

IMPLEMENTASI TEKNIK *CLUSTERING* MENGGUNAKAN ALGORITMA *K-MEANS* PADA SMK YADIKA 3 JAKARTA BERBASIS *WEB*

Yogi Ahmad¹, Titin Fatimah^{2*}

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹2011501398@student.budiluhur.ac.id, ^{2*}titin.fatimah@budiluhur.ac.id
(* : corresponding author)

Abstrak-Penelitian ini bertujuan mengidentifikasi faktor-faktor yang mempengaruhi tingkat ketidakhadiran siswa di SMK Yadika 3 Jakarta menggunakan teknik *clustering* algoritma *K-Means*. Data penelitian diambil dari siswa kelas 10, 11, dan 12 pada tahun ajaran 2023-2024. Algoritma *K-Means* digunakan karena keunggulannya dalam mengelompokkan data tanpa label kelas, sehingga dapat mengidentifikasi pola ketidakhadiran yang tidak terlihat sebelumnya. Teknik *elbow* digunakan untuk menentukan jumlah *cluster* optimal. Hasil pengujian ini menggunakan data kehadiran siswa dan faktor siswa dengan jumlah 936 baris data. Pada pengujian evaluasi *elbow* dengan 10 *cluster*, 9 *cluster* merupakan *cluster* paling optimal dengan memiliki nilai selisih *elbow* SSE sebesar 958.37 yang merupakan nilai selisih terendah pada *cluster*. Pada setiap anggota *cluster*, yang memiliki anggota paling sedikit yaitu C1 yang memiliki anggota sebanyak 46 siswa dan anggota *cluster* terbanyak yaitu C6 yang memiliki anggota sebanyak 194 siswa. Dari hasil pengujian identifikasi faktor siswa kelompok yang memiliki faktor yang buruk, hampir keseluruhan adalah kelompok C9 dan hasil pengujian yang memiliki faktorsedikit yaitu kelompok C4 yang hanya buruk pada faktor motivasi belajar siswa dan kelompok C7 yang hanya buruk pada faktor kesehatan siswa. Penelitian tersebut dapat dikembangkan pada sekolah untuk memberikan arahan serta intervensi kepada siswa dan lingkungan di sekolah SMK Yadika 3 Jakarta seperti membuat kebijakan pada siswa yaitu membuat program motivasi, program kesehatan dan kebersihan, layanan konseling dan menyediakan fasilitas transportasi sekolah guna mencegah penyebab ketidakhadiran pada siswa.

Kata Kunci: *Data Mining, K-means Clustering, Elbow, Siswa, Faktor*

IMPLEMENTATION OF CLUSTERING TECHNIQUES USING THE *K-MEANS* ALGORITHM IN WEB-BASED SMK YADIKA 3 JAKARTA

Abstract-This study aims to identify the factors that affect the level of student absenteeism in SMK Yadika 3 Jakarta using the *k-Means* algorithm clustering technique. The study Data were taken from students in Grades 10, 11, and 12 in the 2023-2024 school year. The *k-Means* algorithm is used because of its advantage in grouping data without class labels, so that it can identify patterns of absence not seen before. Elbow technique is used to determine the optimal number of clusters. The results of this test using student attendance data and student factors with a total of 936 rows of data. In elbow evaluation test with 10 clusters, 9 clusters are the most optimal clusters with elbow difference value with SSE method 958.37 which is the lowest difference value from other clusters. In each cluster member, which has the fewest members is C1 which has members as many as 46 students and the most cluster members is C6 which has members as many as 194 students. From the test results of the identification of student factors, the group that has a bad factor is almost the entire C9 group and the test results that have a little factor is the C4 group that is only bad at the factor of student learning motivation and C7 group that is only bad at the factor of student health. The research can be developed in schools to provide direction and intervention to students and the environment at SMK Yadika 3 Jakarta such as making policies on students, namely creating motivational programs, health and hygiene programs, counseling services and providing school transportation facilities to prevent the causes of absenteeism in students.

Keywords: *Data Mining, , K-means Clustering, Elbow, students, factor*

1. PENDAHULUAN

Pertumbuhan teknologi informasi dimana saat ini berkembang secara cepat menuntut kita untuk mampu memanfaatkan teknologi informasi secara maksimal, memanfaatkan teknologi informasi digunakan dalam berbagai aspek kehidupan manusia untuk menciptakan alur kerja yang mudah digunakan dan menyediakan informasi untuk kebutuhan spesifik yang lebih rinci, mudah mengikuti, dan efisiensi. Pemanfaatan teknologi juga akan meningkatkan keakuratan data dan informasi [9].

Data mining adalah metode mengekstraksi data dari database besar menggunakan kombinasi bidang ilmiah seperti statistika, matematika, inteligensi buatan, dan *machine learning* untuk menemukan pola serta mengenali data informasi yang bermanfaat. Salah satu metode yang digunakan dalam penelitian ini yaitu metode *clustering*. *Clustering* merupakan sebuah proses pengelompokan data secara otomatis tanpa informasi label kelas. Pendekatan algoritma *k-means*, salah satu algoritma clustering yang paling populer, digunakan karena keunggulannya yaitu mudah diimplementasikan [5].

Sekolah merupakan tempat pendidikan setelah lingkungan keluarga, terkadang di sekolah siswa mempunyai perbuatan-perbuatan negatif yang dapat mempengaruhi perkembangannya mentalnya, oleh karena itu kedisiplinan dari para siswa memiliki signifikansi yang sangat besar karena dapat menjadi salah satu faktor kunci dalam mencapai keberhasilan di sekolah dan masa depan. SMK Yadika 3 Jakarta, seperti banyak institusi pendidikan menengah kejuruan lainnya di Indonesia, SMK Yadika 3 Jakarta yang berlokasi di Tegal Alur Jakarta Barat mempunyai jurusan di berbagai bidang seperti Teknik Otomotif, Teknik Sepeda Motor, dan Teknik Informatika dan Jaringan.

Di SMK Yadika 3 Jakarta, evaluasi siswa tidak semata-mata didasarkan pada prestasi akademik maupun non-akademik namun juga pada tingkat kehadiran para siswa di sekolah. Catatan kehadiran siswa, termasuk seluruh ketidakhadirannya, menjadi dasar evaluasi ini. Namun dalam proses pelaksanaannya, pihak sekolah masih kesulitan mengidentifikasi siswa yang tingkat ketidakhadirannya tinggi. Hal ini disebabkan banyaknya siswa yang absen kronis sehingga sulit menentukan secara akurat siswa mana yang benar-benar mengalami masalah absensi kronis dan siswa mana yang tidak.

Tingkat absensi yang tinggi selalu menggunakan perkiraan. Hal ini membuat proses identifikasi tingkat ketidakhadiran siswa yang tinggi menjadi tidak objektif dan tidak terukur. Kurangnya sistem yang tepat dalam menentukan kriteria ketidakhadiran siswa yang tinggi menjadi penyebab masalah ini. Hal ini menyebabkan penilaian kehadiran siswa di sekolah menjadi tidak konsisten dan tidak akurat.

Perbedaan yang terdapat pada penelitian ini dengan penelitian sebelumnya adalah mengidentifikasi faktor ketidakhadiran siswa yang mempelajari apa yang menjadi faktor penyebab siswa yang sering tidak hadir dengan memahami karakteristik setiap siswa yang dikelompokkan dengan Teknik *clustering k-means*. Penelitian ini akan memanfaatkan teknik *elbow* untuk menetapkan jumlah *cluster* yang terbaik untuk mengidentifikasi faktor-faktor pengaruh tingkat ketidakhadiran di sekolah SMK Yadika 3 Jakarta.

Usulan penelitian tugas akhir ini dapat diharapkan memberikan pengetahuan yang lebih dalam tentang keterkaitan antara faktor-faktor yang teridentifikasi dengan tingkat ketidakhadiran siswa. Temuan dari penelitian ini dapat dimanfaatkan untuk merancang program penerapan intervensi dengan tepat sasaran dan efektif untuk mengatasi masalah ketidakhadiran siswa di SMK Yadika 3 Jakarta dan sekolah lainnya.

Clustering atau pengelompokan adalah teknik *data mining* yang digunakan untuk menganalisis data guna memecahkan masalah dalam mengelompokkan data, lebih khusus lagi membagi kumpulan data menjadi beberapa *subset* [4].

Clustering K-Means adalah salah satu teknik *clustering* yang paling umum digunakan dalam analisis data. Algoritma ini bekerja dengan cara membagi data menjadi K kelompok yang sebelumnya ditetapkan dimana K adalah jumlah kelompok yang diinginkan.[8] Algoritma *K-Means* adalah suatu metode pengelompokan data yang mempunyai karakteristik yang sama ke dalam kelompok yang sama serta data yang mempunyai karakteristik yang berbeda ke dalam kelompok data yang berbeda.[11]

Menurut [6] perhitungan dalam *K-means* adalah sebagai berikut:

- Menentukan k sebagai jumlah klaster yang akan dibentuk.
- K *centroid* (titik pusat klaster) awal akan dibangkitkan secara acak.

$$v = \sum_{k=1}^n \frac{X_k}{N} \quad (1)$$

$k = 1, 2, 3 \dots n$

Keterangan:

- v = *Centroid* yang akan ditentukan
- X_k = Titik data ke- k dalam *dataset*
- N = Jumlah total titik data

- c. Jarak setiap objek ke masing-masing *centroid* dihitung dari masing-masing kluster.
- d. Jarak antara objek dan *centroid* dihitung menggunakan *Euclidian Distance*.

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

Keterangan:

- $d(x, y)$ = Jarak antara titik x dan titik y
- $|x - y|$ = Nilai absolut dari selisih antara titik x dan y
- $\sum_{i=1}^n$ = Penjumlahan untuk setiap dimensi dari titik data
- variabel x indeks i = Nilai pada posisi ke i dalam vektor x
- variabel y indeks i = Nilai pada posisi ke i dalam vektor y

- e. Alokasikan setiap objek pada *centroid* terdekat.
- f. Lakukan iterasi, setelah itu dengan menggunakan persamaan tentukan posisi *centroid* baru. Jika posisi *centroid* tidak sama, maka ulangi langkah ke 3.

Metode *Elbow* merupakan suatu teknik menghitung jumlah *cluster* yang ideal dengan menghitung nilai *Sum of Square Error* (SSE) dari masing-masing *cluster*. Semakin besar perbedaan nilai SSE dengan *cluster* berikutnya sehingga membentuk titik sudut siku, semakin banyak *cluster* yang ada [1]. Dalam menentukan jumlah *cluster* yang optimal, maka rumus pada metode *elbow* dengan menggunakan rumus SSE adalah sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{xi \in C_k} |xi - ck|^2 \quad (3)$$

Keterangan :

- K = Jumlah kluster
- k = Indeks yang menggantikan nilai dari 1 hingga K dalam penjumlahan
- C_k = Kluster ke- k
- xi = Titik data ke- i dalam kluster C_k
- ck = *Centroid* (pusat) dari kluster C_k
- $||xi - ck||^2$ = Jarak *Euclidean* kuadrat antara titik data xi dan *centroid* ck

2. METODE PENELITIAN

2.1 Data Penelitian

Data penelitian merupakan informasi atau data bersih yang belum dimanipulasi dan tidak mengandung campur tangan pihak ketiga sehingga dapat digunakan untuk analisis penelitian. Dengan penelitian di SMK Yadika 03 Jakarta yang merupakan sekolah menengah kejuruan standar nasional yang terletak di DKI Jakarta. SMK Yadika 03 memiliki siswa sebanyak 936 siswa pada tahun ajaran 2023/2024 yang terdiri dari 342 siswa kelas 10, 329 siswa kelas 11, dan 263 siswa kelas 12. Data yang penulis dapat yaitu data kehadiran dalam bentuk excel yang jika digabungkan berisi 936 baris data. Data faktor-faktor yang terdiri dari 5 faktor yang didapat dari wawancara hingga disetujui pihak sekolah dan sudah diberikan penilaian siswa dalam faktor oleh pihak sekolah berbentuk skor 1 sampai 5 yang mana 1 merupakan sangat buruk dan 5 merupakan sangat baik. Adapun data yang diberikan SMK Yadika 3 Jakarta meliputi atribut-atribut pada Tabel 1 dibawah ini :

Tabel 1. Atribut Data Penelitian

Atribut	Keterangan
Nama	Nama siswa
SKT	Absen sakit
IZN	Absen izin
ALP	Absen tanpa keterangan
KTN	Faktor kesehatan siswa
MB	Faktor motivasi belajar siswa
TP	Faktor transportasi siswa ke sekolah
MP	Faktor model pembelajaran siswa
HG	Faktor hubungan dengan guru pada siswa

2.2 Metode Pengumpulan Data

Dalam melakukan penelitian, penulis mengumpulkan data dengan tujuan untuk memperoleh informasi yang relevan terkait dengan topik yang dibahas, yaitu:

a. Wawancara

Wawancara dilakukan dengan Wakil Kepala Sekolah SMK Yadika 3 Jakarta untuk memperoleh informasi mengenai kriteria faktor-faktor penyebab ketidakhadiran siswa dengan memilah faktor-faktor yang akan diidentifikasi dan sudah disetujui oleh wakil kepala sekolah dan pihak sekolah. Wawancara ini dilakukan untuk memperoleh wawasan lebih dalam tentang menganalisa faktor yang mungkin terjadi pada siswa dalam kehadiran sekolah serta dapat memberikan solusi untuk membantu pihak sekolah untuk membuat kebijakan baru yang lebih baik.

b. Dokumentasi

Dalam dokumentasi, penulis meminta kepada Wakil Kepala Sekolah untuk menyediakan data yang diperlukan sebagai input dalam proses pengembangan sistem aplikasi.

c. Tinjauan Pustaka

Tinjauan pustaka dilakukan untuk mengumpulkan informasi dan pengetahuan tentang *data mining* dan *K-Means Clustering* dari jurnal-jurnal yang akan digunakan sebagai referensi dalam penelitian.

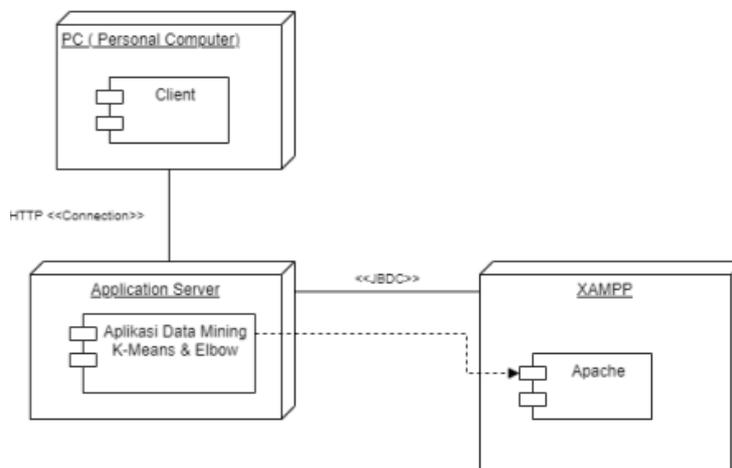
3. HASIL DAN PEMBAHASAN

3.1 Lingkungan Percobaan

Sebelum mengimplementasikan dan menguji coba sistem, perlu dilakukan persiapan yang matang agar proses tersebut berjalan sesuai dengan rencana dan harapan.

3.1.1 Deployment Diagram

Deployment diagram digunakan untuk memvisualisasikan hubungan antara perangkat lunak (software) dan perangkat keras (*hardware*). Seperti Gambar 1 dibawah ini:



Gambar 1. Deployment Diagram

3.2 Hasil Pengujian Nilai Elbow

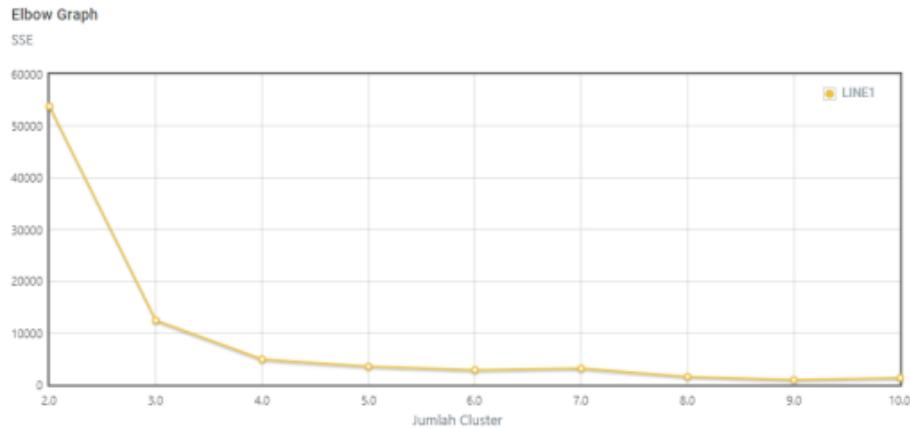
Uji coba ini dilakukan dengan data kehadiran dan faktor siswa SMK Yadika 3 Jakarta dan dilakukan dengan memproses data menggunakan algoritma *clustering K-Means* dengan jumlah 10 *cluster*. Hasil uji coba ditunjukkan pada tabel 2 sebagai berikut:

Tabel 2. Hasil Pengujian *Clustering*

No	Jumlah <i>Cluster</i>	Hasil <i>Clustering</i>	Hasil Evaluasi <i>Elbow</i> (SSE)	Selisih Hasil <i>Elbow</i> (SSE)
1	2	C1 = 265 C2 = 671	53807,76	53807,76
2	3	C1 = 152 C2 = 325 C3 = 459	41352,10	12455,65
3	4	C1 = 99 C2 = 191 C3 = 236 C4 = 410	36435,51	4916,59
4	5	C1 = 70 C2 = 143 C3 = 172 C4 = 374 C5 = 177	32896,85	3538,65
5	6	C1 = 66 C2 = 115 C3 = 158 C4 = 386 C5 = 135 C6 = 76	30041,75	2855,09
6	7	C1 = 47 C2 = 70 C3 = 90 C4 = 136 C5 = 85 C6 = 376 C7 = 132	26882,75	3159,00
7	8	C1 = 49 C2 = 42 C3 = 79 C4 = 80 C5 = 160 C6 = 71 C7 = 325 C8 = 130	25317,74	1565,00
8	9	C1 = 46 C2 = 68 C3 = 76 C4 = 52 C5 = 119 C6 = 194 C7 = 87 C8 = 192 C9 = 102	24359,36	958,37
9	10	C1 = 45 C2 = 36 C3 = 59 C4 = 70 C5 = 65 C6 = 119 C7 = 152 C8 = 123 C9 = 160 C10 = 107	23028,65	1330,71

Dengan menggunakan algoritma *K-Means*, peneliti menguji data kehadiran siswa di SMK Yadika 3 Jakarta dan menemukan bahwa dengan pengujian jumlah *cluster* dari 2 hingga 10 terdapat perbedaan jumlah siswa dalam setiap *cluster*, serta nilai SSE *elbow* dan selisih *elbow* antara *cluster* yang menurun secara signifikan, yang menunjukkan pengelompokan *cluster* dari hasil selisih *elbow* yang terkecil yang paling baik dan optimal. Pada

pengujian ini jumlah *cluster* yang paling optimal adalah *cluster* 9 yang memiliki nilai selisih *elbow* dengan metode SSE sebesar 958,37 dan pada *cluster* 10 terdapat peningkatan pada selisih *elbow* sebesar 1330,71 . Pada Gambar 2 yaitu dari hasil grafik pada nilai selisih *elbow* sebagai berikut :



Gambar 2. Hasil *Elbow* dengan 10 *Cluster*

3.3 Hasil Pengelompokkan Faktor Siswa

Dari data yang diidentifikasi dengan algoritma *K-means clustering*, berikut ini adalah hasil pengujian dengan menggunakan *cluster* yang optimal yaitu *cluster* 9 pada pengelompokkan siswa dengan 9 *cluster* pada tabel 3 di bawah ini :

Tabel 3. Hasil Pengelompokkan Faktor Siswa

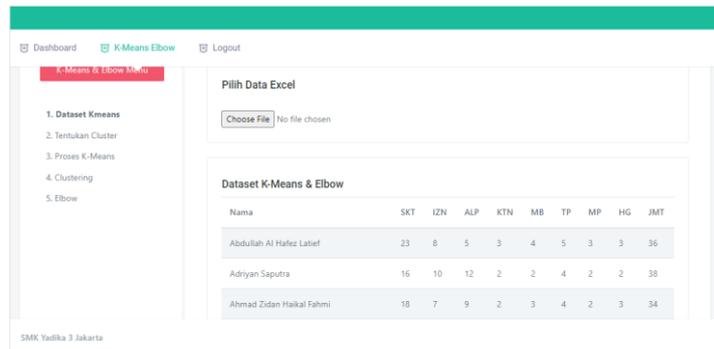
Cluster	Keterangan
C1	Siswa di kelompok ini memiliki faktor kesehatan yang cukup, faktor motivasi belajar yang cukup, faktor kemudahan transportasi yang cukup, faktor model pembelajaran yang buruk, faktor hubungan dengan guru yang buruk.
C2	Siswa di kelompok ini memiliki faktor kesehatan yang cukup, faktor motivasi belajar yang cukup, faktor kemudahan transportasi yang buruk, faktor model pembelajaran yang cukup, faktor hubungan dengan guru yang buruk.
C3	Siswa di kelompok ini memiliki faktor kesehatan yang buruk, faktor motivasi belajar yang buruk, faktor kemudahan transportasi yang cukup, faktor model pembelajaran yang buruk, faktor hubungan dengan guru yang cukup.
C4	Siswa di kelompok ini memiliki faktor kesehatan yang cukup, faktor motivasi belajar yang buruk, faktor kemudahan transportasi yang cukup, faktor model pembelajaran yang cukup, faktor hubungan dengan guru yang cukup.
C5	Siswa di kelompok ini memiliki faktor kesehatan yang cukup, faktor motivasi belajar yang buruk, faktor kemudahan transportasi yang buruk, faktor model pembelajaran yang cukup, faktor hubungan dengan guru yang cukup.
C6	Siswa di kelompok ini memiliki faktor kesehatan yang buruk, faktor motivasi belajar yang buruk, faktor kemudahan transportasi yang cukup, faktor model pembelajaran yang cukup, faktor hubungan dengan guru yang buruk.
C7	Siswa di kelompok ini memiliki faktor kesehatan yang buruk, faktor motivasi belajar yang cukup, faktor kemudahan transportasi yang cukup, faktor model pembelajaran yang cukup, faktor hubungan dengan guru yang cukup.
C8	Siswa di kelompok ini memiliki faktor kesehatan yang cukup, faktor motivasi belajar yang cukup, faktor kemudahan transportasi yang buruk, faktor model pembelajaran yang buruk, faktor hubungan dengan guru yang cukup.
C9	Siswa di kelompok ini memiliki faktor kesehatan yang buruk, faktor motivasi belajar yang buruk, faktor kemudahan transportasi yang cukup, faktor model pembelajaran yang buruk, faktor hubungan dengan guru yang buruk.

Dari hasil pengujian faktor-faktor yang mempengaruhi ketidakhadiran siswa berdasarkan hasil pengujian dari analisis *k-means* yang dapat disimpulkan yaitu C9 yang merupakan kelompok siswa yang memiliki paling banyak faktor buruk. Hasil yang memiliki paling sedikit buruk pada pengujian faktor siswa adalah anggota C4

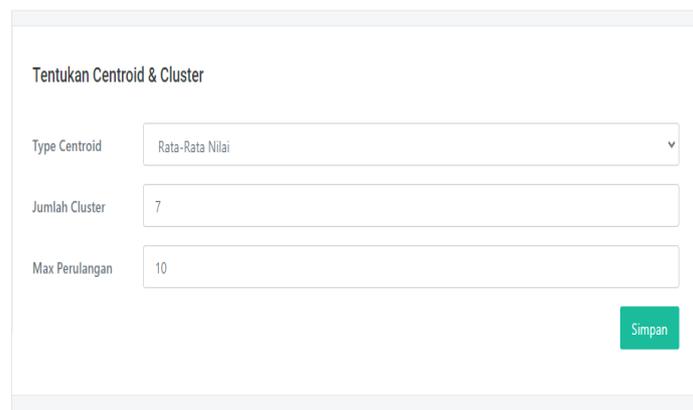
yang hanya memiliki faktor buruk pada motivasi belajar siswa dan anggota C7 yang hanya memiliki faktor buruk pada kesehatan siswa.

3.4 Tampilan Layar Aplikasi

Pada gambar 3. merupakan tampilan *import dataset*. Pada gambar 4 merupakan tampilan tentukan *cluster*. Pada gambar 5 merupakan tampilan proses *k-means*.



Gambar 3. Tampilan *import dataset*



Gambar 4. Tampilan tentukan *cluster*

Perulangan 1 - Hitung Euclidean Distance										
Abdullah Al Hafez Latief	28.573082248551	32.474721563932	32.610350383048	36.016020172509	38.464214838252	:				
Adriyan Saputra	27.626293163553	31.864382685665	32.069686665938	36.182229188741	38.336144853489	:				
Ahmad Zidan Haikal Fahmi	24.044178541172	28.166272098709	28.348613851986	32.182849448407	34.505076448765	:				
alfito Pratama	12.070198254331	15.257347250834	15.359931296024	18.271296512037	20.752708921989	:				
Alvino	10.480407417186	14.653564195786	14.822402262169	18.86649818707	20.920681814912	:				

Gambar 5. Tampilan proses *k-means*

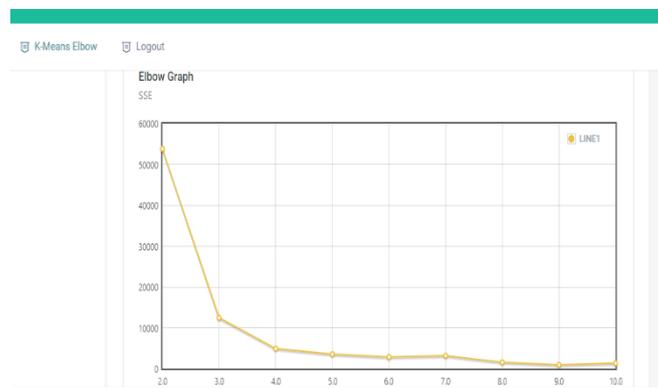
Pada gambar 6 merupakan tampilan hasil pengelompokkan *clustering*. Pada gambar 7 merupakan tampilan hasil *elbow*.

6	3.02	3.01	3.00	3.07	2.93
7	2.93	2.96	3.03	2.98	2.79

Jumlah Cluster

Cluster	Jumlah	Keterangan
1	57	Siswa di kelompok ini memiliki faktor kesehatan yang cukup, faktor motivasi belajar yang cukup, faktor kemudahan transportasi yang cukup, faktor model pembelajaran yang buruk, faktor hubungan dengan guru yang buruk.
2	67	Siswa di kelompok ini memiliki faktor kesehatan yang cukup, faktor motivasi belajar yang buruk, faktor kemudahan transportasi yang buruk, faktor model pembelajaran yang cukup, faktor hubungan dengan guru yang cukup.

Gambar 6. Tampilan hasil pengelompokan *clustering*



Gambar 7. Tampilan hasil *elbow*

4. KESIMPULAN

Dari hasil faktor siswa yang di hasilkan dari pengujian pada faktor siswa pada 9 *cluster* yaitu C1 sampai C9 bahwa anggota C9 yang merupakan kelompok siswa yang memiliki paling banyak faktor buruk dan Hasil yang memiliki paling sedikit buruk pada pengujian faktor siswa adalah anggota C4 yang hanya memiliki faktor buruk pada motivasi belajar siswa dan anggota C7 yang hanya memiliki faktor buruk pada kesehatan siswa. Jumlah *cluster* yang optimal untuk mengelompokkan data kehadiran siswa SMK Yadika 3 adalah 9 *cluster* yang memiliki nilai selisih *elbow* dengan metode SSE sebesar 958.37. Berdasarkan hasil pengujian, SMK Yadika 3 Jakarta membuat kebijakan yang tepat seperti menyediakan fasilitas transportasi, program pemeriksaan kesehatan, layanan konseling, dan seminar motivasi untuk meningkatkan kehadiran dan kinerja siswa.

DAFTAR PUSTAKA

- [1] P. R. Adella, V. Apriade and R. Taufik, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, no. 1, pp. 164-172, 2023.
- [2] Y. Aditia and A. Rio, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20-28, 2023.
- [3] D. Irwan, r. sholeh, V. A. Sandy, H. H. Nirwana and F. R. Anang, "Implementasi Metode Clustering Pada Data Penerimaan Siswa Baru Menggunakan Algoritma K-Means (Studi Kasus SMK Darul Falah Kertagena Tengah)," *Jurnal Insand Comtech*, vol. 7, no. 2, pp. 1-7, 2022.
- [4] N. Ismasari, P. Reza and Suherman, "Implementasi Algoritma K-Means Dalam Mengkategorikan Produk Terlaris Dan KurangLaris Pada Toko Alfamart Cikarang," *Jurnal Pelita Teknologi*, vol. 16, no. 1, pp. 79-87, 2021.
- [5] A. K. M, J. Asep and I. A. R, "Pengaruh Pendapatan Orang Tua terhadap Hasil Belajar Siswa Menggunakan Algoritma K-Means Clustering," *Jurnal Pendidikan Tambusai*, vol. 7, no. 3, pp. 31434-31444, 2023.
- [6] B. A. M. Muhammad, M. C. Yusup and E. Ultach, "Penerapan Algoritma K-Means dan Decision Tree Dalam Analisis Prestasi Siswa Sekolah Menengah Kejuruan," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, p. 355-364, 2022.

- [7] S. Muhammad and A. Khurotul, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow Dan Silhouette Pada Model Clustering Dengan Menggunakan Algoritma K Means," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 8, no. 1, pp. 56-65, 2023.
- [8] H. Nirwana, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, pp. 46-56, 2024.
- [9] I. M. Qorik and S. B. Eko, "Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 3, no. 4, p. 522-528, 2022.
- [10] Rasiban and P. R. M. Samuel, "Penerapan Data Mining Untuk Memprediksi Penerimaan Peserta Didik Baru Jalur Prestasi Akademik Di SMA Negeri 13 Jakarta Dengan Menggunakan Algoritma Random Forest," *INNOVATIVE: Journal Of Social Science Research*, vol. 3, no. 4, pp. 10065-10079, 2023.
- [11] M. Rifaldie and Z. Hadi, "Penerapan Algoritma K-Means dalam Penentuan Siswa Bermasalah Berdasarkan Running Record (Studi Kasus: SMK Averus Jakarta)," *JRIIN: Jurnal Riset Informatika dan Inovasi*, vol. 1, no. 7, pp. 720-729, 2024.
- [12] D. Sofia and Oktaviawati, "Penerapan Algoritma C4.5 untuk Pehamanan Siswa SMK Pada Pelajaran Kompetensi Keahlian," *INTERNAL (Information System Journal)*, vol. 5, no. 2, pp. 116-125, 2022.
- [13] W. Sri and Y. Sonia, "Penerapan Data Mining Untuk Memprediksi Siswa Berprestasi Dengan Menggunakan Algoritma K Nearest Neighbor," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 3, pp. 2598-2611, 2022.
- [14] Suryani and Mustakim, "Estimasi Keberhasilan Siswa dalam Pemodelan Data Berbasis Learning," *Bulletin of Informatics and Data Science*, vol. 1, no. 2, pp. 81-88, 2022.