisih	Kuadrat
1.	524,2	274.750,7
2.	6,2	38,0
...
11.	-134,8	18.180,0
Jumlah	-528,8	279664,7

Dari Tabel 9 dapat diketahui jumlah keseluruhan dari selisih adalah $-528,8$ dan total dari kuadrat adalah 279.664,7, sehingga nilai SSE yang merupakan nilai total kuadratnya adalah 279.664,7. Untuk menentukan jumlah kluster optimal, maka hasil nilai SSE dihitung selisihnya.

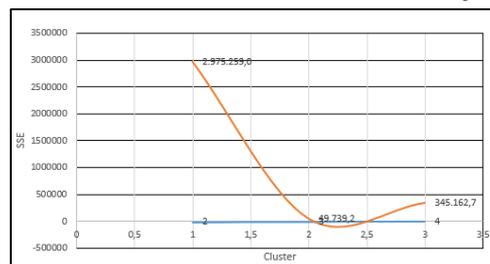
Tabel 10 Hasil Selisih Nilai SSE Data Uji 1

Jumlah Kluster	SSE	Selisih
2.	1.289.019,8	1.289.019,8
3.	1.000.105,4	-288.914,4
4.	628.733,6	371.371,8

Tabel 10 diketahui bahwa nilai selisih SSE di data uji1 untuk 2 kluster = 1.289.019,8 untuk 3 kluster = -288.914,4 dan untuk 4 kluster = 371.371,8. Berdasarkan hasil selisih SSE untuk semua data uji, dari data uji1, data uji2 dan data uji3 diperoleh grafik pada Gambar 3.



Gambar 3 Grafik Hasil Selisih Data Uji 2



Gambar 4 Grafik Hasil Selisih Data Uji 3

Sesuai pengertian dari *Elbow* itu sendiri, yaitu teknik yang digunakan untuk menghasilkan informasi dalam menentukan jumlah kluster terbaik dengan menganalisis perbandingan persentase variansi yang dijelaskan oleh jumlah kluster yang berbeda dan dilakukan dengan cara memeriksa titik di mana penambahan kluster tambahan memberikan penurunan yang tidak signifikan dalam penjelasan variasi, sering kali menyerupai "siku" pada grafik [15], maka jumlah kluster optimal adalah 3.

3.3 K-Means Clustering

Pada langkah ini, dengan menggunakan Tabel 4 sebagai data awal proses perhitungan *K-Means* dengan jumlah kluster (k) = 3 sesuai hasil uji dengan metode *Elbow* di atas. Langkah awal ditentukan terlebih dahulu nilai titik pusat (*centroid*) nya secara acak, misal klaster ke1, klaster ke2, dan klaster ke3 masing-masing diambil dari data ke3, data ke13, dan data ke19.

Dengan menerapkan perhitungan menentukan jarak terdekat berdasarkan persamaan (2) *Euclidean Distance*. Hasil perhitungannya terlampir di tabel 15. Perhitungan dengan menerapkan persamaan (2) sebagai berikut:

- Iterasi1- Klaster1
 $\text{SQRT} ((172-91)^2+(3203-91)^2+(9-91)^2) = 1131,913$
 $\text{SQRT} ((101-91)^2+(1726-91)^2+(3-91)^2) = 348,14508$
 $\text{SQRT} ((91-91)^2+(2074-91)^2+(4-91)^2) = 0$
- Iterasi1- Klaster2
 $\text{SQRT} ((172-23)^2+(3203-523)^2+(9-1)^2) = 2684,15$
 $\text{SQRT} ((101-23)^2+(1726-523)^2+(3-1)^2) = 1205,528$
 $\text{SQRT} ((91-23)^2+(1726-523)^2+(4-1)^2) = 1552,493$
- Iterasi1- Klaster3
 $\text{SQRT} ((172-15)^2+(3203-166)^2+(1-1)^2) = 3041,066$
 $\text{SQRT} ((101-15)^2+(1726-166)^2+(1-1)^2) = 1562,37$
 $\text{SQRT} ((91-15)^2+(1726-166)^2+(1-1)^2) = 1909,515$

Hasil di atas terlampir pada tabel 15 di bawah ini.

Tabel 11 Jarak Terdekat Iterasi1

Wilayah	Klaster1	Klaster2	Klaster3	Minimum	Klaster ke-
---------	----------	----------	----------	---------	-------------

Kec. Muntilan	1131,913	2684,1507	3041,0659	1131,913	1
Kec. Mertoyudan	348,14508	1205,5277	1562,37	348,1451	1
Kec. Salaman	0	1552,4928	1909,5154	0	1
...
Kec. Ngablak	1247,852	304,795	661,6804	304,795	2

Merujuk Tabel 11 tersebut, ditentukan nilai minimum untuk setiap data/wilayah, sehingga akan terbentuk klaster, pada data klastering iterasi 1 terdapat 7 data untuk klaster1, 6 data untuk klaster2 dan 8 data untuk klaster3. Selanjutnya menentukan nilai *centroid* baru untuk iterasi2, yaitu *centroid1* untuk klaster1= 108,8571, *centroid1* untuk klaster2 = 2058,286 *centroid1* untuk klaster3 = 4,142857, kemudian *centroid2* untuk klaster1 = 46,33333 *centroid2* untuk klaster2 = 902,3333 *centroid2* untuk klaster3 = 1,833333 dan *centroid3* untuk klaster1= 9,25 *centroid3* untuk klaster2= 141,75 *centroid3* untuk klaster3 = 0,75. Berdasarkan data centroid baru tersebut akan digunakan untuk proses perhitungan K-Means untuk iterasi2 dan selanjutnya, sehingga keseluruhan jumlah iterasi adalah 7, karena di saat iterasi6 dan 7 data per klaster tidak berubah, hasil iterasi proses K-Means seperti Tabel 12.

Tabel 12 Hasil Proses Iterasi Perhitungan K-Means

Iterasi ke-	Klaster1	Klaster2	Klaster3
1	7	6	8
2	5	8	8
3	4	7	10
4	2	17	2
5	2	13	6
6	2	11	8
7	2	11	8

Sehingga hasil akhir dari proses perhitungan K-means dilampirkan dalam Tabel 13.

Tabel 13 Hasil Akhir Perhitungan K-Means

No	Wilayah	Klaster1	Klaster2	Klaster3
1	Kec. Muntilan	1		
2	Kec. Mertoyudan			1
3	Kec. Salaman			1
4	Kec. Mungkid			1
5	Kec. Grabag		1	
6	Kec. Salam	1		
7	Kec. Secang			1
8	Kec. Borobudur			1
9	Kec. Bandongan			1
10	Kec. Sawangan		1	
11	Kec. Tegalrejo			1
12	Kec. Dukun		1	
13	Kec. Tempuran		1	
14	Kec. Srumbung		1	
15	Kec. Candimulyo		1	
16	Kec. Kajoran		1	
17	Kec. Pakis		1	
18	Kec. Windusari			1

19	Kec. Ngluwar		1	
20	Kec. Kaliangkrik		1	
21	Kec. Ngablak		1	

Pada Tabel 13 menunjukkan setiap kluster terdiri dari berbagai kecamatan, sehingga kluster dikategorikan menjadi kategori banyak, cukup dan kurang.

4. KESIMPULAN

Proses perhitungan untuk menyelesaikan kasus dalam penelitian ini menggunakan metode Elbow dan K-Means. Metode Elbow digunakan untuk menentukan jumlah kluster optimal, yang selanjutnya jumlah kluster tersebut akan diproses klustering dengan K-Means. Melalui perhitungan Elbow diperoleh jumlah optimal kluster adalah 3, dan dengan perhitungan K-Means diperoleh hasil sebagai berikut:

- Kluster 1 dengan kategori kurang terdiri dari 2 kecamatan, yaitu kecamatan Muntilan dan Salam.
- Kluster2 dengan kategori banyak atau lebih terdapat 11 kecamatan yaitu kecamatan Grabag, Sawangan, Dukun, Tempuran, Srumbung, Candimulyo, Kajoran, Pakis, Ngluwar, Kaliangkrik, dan Ngablak.
- Sedangkan kluster3 dengan kategori cukup terdapat 8 kecamatan, yaitu Mertoyudan, Salaman, Mungkid, Secang, Borobudur, Bandongan, Tegalrejo dan Windusari.

5. SARAN

Penelitian ini jauh dari kesempurnaan, penelitian selanjutnya bisa untuk dikembangkan adanya model klustering lainnya untuk implementasi data mining secara online atau *mobile*, dengan penambahan kriteria atau penggabungan kasus lainnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim Redaksi Jurnal Teknik Politeknik Negeri Sriwijaya yang telah memberi memberi kesempatan, sehingga artikel ilmiah ini dapat diterbitkan.

DAFTAR PUSTAKA

- [1] I. Rohmah Susiani and N. Diny Abadiah, "Kualitas Guru dalam Meningkatkan Mutu Pendidikan di Indonesia," *MODELING*, vol. 8, pp. 292–298, 2021.
- [2] Y. Pujilestari, D. Prodi, P. Pancasila, and D. Kewarganegaraan, "Dampak Positif Pembelajaran Online Dalam Sistem Pendidikan Indonesia Pasca Pandemi Covid-19," *ADALAH Buletin Hukum & Keadilan*, vol. 4, pp. 49–56, 2020, doi: 10.15408/adalah.v4i1.15394.
- [3] R. Diki Maulansyah, D. Febrianty, and M. Asbari, "Peran Guru dalam Peningkatan Mutu Pendidikan: Penting dan Genting!," *JOURNAL OF INFORMATION SYSTEMS AND MANAGEMENT*, vol. 02, no. 05, 2023, [Online]. Available: <https://jisma.org>
- [4] M. Parid, A. Laili, and S. Alif, "Pengelolaan Sarana dan Prasarana Pendidikan," *Jurnal Pendidikan dan Pemikiran Islam Tafhim Al'Ilmi*, vol. Februari, pp. 266–275, 2020.
- [5] H. Rosa and P. P. Pinaryanto, "The Teacher Placement using K-Means Clustering and Genetic Algorithm," in *ADI International Conference Series*, 2022, pp. 43–51.
- [6] F. Idris, F. Azmi, and P. S. Daru Kusuma, "Pengelompokan Data Guru di Indonesia Menggunakan K-Means Clustering Teacher Data Grouping in Indonesia Using K-Means Clustering," in *e-Proceeding of Engineering*, 2019, pp. 5648–5653.

-
- [7] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster." *IOP Conf Ser Mater Sci Eng*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
- [8] N. A. Maori, "METODE ELBOW DALAM OPTIMASI JUMLAH CLUSTER PADA K-MEANS CLUSTERING," *Jurnal SIMETRIS*, vol. 14, 2023.
- [9] V. A. Permadi, S. P. Tahalea, and R. P. Agusdin, "K-Means and Elbow Method for Cluster Analysis of Elementary School Data," *Progres Pendidikan*, vol. 4, no. 1, pp. 50–57, 2023, doi: 10.29303/prospek.v4i1.328.
- [10] N. T. Hartanti, "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 82–89, Aug. 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
- [11] N. T. Hartanti, "Mengukur Tingkat Pemahaman Mahasiswa pada Mata Kuliah Pemrograman dengan Algoritma K-Means Clustering Measuring Students ' Level of Understanding in Programming Courses with the K-Means Clustering Algorithm," *Jurnal Sisfotenika*, vol. 12, no. 1, pp. 62–73, 2022.
- [12] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, "COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm," *J Phys Conf Ser*, vol. 1783, no. 1, 2021, doi: 10.1088/1742-6596/1783/1/012027.
- [13] N. Syahfitri, E. Budianita, A. Nazir, and I. Afrianty, "KLIK: Kajian Ilmiah Informatika dan Komputer Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid," *Media Online*, vol. 4, no. 3, pp. 1668–1675, 2023, doi: 10.30865/klik.v4i3.1525.
- [14] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [15] S. Juanita and R. D. Cahyono, "K-Means Clustering With Comparison of Elbow and Silhouette Methods for Medicines Clustering Based on User Reviews," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 1, pp. 283–289, 2024, [Online]. Available: <https://doi.org/10.52436/1.jutif.2024.5.1.1349>