



Penerapan Data Mining Menggunakan K-Means Clustering Dalam Mengelompokkan Tingkat Kesulitan Mata Pelajaran

Fahrul Fauji, Lia Farokhah*

Fakultas Teknologi dan Desain, Teknik Informatika, Institut Teknologi dan Bisnis Asia, Malang
Jl. Soekarno Hatta-Rembuksari No.1 A, Mojolangu, Kec. Lowokwaru, Kota Malang, Jawa Timur, Indonesia

Email: ¹farul123654@gmail.com, ²*lia.farokhah@gmail.com

Email Penulis Korespondensi: lia.farokhah@gmail.com

Submitted: 10/02/2025; Accepted: 14/04/2025; Published: 15/04/2025

Abstrak—SMA Kertanegara Malang berkomitmen untuk memberikan dan meningkatkan kualitas pendidikan kepada siswanya. Salah satu tantangan utama yang sering dihadapi adalah adanya mata pelajaran tertentu yang dianggap sulit oleh siswa. Mata pelajaran yang sulit sering menjadi penghambat bagi siswa untuk mencapai nilai yang memuaskan. Salah satu pendekatan yang dapat dilakukan adalah Data Mining. Teknik analisis yang digunakan adalah algoritma K-Means metode Clustering. Metode Elbow membantu dalam menentukan jumlah cluster yang tepat untuk data yang telah diproses. Hasil dari metode Elbow pada penelitian ini adalah nilai cluster yang paling optimal yaitu $K=3$ berdasarkan nilai perolehan Within-Cluster Sum Of Square sebesar 104.5298167. Data yang digunakan adalah nilai rapor siswa kelas XII tahun angkatan 2021, 2022 dan 2023 berjumlah 163 data. Tujuan penelitian ini adalah untuk menentukan kelompok mata pelajaran apa saja yang dianggap sulit. Hasil penelitian setelah 163 data ditransformasi menjadi 30 data didapatkan 3 cluster optimal melalui metode Elbow, yaitu Cluster_1 dengan kategori pelajaran sulit berisi 8 mata pelajaran. Cluster_2 dengan kategori pelajaran tingkat kesulitan sedang berisi 10 mata pelajaran. Cluster_3 dengan kategori pelajaran mudah berisi 12 mata pelajaran. Hasil pengelompokkan ini dapat digunakan para guru di SMA Kertanegara Malang untuk memberikan pendampingan lebih kepada siswa, terutama pada mata pelajaran yang dikategorikan sulit.

Kata Kunci: Mata Pelajaran; Pengelompokkan; Data Mining; K-Means; Clustering

Abstract—SMA Kertanegara Malang is committed to providing and improving the quality of education to its students. One of the main challenges that is often faced is that there are certain subjects that students consider difficult. Difficult subjects often become an obstacle for students to achieve satisfactory grades. One approach that can be taken is Data Mining. The analysis technique used is the K-Means algorithm, Clustering method. The Elbow method helps in determining the right number of clusters for the data that has been processed. The results of the Elbow method in this research are the most optimal cluster value, namely $K=3$ based on the Within-Cluster Sum Of Square value of 104.5298167. The data used are the report cards of class The aim of this research is to determine what groups of subjects are considered difficult. The research results after 163 data were transformed into 30 data obtained 3 optimal clusters using the Elbow method, namely Cluster_1 with a difficult subject category containing 8 subjects. Cluster_2 with the medium difficulty level lesson category contains 10 subjects. Cluster_3 with the easy lesson category contains 12 subjects. The results of this grouping can be used by teachers at SMA Kertanegara Malang to provide more assistance to students, especially in subjects that are categorized as difficult.

Keywords: Eyesight; Grouping; Data Mining; K-Means; Clustering;

1. PENDAHULUAN

SMA Kertanegara Malang adalah salah satu lembaga pendidikan di Kota Malang yang beralamatkan di jalan Cengger Ayam No 1-5, Desa Tulusrejo, Kecamatan Lowokwaru, Kota Malang. SMA Kertanegara Malang berkomitmen untuk memberikan dan meningkatkan kualitas pendidikan kepada siswanya. Akan tetapi seperti halnya sekolah-sekolah lain di Kota Malang, SMA Kertanegara juga menghadapi berbagai tantangan dalam proses pembelajaran. Salah satu tantangan utama yang sering dihadapi adalah adanya mata pelajaran tertentu yang dianggap sulit oleh siswa. Kesulitan ini dapat mempengaruhi motivasi belajar siswa terhadap hasil pembelajaran yang dicapai. Mata pelajaran yang sulit sering menjadi penghambat bagi siswa untuk mencapai nilai yang memuaskan. Kesulitan belajar peserta didik dapat disebabkan oleh dua faktor, yaitu faktor internal dan faktor eksternal. Faktor internal adalah faktor yang berasal dari dalam diri peserta didik itu sendiri, misalnya kesehatan, motivasi, bakat, minat, kemampuan intelektual, dan kemampuan mengingat. Kemudian faktor eksternal adalah faktor yang datang dari luar peserta didik, misalnya faktor guru, faktor orang tua, faktor sarana prasarana, dan teman [1]. Identifikasi mata pelajaran yang sulit melalui nilai rapor perlu dilakukan agar SMA Kertanegara Malang dapat membuat strategi pembelajaran yang lebih baik terutama untuk siswa kelas XII.

Dalam menghadapi tantangan ini, teknologi dapat menjadi solusi yang tepat. Salah satu pendekatan yang dapat dilakukan adalah Data Mining. Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar [2]. Dengan menggunakan pendekatan Data Mining, penelitian ini dapat menganalisis data nilai rapor siswa untuk mengidentifikasi mata pelajaran yang sulit.

Salah satu teknik Data Mining yang dapat digunakan untuk pengelompokkan adalah clustering. Teknik clustering adalah suatu metode dalam analisa data untuk menyelesaikan persoalan tentang pengelompokkan data [3]. Sedangkan algoritma yang digunakan dalam metode clustering adalah K-Means. K-Means adalah metode



analisis data atau metode data mining yang melakukan pemodelan tanpa pengawasan yang mengklusterisasikan berdasarkan partisi dan melakukan clustering melalui proses iterasi berkelanjutan sampai dengan bertemu kondisi akhir, proses iterasi berhenti dan hasil clustering adalah output [4][5]. K-Means membagi data menjadi kelompok berdasarkan jarak ke pusat kluster [6]. Dalam penelitian ini algoritma K-Means dipilih untuk melakukan pengelompokan mata pelajaran yang sulit dikarenakan mudah diaplikasikan ke dalam data berjumlah kecil ataupun besar.

Penelitian yang berkaitan dengan clustering menentukan tingkat kesulitan mata pelajaran pernah dilakukan oleh Saputra dan Nataliani pada tahun 2021. Fokus masalah dalam penelitian tersebut adalah melakukan analisa penerapan Data Mining dalam menentukan kelompok siswa berprestasi. Metode penyelesaian yang digunakan adalah K-Means clustering dengan penerapan Data Mining. Hasil dari penelitian adalah pembentukan 3 kelompok cluster yaitu cluster 1 anggota kategori rendah sebanyak 12 anggota, cluster 2 merupakan anggota kategori cukup sebanyak 5 anggota dan cluster 3 merupakan kategori tinggi sebanyak 8 anggota. Hasil tersebut dapat dimanfaatkan sekolah untuk menganalisis prestasi dari murid juga bisa membantu guru maupun wali kelas dalam membentuk kelas yang ideal [7].

Penelitian serupa dilakukan oleh Silvia Ningsih pada tahun 2022. Fokus masalah dalam penelitian tersebut adalah untuk menentukan kelas kelompok bimbingan belajar tambahan dalam menghadapi Ujian Nasional. Metode yang digunakan adalah K-Means Clustering. Hasil dari penelitian tersebut adalah metode Clustering dengan algoritma K-Means dapat digunakan untuk mengelompokkan data siswa berdasarkan nilai ujian semester mata pelajaran Ujian Nasional, yaitu kemampuan siswa pintar, siswa sedang dan siswa kurang pintar. sehingga dapat mengetahui siswa yang mana saja yang akan diberi belajar tambahan agar dapat mencapai nilai standar kelulusan Ujian Nasional [8].

Penelitian lainnya dilakukan oleh Muhammad Azzam Al Fauzie ,Dkk pada tahun 2023. Fokus masalah dalam penelitian tersebut adalah untuk mencari metode pembelajaran yang bisa mengimbangi nilai akademik siswa dan ekstrakurikuler yang diikutinya dan mengetahui nilai rata-rata siswa yang aktif dalam mengikuti ekstrakurikuler. Metode yang digunakan adalah K-Means Clustering. Hasil dari penelitian tersebut adalah terbentuknya 8 kelompok cluster yaitu pintar (C2 dan C3), sedang (C5,C6 dan C7) dan cukup (C1, C4 dan C8), dalam hal ini nilai akademik siswa tecukup adalah Bahasa Inggris yang rata-rata nilai di bawah 80. Hasil tersebut dapat digunakan oleh pihak sekolah untuk mengetahui solusi bagi siswanya agar nilai rata-rata memenuhi standar kelulusan [9].

Penelitian berikutnya dilakukan oleh Asmana, Dkk pada tahun 2022. Masalah yang ada pada penelitian tersebut adalah SMK Wahidin Kota Cirebon masih sering menemui strategi yang tidak tepat dalam promosi penerimaan siswa dikarenakan tidak memiliki data induk dalam bentuk pengelompokan data asal sekolah siswa yang mendaftar di SMK Wahidin Kota Cirebon yang sudah dikaji secara ilmiah. Metode penyelesaian menggunakan K-Means Clustering. Hasil dari penelitian tersebut adalah algoritma K-Means dapat menganalisis clustering data asal calon siswa mencari informasi tentang penerimaan peserta didik baru di SMK Wahidin Kota Cirebon sebanyak 774 siswa [10].

Penelitian selanjutnya dilakukan oleh Vikki Darsono, Dkk pada tahun 2022. Metode yang digunakan adalah K-Means clustering dengan penerapan Data Mining. Masalah yang ada pada penelitian tersebut adalah kebingungan siswa dalam hal memilih bidang studi di perguruan tinggi. Hasil dari penelitian tersebut adalah terjadi 15 kali iterasi dengan cluster sebanyak 10 pada Cluster C1 sebanyak 32 siswa direkomendasikan masuk Bidang Keagamaan, C2 sebanyak 12 siswa direkomendasikan masuk Bidang Hukum, C3 sebanyak 5 siswa direkomendasikan masuk Bidang Sosial dan Politik, C4 sebanyak 10 siswa direkomendasikan masuk Bidang Komunikasi, C5 sebanyak 45 siswa direkomendasikan ke Bidang Ekonomi dan Bisnis, C6 sebanyak 44 siswa direkomendasikan masuk Bidang Pendidikan, C7 sebanyak 36 siswa direkomendasikan masuk Bidang Sains dan Teknologi, C8 sebanyak 39 siswa direkomendasikan masuk Bidang Seni dan Budaya, C9 sebanyak 9 siswa direkomendasikan masuk Bidang Teknik, C10 sebanyak 1 siswa direkomendasikan masuk Bidang Kedokteran dan Kesehatan [11]. Tujuan penelitian ini adalah untuk mengelompokkan mata pelajaran apa saja yang sering dianggap sulit oleh siswa dengan menggunakan teknik K-Means Clustering, melalui penentuan nilai cluster optimal menggunakan metode Elbow. Hasil akhir dari clustering adalah terbentuknya kelompok cluster yang berisi nama mata pelajaran dari tingkat kesulitan tinggi hingga mudah.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian adalah langkah-langkah yang diperlukan untuk memberikan pedoman dalam pengembangan penelitian ini. Tahapan ini menguraikan langkah-langkah yang diambil dalam penelitian ini. Langkah-langkah dari tahapan penelitian sebagai berikut:

a. Pengumpulan data

Data yang digunakan dalam penelitian ini diperoleh dari nilai rapor siswa kelas 12 angkatan tahun 2021, 2022 dan 2023. Data diambil langsung dari SMA Kertanegara Malang melalui proses perizinan dengan pihak sekolah.

- b. Preprocessing data
Data yang terkumpul akan dilakukan proses pengolahan data awal untuk membersihkan atribut atau nilai yang hilang agar dapat menghasilkan data yang lebih akurat. Proses pengolahan data awal yang digunakan pada penelitian ini adalah data transformation
- c. Metode Elbow
Metode elbow digunakan untuk menentukan jumlah k-cluster terbaik pada proses clustering melalui perhitungan SSE.
- d. Penerapan algoritma K-Means Clustering
Setelah mendapatkan nilai K optimal dari metode Elbow, selanjutnya data yang sudah melewati proses preprocessing akan dianalisis menggunakan algoritma K-Means clustering untuk melihat hasil masing-masing cluster.
- e. Evaluasi
Evaluasi dilakukan dengan menggunakan hasil clustering dari penentuan nilai K optimal yang diperoleh dari metode Elbow.

Gambar 1 menunjukkan alur proses tahap penelitian.



Gambar 1. Alur Tahapan Penelitian

2.2 Data Mining

Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Proses data mining terdiri dari pengumpulan data, ekstraksi data, analisa data, dan statistik data. Terdapat beberapa istilah lain yang memiliki makna sama dengan data mining, yaitu Knowledge discovery in databases (KDD), ekstraksi pengetahuan (knowledge extraction), analisa data atau pola (data atau pattern analysis), kecerdasan bisnis (business intelligence) dan data archaeology dan data dredging[12]. Data mining merupakan suatu proses yang dilakukan dengan cara mengekstraksi dasar data menjadi informasi yang mempunyai nilai penting, dengan data pola yang terbentuk dari data, data dasar disebut dengan Penemuan Pengetahuan dalam Database yang dicari menggunakan pembelajaran berbasis berulang untuk menemukan pola yang tetap dan mudah digunakan oleh pengguna dalam pengambilan keputusan [13].

Knowledge Discovery in Database (KDD) dan data mining seringkali menggunakan proses penggalian informasi tersembunyi secara bergantian untuk menjelaskan suatu basis data yang besar. KDD dan data mining kedua istilah tersebut memiliki konsep yang berbeda, namun saling berkaitan satu sama lain[14]. KDD akan membantu untuk mengidentifikasi dan mengamati hal-hal yang ingin diketahui, dan mengolah data menjadi informasi dan berkaitan satu sama lain. Proses KDD salah satunya adalah Data Mining. Berikut proses KDD secara garis besar[15]:

- a. Data selection
Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.
- b. Preprocessing
Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning dengan tujuan untuk membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.
- c. Transformation
Proses coding pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam database.
- d. Data mining
Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.
- e. Interpretation / evaluation
pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya atau tidak.

2.3 Algoritma K-Means

Algoritma K-Means merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah K cluster yang sudah ditetapkan di awal [16]. K-Means merupakan metode klasterisasi yang sering



digunakan diberbagai bidang karena penggunaannya sederhana, mudah untuk di implementasikan, mampu untuk mengklaster data yang besar. Algoritma K-Means merupakan metode berbasis jarak yang membagi data kedalam sejumlah cluster dan dalam setiap tahapan tertentu setiap objek harus masuk dalam kelompok, pada tahap selajutnya objek dapat berpindah ke kelompok lain [17]. Algoritma K-means merupakan algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k clustersehingga tingkat kemiripan antar anggota dalam suatu cluster tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah [18]. Langkah-langkah yang dilakukan dalam proses K-Means adalah sebagai berikut [19]:

- a. Menentukan nilai K, di mana nilai K merepresentasikan jumlah cluster.
- b. Menentukan titik secara acak sebagai centroid atau pusat cluster untuk inialisasi cluster awal.
- c. Menghitung jarak setiap titik dengan centroid. Rumus yang digunakan adalah Euclidean Distance.

$$d_{(a,b)} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \tag{1}$$

$d_{(a,b)}$ merupakan jarak antara objek a dan b. n merupakan dimensi data. a_i merupakan koordinat dari objek a pada dimensi data n. b_i merupakan koordinat dari objek b pada dimensi data n.

- d. Menetapkan setiap titik ke cluster dengan jarak terdekat.
- e. Menghitung rata-rata dari setiap cluster untuk menjadi centroid baru menggunakan rumus:

$$C_j = \frac{1}{|S_a|} \sum_{x_b \in S_a} x_b \tag{2}$$

C_j adalah centroid baru untuk a. S_a merupakan kumpulan data dalam cluster a. x_b adalah data dari cluster a. $|S_a|$ merupakan jumlah data dari cluster a.

- f. Mengulangi langkah ketiga hingga langkah kelima untuk titik tengah yang baru dari setiap cluster.

2.4 Metode Elbow

Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster (K) yang akan membentuk siku pada suatu titik [20]. Metode Elbow ini memberikan ide atau gagasan dengan cara memilih nilai cluster dan kemudian menambah nilai cluster tersebut untuk dijadikan model dalam penentuan cluster terbaik [21]. Cara untuk menggunakan metode Elbow adalah dengan mengevaluasi hasil pengujian K-means menggunakan SSE (Sum of Square Error. SSE banyak digunakan untuk mendapatkan jumlah cluster yang optimal. Metode ini menguji nilai SSE dengan jumlah cluster yang berbeda-beda kemudian mencari nilai SSE dengan selisih terbesar atau yang membentuk sudut paling siku pada grafik elbow dalam menentukan jumlah cluster terbaik [22]. Jika nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai cluster tersebut yang terbaik [23]. Rumus SSE sebagai berikut [24]:

$$SSE = \sum_{k=1}^K \sum_{x_1 \in S_k} |x_1 - C_k|^2 \tag{3}$$

K adalah Jumlah cluster dalam algoritma K-Means. S_k adalah Sekumpulan data yang termasuk dalam klaster ke-K. x_1 merupakan sebuah titik dalam cluster S_k . C_k merupakan centroid dari cluster K. $|x_1 - C_k|^2$ adalah Jarak Euclidean antara titik data x_1 dan centroid C_k . $\sum_{k=1}^K \sum_{x_1 \in S_k}$ digunakan untuk menjumlahkan semua jarak setiap cluster.

3. HASIL DAN PEMBAHASAN

3.1 Data Selection

Dalam proses ini melakukan data pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi

Tabel 1. Hasil Data Selection

No	Kelompok A (Umum)						Kelompok B (Umum)				
	SMT	PA	PPKN	Matematika	B. Ind	S.Ind	B. Ing	SB	PJOK	PKWU	B. Jawa
1	I	85	87	75	75	75	78	78	80	78	85
	II	85	88	75	75	75	78	78	80	78	90
	III	92	92	75	75	75	79	82	93	84	88
	IV	75	92	85	75	76	79	79	80	80	85
	V	88	92	85	76	78	79	90	80	85	84
	VI	89	93	88	84	85	85	90	88	92	93
..
163	I	82	83	81	82	80	80	85	83	83	82
	II	82	83	81	80	80	80	85	83	83	82
	III	85	87	80	84	83	78	84	83	80	81

No	SMT	PA	PPKN	Kelompok A (Umum)				Kelompok B (Umum)			
				Matematika	B. Ind	S.Ind	B.Ing	SB	PJOK	PKWU	B.Jawa
	IV	82	83	81	78	80	80	85	83	83	80
	V	82	83	81	82	80	80	85	83	83	84
	VI	82	83	81	78	80	80	85	83	83	84

Data yang digunakan merupakan data nilai rapor siswa tahun angkatan 2021, 2022 dan 2023 dengan total data sebanyak 163 data nilai dengan format excel, data bersifat primer karena diambil langsung dari SMA Kertanegara melalui perizinan pihak sekolah. Hasil dari data selection terdiri dari kolom keterangan semester berisi semester I sampai dengan VI, kolom mata pelajaran umum (A) berisi Pendidikan Agama, PPKN (Pendidikan Pancasila dan Kewarganegaraan), Matematika, Bahasa Indonesia, Sejarah Indonesia dan Bahasa Inggris. Sedangkan untuk kolom mata pelajaran umum (B) berisi Seni Budaya, PJOK (Pendidikan Jasmani Olahraga dan Kesehatan), PKWU (Prakarya dan Kewirausahaan) dan Bahasa Jawa. Hasil data selection pada Tabel 1.

3.2 Data Transformation

Data transformation adalah proses mengubah data dari satu format ke format lain untuk analisis dan pengambilan keputusan.

Tabel 2. Contoh Pemisahan Data Semester I dan II pada Angkatan Tahun 2021

No	Semester	PA	PPKN	Matematika	B.Ind	S.Ind	B.Ing	SB	PJOK	PKWU	B.Jawa
1	I	85	87	75	75	75	78	78	80	78	85
	II	85	88	75	75	75	78	78	80	78	90
2	I	84	84	85	83	85	84	90	88	92	80
	II	92	85	86	85	90	88	93	95	88	83
3	I	85	83	85	82	85	83	88	85	83	82
	II	89	89	89	85	87	88	89	85	88	82
4	I	90	90	91	92	91	90	89	90	89	87
	II	89	89	89	90	89	89	89	91	89	89
5	I	85	83	85	82	85	83	88	85	83	82
	II	89	89	89	85	87	88	89	85	88	82
..
71	I	85	83	85	82	85	83	88	85	83	82
	II	89	89	89	85	87	88	89	85	88	82

Pada data selection tidak ditemukan data yang rusak maupun data yang nilainya hilang , jadi proses berlanjut ke tahap data transformasi, yaitu mengubah format susunan data agar data dapat dianalisis. Tahap pertama transformasi data adalah memisahkan data pada Tabel 1 kemudian menambah sheet baru pada excel untuk melakukan pemisahan perdua semester, artinya nilai semester I dijadikan satu dengan nilai semester II. Proses pemisahan data juga dilakukan pada nilai semester dengan cara semester III dan IV dijadikan satu kemudian lanjut semester IV dan VI. Contoh pemisahan data pada semester I dan II angkatan tahun 2021, pada Tabel 2.

Setelah data perdua semester dipisahkan, langkah selanjutnya adalah menjumlahkan semester I dengan semester II kemudian dibagi dengan total semester yang ada yaitu 2, pada semester III, IV, V dan VI juga sama dijumlahkan kemudian dibagi perdua semester. Proses berikutnya adalah menjumlahkan semua total nilai rata-rata kemudian dibagi dengan total semua data yaitu untuk angkatan tahun 2021 dibagi dengan 71 total data, untuk tahun angkatan 2022 dibagi dengan 41 total data, untuk tahun angkatan 2023 dibagi dengan 51 total data. Contoh hasil penjumlahan total dibagi dengan semua data semester I dan II tahun angkatan 2021 pada tabel 3.

Tabel 3. Hasil Nilai Rata-Rata Tahun Angkatan 2021

No	PA	PPKN	Matematika	B.Ind	S.Ind	B.Ing	SB	PJOK	PKWU	B.Jawa
1	85	87.5	75	75	75	78	78	80	78	87.5
2	88	84.5	85.5	84	87.5	86	91.5	91.5	90	81.5
3	87	86	87	83.5	86	85.5	88.5	85	85.5	82
4	89.5	89.5	90	91	90	89.5	89	90.5	89	88
5	87	86	87	83.5	86	85.5	88.5	85	85.5	82
..
71	87	86	87	83.5	86	85.5	88.5	85	85.5	82
Total	87.49	86.02	85.54	81.70	84.58	84.42	86.96	85.41	84.37	82.70

Langkah terakhir adalah membuat tempat penyimpanan hasil nilai rata-rata tabel 3 di excel dengan nama file Data_Cluster, karena penelitian ini mengklusterkan mata pelajaran yang sulit maka kolom pertama akan diberi nama mata pelajaran, kemudian kolom 2, 3 dan 4 diberi nama sesuai dengan angkatan masing-masing, nilai masing-masing kolom berasal dari baris dengan nama total yang ada pada Tabel 3. Hasil transformasi data pada Tabel 4.

Tabel 4. Hasil Transformasi Data_Cluster

No	Mata Pelajaran	2021	2022	2023
1	P.A 10	87.49	87.57	83.82
2	P.A 11	88.99	88.78	84.89
3	P.A 12	90.06	89.89	84.68
4	PPKN 10	86.02	86.07	84.62
5	PPKN 11	91.10	91.13	85.70
6	PPKN 12	91.17	91.02	84.93
7	B.Ind 10	85.54	85.41	81.80
8	B.Ind 11	87.27	87.06	81.72
9	B.Ind 12	86.90	86.70	84.35
10	Matematika 10	81.70	81.51	79.40
11	Matematika 11	82.28	82.02	79.22
12	Matematika 12	83.51	83.16	81.45
13	S.Ind 10	84.58	84.41	79.98
14	S.Ind 11	86.73	86.52	81.25
15	S.Ind 12	86.85	86.61	81.53
16	B.Ing 10	84.42	84.26	80.24
17	B.Ing 11	86.54	86.37	79.94
18	B.Ing 12	85.95	85.66	82.16
19	S.B 10	86.96	86.79	84.47
20	S.B 11	87.78	87.29	84.69
21	S.B 12	88.20	87.34	85.89
22	PJOK 10	85.41	85.43	83.56
23	PJOK 11	88.63	88.39	84.06
24	PJOK 12	89.23	88.71	84.02
25	PKWU 10	84.37	84.22	83.41
26	PKWU 11	85.89	85.70	82.40
27	PKWU 12	88.73	88.56	84.26
28	B.Jawa 10	82.7	82.60	82.22
29	B.Jawa 11	82.87	82.67	81.49
30	B.Jawa 12	83.13	83.01	83.70

Kode angka 10, 11 dan 12 yang berada pada kolom mapel menandakan bahwa pelajaran tersebut memiliki kelas masing-masing, seperti untuk kode nomor 10 berarti pelajaran tersebut berlaku untuk siswa kelas 10 angkatan tahun 2021, 2022 dan 2023, begitu juga dengan nomor kode 11 dan 12. Nilai yang ada pada setiap baris di kolom angkatan berasal dari hasil perhitungan rata-rata penjumlahan semua nilai perdua semester dibagi total semua data, yang ada pada Tabel 3.

3.3 Menentukan Nilai K-Optimal Menggunakan Metode Elbow

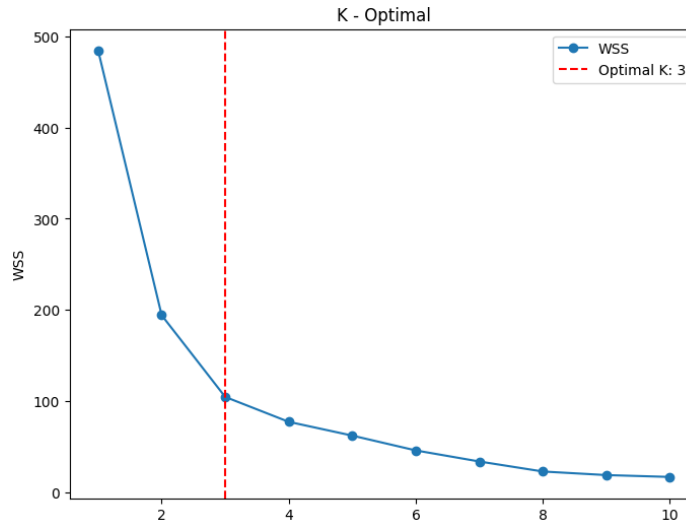
Perhitungan manual menggunakan uji coba jumlah K=2 sampai K=10 dengan penerapan metode Elbow menggunakan perhitungan nilai SSE (Sum of Square Error) pada tiap kluster

Tabel 5. WSS Pada Data_Cluster

Cluster	WSS
1	484.7087633
2	194.1479375
3	104.5298167
4	77.18052381
5	62.08227143
6	45.81875476
7	33.67353048
8	22.68428333
9	18.76693
10	16.718455

Pada dataset yang telah diolah dengan nama Data_Cluster, nilai WSS dapat diamati pada Tabel 5 menunjukkan terjadinya penurunan yang tinggi dari K = 1 ke K = 2, kemudian K= 2 ke K = 3. Akan tetapi, setelah K = 3, penurunan menjadi lebih kecil dan lebih stabil. K= 3 dianggap optimal karena setelah itu penurunan WSS tidak lagi signifikan. Tabel WSS digunakan untuk mengukur seberapa dekat data dalam setiap cluster terhadap centroidnya, semakin kecil nilai WSS maka semakin baik data tersebut terpusat pada centroidnya. Bisa dilihat

pada gambar 2 grafik K ditunjukkan oleh garis merah putus-putus, maka bisa dipastikan K paling optimal berada pada angka 3.



Gambar 2. Grafik K Optimal

3.4 Perhitungan Menggunakan Metode K-Means

Selanjutnya, proses dilakukan dengan melakukan perhitungan menggunakan Metode K-Means. Adapun hasil dari proses dapat dilihat berikut:

Tabel 6. Centroid Awal

	Centroid awal		
C1	82.7	82.60	82.22
C2	84.42	84.26	80.24
C3	89.23	88.71	84.02

Tahap selanjutnya adalah inisiasi centroid awal untuk perhitungan iterasi-1, nilai centroid awal pada Tabel 6. Penentuan titik awal centroid awal sangat berpengaruh pada hasil akhir proses clustering, apabila nilai awal centroid ditentukan secara acak atau asal maka hasil dari perhitungan akan berbeda-beda saat algoritma dijalankan. Metode K-Means Clustering digunakan untuk meminimalisir kesalahan perhitungan terhadap pemilihan centroid, caranya adalah dengan memilih titik acak sebagai centroid awal, kemudian untuk setiap titik data lainnya dihitung menggunakan rumus jarak Euclidean, setelah diterapkan rumus jarak Euclidean pilih centroid berikutnya berdasarkan probabilitas yang lebih tinggi untuk titik yang lebih jauh, kemudian ulangi proses hingga menghasilkan K centroid. Setelah mengetahui Nilai K berdasarkan metode Elbow dan inisiasi nilai centroid awal, kemudian menghitung jarak setiap data terhadap centroid menggunakan rumus Euclidean pada persamaan 1. Hasil perhitungan jarak Euclidean iterasi-1 pada Tabel 7.

Tabel 7. Jarak setiap Data Pada Iterasi-1

Data Ke-i	C1	C2	C3	Cluster
1	7.085548673	5.761718	2.089785	3
2	9.213327303	7.933335	0.905207	3
3	10.64731422	9.122505	1.586474	3
4	5.368733556	5.00205	4.19925	3
5	12.46720899	11.02864	3.48937	3
..
30	1.594804063	3.898487	8.354783	1

Setelah dilakukan perhitungan rata-rata nilai centroid, pada iterasi 2, 3, 4, 5 dan 6 rata-rata centroid masih berubah, baru pada iterasi-7 nilai rata-rata centroid tidak berubah menandakan bahwa proses clustering berhenti pada iterasi-7. Berikut adalah hasil dari jarak setiap data pada iterasi-7, bisa dilihat pada Tabel 8.

Tabel 8. Jarak setiap Data Pada Iterasi-7

Data Ke-i	C1	C2	C3	Cluster
1	6.950538311	2.338728	2.272572	3
2	9.136524657	4.526662	0.260891	3
3	10.4727937	5.82986	1.376504	3
4	5.596839091	2.027552	4.190012	2
5	12.34896384	7.710225	3.158817	3
..
30	2.738116654	4.456417	8.449868	1

Hasil Akhir nilai centroid pada iterasi-7, pada Tabel 9.

Tabel 9. Nilai Akhir Centroid Iterasi-7

Centroid akhir			
C1	83.14875	84.02286	80.9625
C2	86.84833	86.045	82.60083
C3	89.138	88.868	84.694

Berikut adalah hasil K-Means Clustering disertai dengan penentuan kelompok Cluster bisa dilihat pada Tabel 10.

Tabel 10. Hasil K-Means Clustering

No	Mata Pelajaran	2021	2022	2023	Cluster
1	P.A 10	87.49	87.57	83.82	3
2	P.A 11	88.99	88.78	84.89	3
3	P.A 12	90.06	89.89	84.68	3
4	PPKN 10	86.02	86.07	84.62	2
5	PPKN 11	91.10	91.13	85.70	3
6	PPKN 12	91.17	91.02	84.93	3
7	B.Ind 10	85.54	85.41	81.80	2
8	B.Ind 11	87.27	87.06	81.72	2
9	B.Ind 12	86.90	86.70	84.35	2
10	Matematika 10	81.70	81.51	79.40	1
11	Matematika 11	82.28	82.02	79.22	1
12	Matematika 12	83.51	83.16	81.45	1
13	S.Ind 10	84.58	84.41	79.98	1
14	S.Ind 11	86.73	86.52	81.25	2
15	S.Ind 12	86.85	86.61	81.53	2
16	B.Ing 10	84.42	84.26	80.24	1
17	B.Ing 11	86.54	86.37	79.94	2
18	B.Ing 12	85.95	85.66	82.16	2
19	S.B 10	86.96	86.79	84.47	2
20	S.B 11	87.78	87.29	84.69	3
21	S.B 12	88.20	87.34	85.89	3
22	PJOK 10	85.41	85.43	83.56	2
23	PJOK 11	88.63	88.39	84.06	3
24	PJOK 12	89.23	88.71	84.02	3
25	PKWU 10	84.37	84.22	83.41	2
26	PKWU 11	85.89	85.70	82.40	2
27	PKWU 12	88.73	88.56	84.26	3
28	B.Jawa 10	82.7	82.60	82.22	1
29	B.Jawa 11	82.87	82.67	81.49	1
30	B.Jawa 12	83.13	83.01	83.70	1

Hasil pengelompokkan data menggunakan metode K-Means Clustering berhasil mendapatkan informasi, dengan data yang sudah melewati proses transformasi yaitu sebanyak 30 data. Diketahui bahwa nilai tertinggi masuk ke dalam Cluster_3 dengan kategori mata pelajaran paling mudah, nilai sedang masuk ke dalam Cluster_2 dengan kategori mata pelajaran cukup mudah dan nilai terendah masuk ke dalam Cluster_1 dengan kategori mata pelajaran paling sulit.

3.5 Evaluasi

Berdasarkan hasil perhitungan menggunakan Elbow method, nilai optimal K yang diperoleh adalah 3. Pemilihan nilai K = 3 sebagai jumlah cluster yang optimal dilakukan berdasarkan hasil analisis grafik WSS, yang



menunjukkan bahwa setelah $K = 3$, penurunan nilai WSS tidak lagi signifikan, sehingga nilai $K = 3$ dipilih untuk penelitian ini.

4. KESIMPULAN

Berdasarkan hasil pembahasan dan perhitungan pada penelitian mengenai penerapan Data Mining menggunakan algoritma K-Means Clustering, maka dapat diambil kesimpulan pada penelitian ini menggunakan data siswa-siswi yang diambil dari SMA Kertanegara Malang, Jalan Cengger Ayam No 1-5, Desa Tulusrejo, Kecamatan Lowokwaru, Kota Malang, tahun angkatan 2021 sampai dengan tahun angkatan 2023 kelas XII. Data awal berjumlah 163 data, akan tetapi setelah melalui proses pengolahan dan transformasi data, menjadi 30 data dengan 3 atribut yaitu 2021, 2022 dan 2023, serta kolom mata pelajaran. Penentuan K optimal menggunakan metode Elbow terbukti mampu mengelompokkan cluster dengan baik dengan hasil nilai WSS 104.5298167, hasil akhir dari clustering adalah terbentuknya 3 kelompok cluster yaitu Cluster_1 (mata pelajaran paling sulit) berisi Matematika 10, Matematika 11, Matematika 12, Sejarah Indonesia 10, Bahasa Inggris 10, Bahasa Jawa 10, Bahasa Jawa 11 dan Bahasa Jawa 12. Cluster_2 (mata pelajaran kesulitan sedang) berisi Pendidikan agama 10, Pendidikan agama 11, Pendidikan agama 12, PPKN 11, PPKN 12, Seni budaya 11, Seni Budaya 12, PJOK 11, PJOK 12 dan PKWU 12. Cluster_3 (mata pelajaran tingkat paling mudah) berisi PPKN 10, Bahasa Indonesia 10, Bahasa Indonesia 11, Bahasa Indonesia 12, Sejarah Indonesia 11, Sejarah Indonesia 12, Bahasa Inggris 11, Bahasa Inggris 12, Seni budaya 10, PJOK 10, PKWU 10 dan PKWU 11. Hasil dari penelitian ini diharapkan dapat membantu para guru untuk memberikan pendampingan lebih kepada para siswa, khususnya pada kategori pelajaran sulit agar mutu pembelajaran dapat meningkat dan mampu bersaing dengan sekolah lainnya.

REFERENCES

- [1] F. Anggraeni, "Faktor-Faktor Yang Mempengaruhi Kesulitan Belajar Peserta Didik Pada Pembelajaran Pendidikan Pancasila dan Kewarganegaraan Di SMP Negeri 3 Gadingrejo," *J. Soc. Educ.*, vol. 1, no. 2, pp. 108–121, 2020, doi: 10.23960/jips.v1i2.108-121.
- [2] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.
- [3] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.
- [4] N. Afiasari, N. Suarna, and N. Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means," *J. SAINTEKOM*, vol. 13, no. 1, pp. 100–110, 2023, doi: 10.33020/saintekom.v13i1.402.
- [5] D. A. Fakhri, S. Defit, and Sumijan, "Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering," *J. Inf. dan Teknol.*, vol. 3, pp. 160–166, 2021, doi: 10.37034/jidt.v3i3.137.
- [6] J. Faran and R. T. Aldisa, "Perbandingan Algoritma K-Means dan K-Medoids Dalam Pengelompokan Kelas Untuk Mahasiswa Baru Program Magister," *J. Inf. Syst. Res.*, vol. 5, no. 2, pp. 509–519, 2024, doi: 10.47065/josh.v5i2.4753.
- [7] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.
- [8] Silvia Ningsih, "Penerapan Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan," *J. Sist. Inf.*, vol. 3, no. 2, pp. 73–82, 2022, doi: 10.32546/jusin.v3i2.1961.
- [9] M. Azzam Al Fauzie and J. Akhir Putra, "Clustering Data Menggunakan Metode K-Means untuk Rekomendasikan Pembelajaran Akademik bagi Siswa Aktif dalam Ekstrakurikuler," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 642–648, 2023, doi: 10.30865/klik.v4i1.1116.
- [10] A. Asmana, Y. Arie Wijaya, and M. Martanto, "Clustering Data Calon Siswa Baru Menggunakan Metode K-Means Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 552–559, 2022, doi: 10.36040/jati.v6i2.5236.
- [11] V. Darsono, Amroni, and A. Andrianti, "Penerapan Data Mining Algoritma K-Means Untuk Rekomendasi Pemilihan Bidang Studi Perguruan Tinggi Pada Siswa SMKN 1 Kota Jambi," *J. Inform. Dan Rekayasa Komputer (JAKAKOM)*, vol. 2, no. 2, pp. 161–171, 2022, doi: 10.33998/jakakom.2022.2.2.80.
- [12] A. I. Rizmayanti, N. Hidayati, F. S. Nugraha, and W. Gata, "Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decision Tree (Studi Kasus Smk Multicom Depok)," *Swabumi*, vol. 9, no. 1, pp. 9–18, 2021, doi: 10.31294/swabumi.v9i1.8363.
- [13] V. No, M. Qusyairi, Z. Hidayatullah, A. Sandi, "Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow Infotek: Jurnal Informatika dan Teknologi Perkembangan teknologi saat ini berkembang dengan sangat pesat ini terbukti," *Infotek: Jurnal Informatika dan Teknologi*, vol. 7, no. 2, pp. 500–510, 2024.
- [14] Lea and Jenie, "Penerapan Algoritma Apriori Untuk Menentukan Impor Barang Pada PT. Pafa Mandiri Sakti," *J. Process.*, vol. 15, no. 1, pp. 54–61, 2020, doi: 10.33998/processor.2020.15.1.767.
- [15] D. Winarti, M. Kom, E. Revita, and M. Kom, "Penerapan Data Mining untuk Analisa Tingkat Kriminalitas Dengan Algoritma Association Rule Metode FP-Growth," *J. SIMTIKA*, vol. 4, no. 3, pp. 8–22, 2021.
- [16] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center," *Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang*



- Layak Mengikuti Asesment Cent. Untuk Clust. Progr. Sdp, vol. 3, no. 1, pp. 87–93, 2018.
- [17] U. T. Suryadi and Y. Supriatna, “Cluster 2 = 169,” vol. 22, no. April, p. 111, 2019.
- [18] F. Yunita, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru,” *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [19] M. . Rani Rotul Muhima, S.Si. et al., *Kupas Tuntas Algoritma Clustering Konsep, Perhitungan Manual dan Program*. 2018.
- [20] A. Winarta and W. J. Kurniawan, “Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python,” *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 5, no. 1, pp. 113–119, 2021, doi: 10.59697/jtik.v5i1.593.
- [21] E. Muningsih, “Optimasi Jumlah Cluster K-Means Dengan Metode Elbow Untuk Pemetaan Pelanggan,” *Pros. Semin. Nas. Elinvo*, no. September, pp. 105–114, 2017.
- [22] A. P. Riani, A. Voutama, and T. Ridwan, “Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 164, 2023, doi: 10.53513/jsk.v6i1.7351.
- [23] 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.
- [24] N. T. Hartanti, “Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.