

Penerapan Metode *K-Means Clustering* Untuk Pemetaan Kompetensi Peserta Didik Baru Berdasarkan Hasil Tes Diagnostik

Muchammad Okgistiyan^{1*}, Irmayansyah²

^{1,2}Program Studi Sistem Informasi, Fakultas Informatika dan Komputer, Universitas Binaniaga Indonesia

¹email: muchokgistiyan@gmail.com

²email: irmayansyah@unbin.ac.id

*Corresponding Author

ABSTRACT

The mapping of student competencies is an essential step in grouping students based on the results of diagnostic tests. Ineffective grouping can hinder decision-making for students with low scores. This research develops a prototype application to map student competencies using the K-Means Clustering algorithm. The prototype application analyzes data and calculates groupings, followed by reanalysis to evaluate students' competencies in each group. Variables such as literacy and numeracy scores were used as clustering parameters. This mapping aims to provide information that helps schools determine which students require additional learning hours. System feasibility testing by experts resulted in a 100% suitability rating, categorized as "Highly Suitable." User satisfaction testing using PSSUQ also achieved an overall score of 100%, indicating the application's usability. The validity of the clustering was assessed using the Silhouette Coefficient, yielding a score of 0.467, which categorizes the clustering as "Weak."

Keywords: Mapping, K-Means Clustering, Diagnostic Test, Literacy, Numeracy, Silhouette Coefficient.

ABSTRAK

Proses pemetaan kompetensi peserta didik merupakan langkah penting dalam pengelompokan berdasarkan hasil tes diagnostik. Pengelompokan yang tidak efektif dapat menghambat pengambilan keputusan yang tepat terhadap peserta didik yang nilainya rendah. Penelitian ini mengembangkan sebuah prototype aplikasi untuk memetakan kompetensi peserta didik menggunakan algoritma K-Means Clustering. Pada penelitian ini dibuat sebuah prototype aplikasi yang dapat melakukan pemetaan kompetensi peserta didik dengan menerapkan Algoritma K-means Clustering yaitu dengan cara menganalisis data dan melakukan perhitungan pengelompokan, yang kemudian perhitungan pengelompokan dianalisis kembali untuk melihat kompetensi peserta didik di masing-masing kelompok. Variabel yang digunakan mencakup nilai literasi dan numerasi sebagai parameter pengelompokan. Hal ini dilakukan untuk memetakan kompetensi peserta didik agar dapat menyajikan informasi yang bisa menjadi acuan sekolah untuk mengambil kebijakan peserta didik mana yang harus mendapatkan jam tambahan. Pada aplikasi yang dibangun telah dilakukan uji kelayakan oleh ahli sistem dan diperoleh persentase kelayakan sebesar 100% yang dapat dikategorikan ke dalam interpretasi "Sangat Layak". Telah dilakukan juga uji kelayakan oleh pengguna dengan menggunakan kuesioner PSSUQ sesuai dengan kategori PSSUQ diantaranya yaitu nilai kepuasan secara keseluruhan (Overall) sebesar 100%, yang artinya aplikasi ini layak digunakan. Serta telah dilakukan juga uji validitas cluster menggunakan *Silhouette Coefficient* terhadap algoritma K-means yang diterapkan dengan nilai yang diperoleh sebesar 0,467 yang berarti cluster yang dibentuk termasuk kedalam kategori "Lemah".

Kata Kunci: Pemetaan, K-Means Clustering, Tes Diagnostik, Literasi, Numerasi, Silhouette Coefficient.

A. PENDAHULUAN

1. Latar Belakang

Pemerataan pendidikan sedang digalakkan melalui penyebaran jalur Penerimaan Peserta didik Baru (PPDB). Menurut Perwali No. 13 Tahun 2024 tentang Pedoman PPDB Kota Bogor, PPDB terdiri dari beberapa jalur, yaitu Jalur Afirmasi, Jalur Prestasi Akademik/Non-Akademik, Jalur Masalah GTK/Perpindahan Orang Tua, Jalur Anak Berkebutuhan Khusus, dan Jalur Zonasi. Jalur Afirmasi adalah satu jalur penerimaan peserta didik baru yang memberikan penegasan dan penguatan untuk memberikan layanan akses pendidikan kepada kelompok peserta didik yang berasal dari keluarga ekonomi tidak mampu. Jalur Prestasi Akademik/Non-Akademik dan Prestasi Nilai Rapor adalah salah satu jalur penerimaan peserta didik baru atas penghargaan bagi peserta didik di bidang akademik/non-akademik pada tingkat internasional, nasional, provinsi, dan/atau tingkat Kota/Kabupaten, serta prestasi nilai rapor. Jalur Masalah GTK/Perpindahan Orang Tua adalah satu jalur penerimaan peserta didik baru bagi peserta didik yang domisili mengikuti perpindahan tempat tugas orang tua dan/atau bagi anak guru/tenaga kependidikan. Jalur Anak Berkebutuhan Khusus adalah salah satu jalur penerimaan peserta didik baru untuk calon peserta didik baru yang mengalami keterbatasan atau keuarbiasaan, baik itu dalam segi fisik, mental intelektual, sosial, dan emosional. Jalur Zonasi adalah salah satu jalur penerimaan peserta didik baru berdasarkan sistem pembagian wilayah dengan mempertimbangkan letak geografis, wilayah administrasi, dan letak satuan pendidikan terhadap domisili calon peserta didik. Dari sekian banyak jalur penerimaan peserta didik baru, semua peserta didik dengan kemampuan belajar yang berbeda-beda akan berkumpul menjadi satu. Maka dari itu, sekolah mengadakan Tes Diagnostik sebelum pembelajaran di tahun ajaran baru dimulai.

Tes Diagnostik bertujuan untuk melihat kemampuan peserta didik dalam memahami pelajaran. Tes diagnostik yang diadakan di sekolah meliputi tes Literasi dan Numerasi. Paul Black, dkk. (2003, p. 74) menyatakan bahwa, "diagnostic assessments provide essential information about students' starting points, helping teachers to adjust their teaching strategies to better address the diverse needs within the classroom". Dalam kutipan tersebut dijelaskan bahwa pentingnya tes diagnostik untuk memberikan informasi awal tentang kemampuan belajar peserta didik, yang dapat digunakan oleh guru untuk menyesuaikan strategi pembelajaran.

K-Means Clustering K-means adalah algoritma clustering berbasis prototipe yang sederhana dan bersifat partitional (memisahkan data menjadi kelompok). Algoritma ini berusaha menemukan K cluster yang tidak tumpang tindih. Setiap cluster direpresentasikan oleh centroid-nya (centroid dari suatu cluster biasanya merupakan rata-rata dari titik-titik yang ada dalam cluster tersebut) (Junjie Wu, 2012, p. 7).

Data mining adalah proses menemukan pola-pola menarik dan informasi bermanfaat dari kumpulan data yang besar. Proses ini melibatkan berbagai teknik, seperti klasifikasi, *Clustering*, regresi, dan asosiatif, untuk mengekstrak informasi yang berguna dari data yang tidak terstruktur. Jiawei Han, dkk. (2012, p. 29) menyatakan bahwa, data mining dapat didefinisikan sebagai "*the process of discovering interesting patterns and knowledge from large amounts of data*". Jiawei Han, dkk. (2012, p. 35) juga menyatakan bahwa, "data mining adalah salah satu langkah dalam proses Knowledge Discovery in Databases (KDD) yang meliputi pembersihan data, transformasi data, penambahan data, evaluasi pola, dan presentasi pengetahuan".

Penerapan *K-Means Clustering* dalam pendidikan, khususnya untuk pemetaan hasil tes diagnostik peserta didik, memungkinkan guru dan pendidik untuk mengelompokkan peserta didik berdasarkan hasil tes yang mereka peroleh. Hal ini membantu dalam mengidentifikasi pola-pola atau kelompok peserta didik dengan karakteristik serupa dalam hal kompetensi yang dimiliki. Dengan mengelompokkan peserta didik ke dalam klaster, pendidik dapat merancang strategi pembelajaran yang lebih efektif dan disesuaikan dengan kebutuhan masing-masing kelompok.

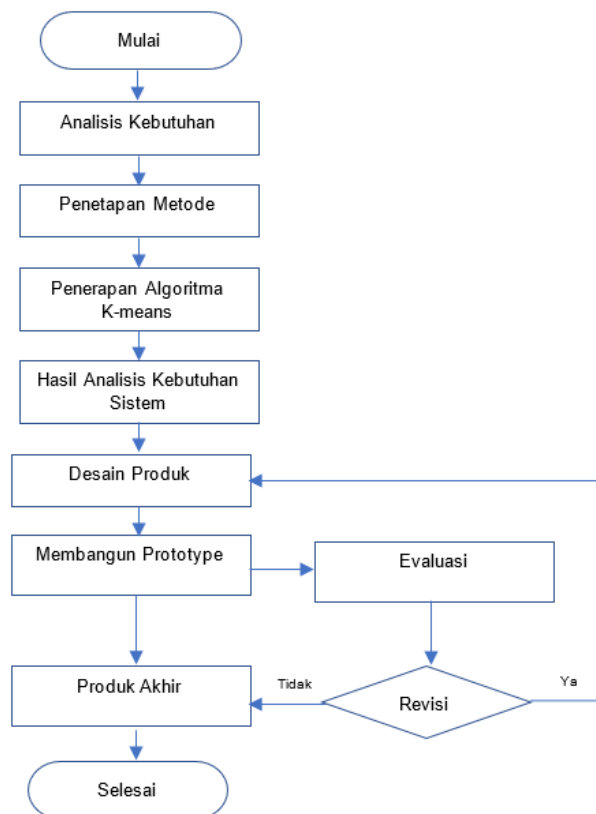
Pada penelitian sebelumnya, dilakukan penelitian menggunakan metode *K-Means* untuk pengelompokan hasil belajar dengan judul "Penerapan *K-Means Clustering* Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow" yang dilakukan oleh Adella Putri Riani, Apriade Voutama, dan Taufik Ridwan. Pada penelitian sebelumnya memiliki fokus yang sama dengan penelitian yang ingin dilakukan saat ini yaitu untuk membantu mengetahui kemampuan peserta didik baru. Namun pada penelitian sebelumnya masih memiliki kekurangan pada penggunaan variabel dan belum ada pengujian hasil *Clustering K-Means*.

Berdasarkan uraian diatas, metode *K-Means Clustering* diharapkan dapat membantu menyelesaikan permasalahan sesuai judul penelitian ini yaitu tentang "Penerapan Metode *K-Means Clustering* Untuk Pemetaan Kompetensi Peserta Didik Baru Berdasarkan Hasil Tes Diagnostik".

B. METODE

1. Tahapan Penelitian

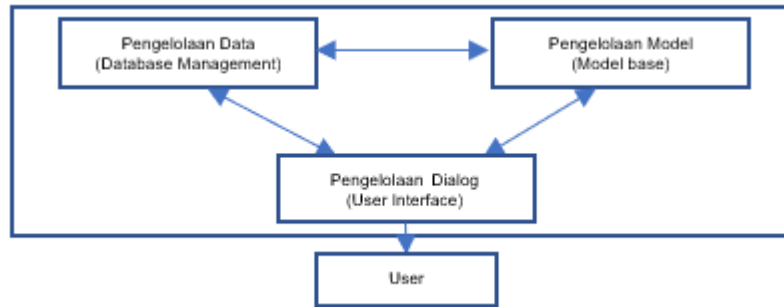
Tahapan penelitian merupakan langkah-langkah dari proses pengembangan yang dilakukan. proses pengembangan dimulai dari analisis kebutuhan, penetapan metode, penerapan algoritma k-means, desain produk, dan membangun prototype. Setelah proses membangun prototype selesai, proses selanjutnya adalah evaluasi dan revisi. Jika prototype sudah sesuai, maka proses berlanjut ke proses produk akhir, menandakan bahwa proses pengembangan telah selesai. Berikut adalah gambaran alur dari proses pengembangan:



Gambar 1. Tahapan Penelitian

2. Metode Penelitian

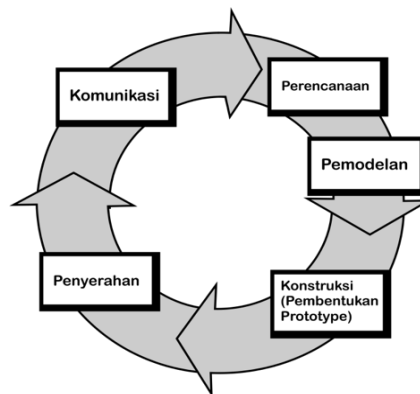
Model teoritis Algoritma *K-Means* yaitu metode untuk membagi data yang ada kedalam satu atau lebih *cluster*. Metode ini membagi data kedalam *cluster* sehingga pada *cluster* tersebut terdapat data yang memiliki karakteristik yang sama. Adapun model konseptual yang digunakan dalam penelitian ini yaitu model Sistem Pendukung Keputusan. Sistem pendukung keputusan atau SPK merupakan sistem yang dirancang untuk membantu dalam pengambilan sebuah keputusan. Pengambilan keputusan digunakan untuk memecahkan suatu permasalahan. Permasalahan yang rumit membutuhkan pengambilan keputusan yang rumit juga. Pada penelitian ini permasalahan yang ada yaitu dalam menentukan pemetaan kompetensi peserta didik baru. Dengan menggunakan konsep Sistem Pendukung keputusan yang terdiri dari tiga komponen.



Gambar 2 Model Konseptual

Sumber: (Riadi dan Muchlisin, 2013)

- Pertama database management, komponen ini merupakan pengorganisasian data yang relevan dengan masalah pada penelitian ini yaitu data hasil tes diagnostik peserta didik baru.
- Kedua Model base, konsep ini merupakan suatu model yang merepresentasikan permasalahan dalam penentuan pemetaan kompetensi peserta didik baru kedalam format kuantitatif. Dengan mengembangkan dan membandingkan solusi alternative, model base ini memungkinkan pengambil keputusan menganalisa secara utuh.
- Ketiga user interface merupakan gabungan antara komponen database management dan model base. User interface menampilkan tampilan sistem yang dimengerti oleh pihak sekolah. Model procedural yang digunakan dalam penelitian ini yaitu model prototyping.



Gambar 3 Alur Prototyping

Sumber: (Pressman dan Maxim, 2019)

Langkah-langkah pada model prototyping adalah yang pertama yaitu melakukan komunikasi antara pengembang dan pihak sekolah tentang tujuan pembuatan sistem dan mengidentifikasi kebutuhan. Kedua, melakukan perencanaan berdasarkan hasil diskusi dengan pihak sekolah. Ketiga, melakukan pemodelan dengan membuat model dan desain yang fokus pada gambar dari segi software apakah dapat dilihat dan digunakan oleh pihak sekolah. Keempat yaitu melakukan konstruksi (pembentukan model prototype) dengan memulai pembuatan prototype. Kelima yaitu melakukan penyerapan prototype kepada pihak sekolah untuk dievaluasi dan pemberian nilai tentang kekurangan atau kebutuhan pihak sekolah terhadap sistem.

3. Uji coba produk

Instrumen pengumpulan data pada penelitian ini terdiri dari 2 instrumen, yaitu pengujian untuk ahli dan pengujian untuk pengguna. Untuk uji ahli menggunakan *Blackbox Testing*, dan untuk pengguna menggunakan PSSUQ. Rumus yang digunakan adalah sebagai berikut;

$$\text{Persentase Kelayakan}(\%) = \frac{\text{Skor yang diobservasi}}{\text{Skor yang diharapkan}} \times 100\%$$

Menurut (Arikunto, 2009, 44) terdapat 5 kategori kelayakan, skala ini akan memperlihatkan rentang dari bilangan persentase. Nilai maksimal yang diharapkan adalah 100% dan minimum 0%.

Tabel 1. Rentang Kategori Kelayakan

Persentase Pencapaian	Interpretasi
< 21%	Sangat Tidak Layak
21% - 40%	Tidak Layak
41% - 60%	Cukup Layak
61% - 80%	Layak
81% - 100%	Sangat Layak

Sumber: (Arikunto 2009)

4. Uji Hasil

Berdasarkan hasil penerapan metode pada sistem dilakukan uji hasil menggunakan metode *Silhouette Coefficient*. Terdapat 3 langkah untuk menghitung Silhouette Coefficient, yaitu menghitung cohesion atau jarak data ke semua data pada cluster yang sama, menghitung separation atau jarak rata-rata semua titik data di cluster tetangga terdekat, dan yang terakhir menghitung silhouette coefficient. Rumus menghitung Silhouette Coefficient adalah sebagai berikut;

- a. Hitung rata-rata jarak dari suatu dokumen misalkan i dengan semua dokumen lain yang ada dalam satu cluster;

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)$$

Dengan j adalah dokumen lain dalam satu cluster A dan $d(i, j)$ adalah jarak antara dokumen i dengan j ;

- b. Hitung rata-rata jarak dari dokumen i tersebut dengan semua dokumen di cluster lain, kemudian diambil nilai terkecilnya;

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j)$$

Dengan $d(i, C)$ adalah jarak rata-rata dokumen i dengan semua objek pada cluster lain C dimana $A \neq C$;

- c. Setelah menghitung $d(i, c)$ untuk semua c , maka diambil nilai terkecil dengan menggunakan persamaan 3;

$$b(i) = \min_{C \neq A} d(i, C)$$

Cluster B yang mencapai minimum (yaitu, $d(i, B) = b(i)$) disebut tetangga dari objek (i). ini adalah cluster terbaik kedua untuk objek (i).

- d. Nilai *silhouette coefficient* adalah sebagai berikut;

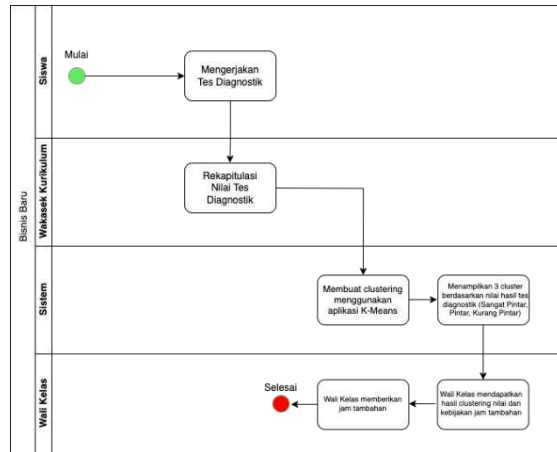
$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

C. HASIL DAN PEMBAHASAN

1. Hasil

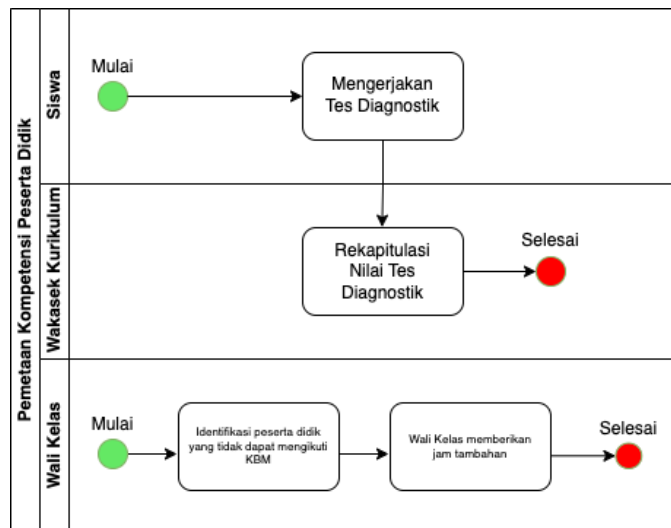
a. Hasil Analisis Proses

Berdasarkan analisis kebutuhan yang telah dilakukan, pengelompokan nilai hasil tes diagnostik akan dilakukan dengan menggunakan metode *K-Means Clustering*. Adapun proses bisnis lama yang berlangsung adalah sebagai berikut:



Gambar 4. Proses Bisnis Lama

Pada proses bisnis lama yang tertera di Gambar 4, terdapat alur yang terputus ketika tes diagnostik telah dilaksanakan. Alur yang dilakukan pada bisnis lama berhenti setelah wakasek kurikulum melakukan rekapitulasi nilai hasil tes diagnostik, dan tidak ada tindak lanjut dalam pengambilan keputusan jam pelajaran tambahan kepada peserta didik yang nilainya dirasa kurang. Jam pelajaran tambahan diadakan berdasarkan hasil observasi Wali Kelas yang mendapati peserta didiknya tidak bisa mengikuti proses KBM di kelas.



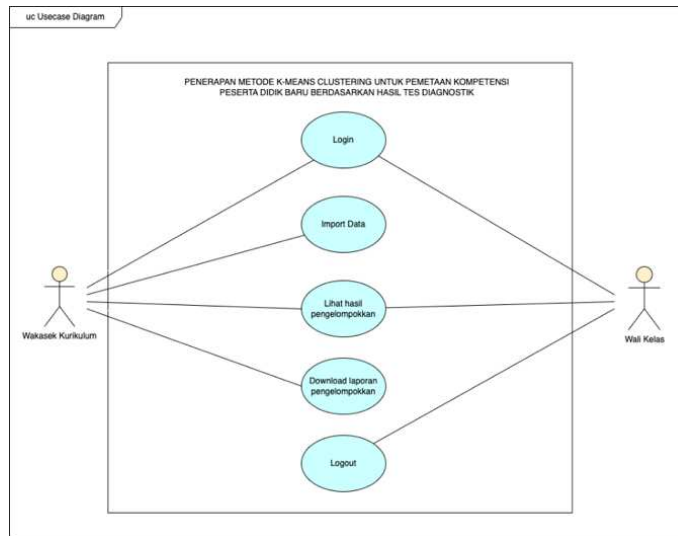
Gambar 5. Proses Bisnis Baru

b. Hasil Analisis Metode

Metode pengelompokan yang diterapkan adalah *K-Means Clustering* untuk mengelompokkan nilai hasil tes diagnostik untuk pemetaan kompetensi peserta didik. Pendekatan ini menggunakan nilai tes diagnostik sebagai dasar untuk mengidentifikasi pola dan membentuk kelompok peserta didik menjadi 2 klaster, yaitu Nilai Rendah dan Nilai Tinggi. Hasil dari pengelompokan ini membantu Wakasek bidang Kurikulum memetakan kompetensi peserta didik dan membuat kebijakan terkait jam tambahan untuk peserta didik yang nilai tes diagnostiknya kurang.

c. Hasil Analisis Kebutuhan Sistem

Pada pengembangan sistem ini pemodelan objek dijelaskan dalam bentuk usecase diagram berdasarkan pada proses pemetaan kompetensi peserta didik. Pada pengembangan sistem ini, terdapat 1 aktor dan 6 usecase dalam pemetaan kompetensi peserta didik berdasarkan nilai hasil tes diagnostik. 1 aktor tersebut yaitu pengguna dan 6 usecase tersebut yaitu usecase login, import data, menampilkan hasil import data, menampilkan hasil clustering, menampilkan laporan hasil clustering, dan logout. Dari usecase tersebut dapat dijelaskan alur kegiatan yang dapat dilakukan yaitu diawali dari pengguna melakukan login terlebih dahulu untuk dapat mengakses sistem. Kemudian pengguna bisa melakukan import data dalam format .csv, melihat data yang telah diimport, melihat hasil clustering, dan melihat laporan hasil clustering. Usecase diagram dapat dilihat pada Gambar 6.

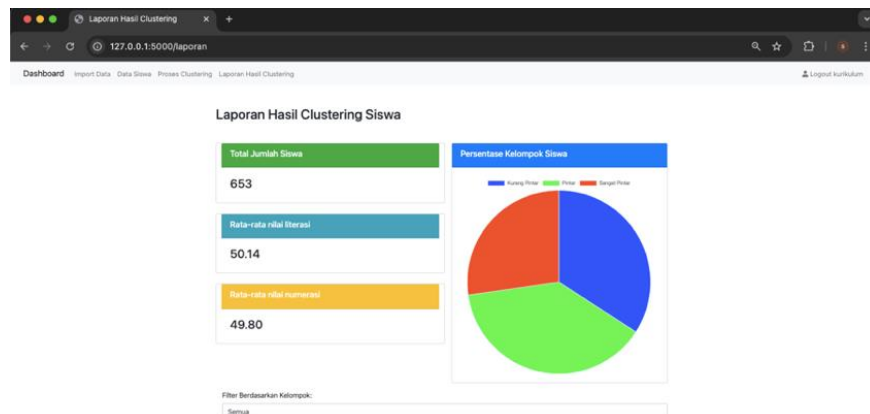


Gambar 6 Use Case Diagram

Pada gambar 3, dapat dilihat bahwa terdapat 2 jenis pengguna, yaitu Wakasek Kurikulum dan Wali Kelas. Wakasek Kurikulum memiliki akses penuh terhadap sistem, sementara Wali Kelas hanya dapat melihat laporan hasil clustering.

d. Produk Akhir

Pada halaman ini terdapat card yang memuat keterangan total siswa, rata-rata nilai literasi, dan rata-rata nilai numerasi. Pada halaman ini juga ditampilkan pie chart yang memuat presentasi tiap cluster. Pada halaman ini juga terdapat filter untuk memilih cluster mana yang akan ditampilkan. Terdapat juga button fungsi untuk mengcopy, mendownload file dengan format excel, .csv, dan .pdf, dan juga fungsi untuk print jika user ingin langsung mencetak laporan hasil clustering. Tampilan halaman proses clustering dapat dilihat pada gambar 7.



Gambar 7 Tampilan Aplikasi

D. PEMBAHASAN

1. Perhitungan Iterasi K-Means

a. Perhitungan Metode K-Means

Cluster pemetaan kompetensi peserta didik dibuat berdasarkan nilai hasil tes diagnostik. Dalam nilai hasil tes diagnostik terdapat 2 variabel, yaitu nilai Literasi dan Numerasi. Algoritma K-means secara umum digunakan untuk membagi data kedalam beberapa kelompok. Berikut tahapan algoritma K-means.

b. Persiapan Data

Pada perhitungan ini, dataset yang digunakan yaitu menggunakan variabel nilai literasi dan numerasi. Dataset yang telah dikumpulkan tersebut terdiri dari 654 baris dan 2 kolom.

Tabel 2. Data Peserta Didik

No.	NAMA LENGKAP (Sesuai Akte)	Literasi	Numerasi
1	Ananda Dinar Ramadhani	44	53
2	Abdad Farras Orlando	84	49
3	Abdullah hisyam	21	51
4	Abiyu Dzaky Putra Widiyanto	65	29
5	Adelia Apriliyanti	36	45
6	Adelia Apriliyanti	21	77
7	Adelia Dwi Batrisya Yusup	49	30
8	Adelia Malona	86	38
9	Adika Dhia Wicaksono	77	58
10	Adila Dwi Ariyani	37	40
No.	NAMA LENGKAP (Sesuai Akte)	Literasi	Numerasi
...
654	Zulfiandra Satria Wardhani	20	62

- c. Menentukan Jumlah *Cluster*
 Berdasarkan data dari Tabel 1 akan dibentuk 2 klaster, yaitu; Klaster 1 merupakan klaster Nilai Rendah dan klaster 2 merupakan klaster Nilai Tinggi.
- d. Inisiasi Pusat Cluster Awal
 Data nilai hasil diagnostik akan dibagi kedalam 2 klaster sehingga diperlukan 2 titik centroid awal untuk Klaster 1(C1) dan Klaster 2 (C2). Langkah pertama untuk menghitung centroid awal adalah menghitung nilai rentang masing-masing variabel cluster dengan rumus sebagai berikut:

$$\frac{\text{min} - \text{max}}{2}$$

$$\text{Variabel literasi} = \frac{95 - 10}{2} = 42.5$$

$$\text{Variabel numerasi} = \frac{91 - 16}{2} = 37.5$$

Setelah mendapatkan nilai rentang masing-masing variabel, tahapan selanjutnya adalah menghitung nilai batas cluster dengan rumus sebagai berikut:

$$\text{min} + \text{nilai rentang}$$

$$\text{Variabel literasi} = 10 + 42.5 = 52.5$$

$$\text{Variabel numerasi} = 16 + 37.5 = 53.5$$

Tahapan selanjutnya yaitu menghitung nilai centroid masing-masing cluster dengan rumus sebagai berikut:

Variabel Literasi (x)

Variabel Numerasi(y)

$$C1 = a + \frac{b-a}{2} = 10 + \frac{52.5-10}{2} = 31.25$$

$$C1 = a + \frac{b-a}{2} = 16 + \frac{53.5-16}{2} = 34.75$$

$$C2 = a + \frac{b-a}{2} = 52.5 + \frac{52.5-10}{2} = 73.75$$

$$C2 = a + \frac{b-a}{2} = 53.5 + \frac{53.5-16}{2} = 72.25$$

Dari perhitungan dengan rumus di atas, diperoleh Centroid awal nilai literasi (x) yaitu C1=31.25 dan C2=73.75. Untuk centroid awal nilai numerasi(y), diperoleh nilai C1=34.75 dan C2=72.25. Nilai Centroid tiap variabel dapat dilihat pada tabel berikut:

Tabel 3. Nilai Centroid Awal

Centroid Awal	Literasi (x)	Numerasi (y)
C1	31.25	34.75
C2	73.75	72.25

- e. Hitung Jarak Dengan Centroid
 Perhitungan jarak data dengan centroid dilakukan untuk menentukan jarak terpendek yang akan digunakan untuk menentukan pengelompokan klaster. *Euclidean Distance* digunakan untuk menghitung jarak data dengan centroid. *Euclidean Distance* merupakan perhitungan jarak dari 2 titik

dan untuk mempelajari hubungan antara sudut dan jarak. Rumus *Euclidean Distance* adalah sebagai berikut:

$$d(p, q) = \sqrt{(q_i - p_i)^2}$$

Keterangan:

d = jarak/distance

p = titik p (data siswa)

q = centroid

Iterasi 1

Perhitungan jarak dengan klaster 1:

Data ke-1:

$$d(p, q) = \sqrt{(q_i - p_i)^2}$$

$$d(p, q) = \sqrt{(31.25 - 44)^2 + (34.75 - 53)^2}$$

$$= 22.26263686$$

Tabel 4. Hasil Perhitungan Jarak Dengan Klaster 1

Data	Hasil Perhitungan
Data ke-2	54.64087298
Data ke-3	19.21262606
Data ke-4	34.23631113
Data ke-5	11.29712353
Data ke-6	43.47556785
Data ke-7	18.37457483
Data ke-8	54.84637636
Data ke-9	51.31885618
Data ke-10	13.7716096
Data	Hasil Perhitungan
...	...
Data ke-654	29.48092604

Perhitungan jarak dengan klaster 2:

Data ke-1:

$$d(p, q) = \sqrt{(q_i - p_i)^2}$$

$$d(p, q) = \sqrt{(73.75 - 44)^2 + (72.25 - 53)^2}$$

$$d(p, q) = 35.43479928$$

Tabel 5. Hasil Perhitungan Jarak Dengan Klaster 2

Data	Hasil Perhitungan
Data ke-2	25.40915189
Data ke-3	56.86936785
Data ke-4	44.12623936
Data ke-5	46.55775983
Data ke-6	52.96343078
Data ke-7	48.96554911
Data ke-8	36.37478522
Data ke-9	14.61591598
Data ke-10	48.89401804
Data	Hasil Perhitungan
...	...
Data ke-654	54.7185983

Hasil clustering iterasi ke-1 dapat dilihat pada Tabel 5.

Tabel 6. Hasil Clustering Iterasi ke-1

No.	NAMA LENGKAP (Sesuai Akte)	Jenis Kelamin	Literasi	Numerasi	Jarak Euclid Cluster 1	Jarak Euclid Cluster 2	Cluster
1	Ananda Dinar Ramadhani	Perempuan	44	53	22.26263686	35.43479928	Nilai Rendah
2	Abdad Farras Orlando	Laki-laki	84	49	54.64087298	25.40915189	Nilai Tinggi
3	Abdullah hisyam	Laki-laki	21	51	19.21262606	56.86936785	Nilai Rendah
4	Abiyu Dzaky Putra Widiyanto	Laki-laki	65	29	34.23631113	44.12623936	Nilai Rendah
5	Adelia Apriliyanti	Perempuan	36	45	11.29712353	46.55775983	Nilai Rendah
6	Adelia Azrilla	Perempuan	21	77	43.47556785	52.96343078	Nilai Rendah
7	Adelia Dwi Batrisya Yusup	Perempuan	49	30	18.37457483	48.96554911	Nilai Rendah
8	Adelia Malona	Perempuan	86	38	54.84637636	36.37478522	Nilai Tinggi
9	Adika Dhia Wicaksono	Perempuan	77	58	51.31885618	14.61591598	Nilai Tinggi
10	Adila Dwi Ariyani	Perempuan	37	40	7.786205751	48.89401804	Nilai Rendah
No.	NAMA LENGKAP (Sesuai Akte)	Jenis Kelamin	Literasi	Numerasi	Jarak Euclid Cluster 1	Jarak Euclid Cluster 2	Cluster
...
654	Zulfiandra Satria Wardhani	Laki-laki	20	62	29.48092604	54.7185983	Nilai Rendah

Pada proses iterasi ini menggunakan rumus yang sama hingga nilai centroid tidak berubah. Nilai centroid tidak berubah ketika sudah sampai pada proses iterasi ke-6. Hasil pengelompokkan dan nilai centroid iterasi ke-6 dapat dilihat pada tabel berikut.

Tabel 7. Hasil Iterasi ke-6

No.	NAMA LENGKAP (Sesuai Akte)	Literasi	Numerasi	Jarak Euclid Cluster 1	Jarak Euclid Cluster 2	Cluster
1	Ananda Dinar Ramadhani	44	53	13.45234616	31.42054593	Nilai Rendah
2	Abdad Farras Orlando	84	49	53.13758985	8.790840444	Nilai Tinggi
3	Abdullah hisyam	21	51	9.917671694	54.2433614	Nilai Rendah
4	Abiyu Dzaky Putra Widiyanto	65	29	40.10180553	22.88190754	Nilai Tinggi
5	Adelia Apriliyanti	36	45	7.203092125	39.47593315	Nilai Rendah
6	Adelia Azrilla	21	77	28.69265272	60.80959665	Nilai Rendah
7	Adelia Dwi Batrisya Yusup	49	30	27.03647578	32.66093419	Nilai Rendah
8	Adelia Malona	86	38	56.43066251	15.74076979	Nilai Tinggi
9	Adika Dhia Wicaksono	77	58	46.80549975	8.711575734	Nilai Tinggi
10	Adila Dwi Ariyani	37	40	11.7785052	39.3779627	Nilai Rendah
No.	NAMA LENGKAP (Sesuai Akte)	Literasi	Numerasi	Jarak Euclid Cluster 1	Jarak Euclid Cluster 2	Cluster
...
654	Zulfiandra Satria Wardhani	20	62	16.14924301	56.62513255	Nilai Rendah

Tabel 8. Nilai centroid akhir

Centroid	Literasi (x)	Numerasi (y)
C1	30.87297297	50.0594595
C2	75.22183099	49.471831

2. Menghitung *Silhouette Coefficient*

Jika nilai $a(i)$ dan $b(i)$ sudah didapat, tahapan terakhir dari uji ahli ini yaitu menghitung nilai *Silhouette Coefficient*($s(i)$). Perhitungan *silhouette coefficient* adalah sebagai berikut:

Data ke-1:

$$s(i) = \frac{35.32903702 - 21.69860583}{\max(21.69860583, 35.32903702)} = 0.386$$

Untuk data ke-2 hingga data ke-654 masih menggunakan rumus dan proses yang sama dengan data ke-1 untuk mencari *Silhouette Coefficient*. Dari hasil perhitungan *Silhouette Coefficient*, didapat nilai rata-rata sebesar 0,467. Nilai ini masuk kedalam kategori sebaran cluster lemah.

E. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, maka kesimpulan yang didapatkan adalah sebagai berikut:

1. Pengelompokan kompetensi berdasarkan hasil tes diagnostik menghasilkan dua kategori, yaitu Nilai Rendah dan Nilai Tinggi. Untuk menentukan hasil clustering menggunakan dua variabel, yaitu variabel Literasi dan Numerasi. Hal ini membantu pihak sekolah memahami kemampuan awal peserta didik untuk menyusun strategi pembelajaran yang lebih terarah.
2. Penerapan algoritma K-Means Clustering untuk pemetaan Kompetensi peserta didik baru belum akurat dan efektif dikarenakan hasil evaluasi algoritma *K-Means Clustering* dengan nilai *Silhouette Coefficient* menghasilkan nilai 0,467. Nilai ini menunjukkan bahwa struktur klasterisasi berada dalam kategori lemah, dengan kualitas pemisahan cukup baik dan masih ada ruang untuk diperbaiki.

F. DAFTAR PUSTAKA

- [1] Aji, Wahyu Prasetyo. "Analisa Pengelompokkan Data Nilai Rapot Siswa Menggunakan Pendekatan Metode K-Means di SMK Ponpes Manba'ul Ullum Cirebon." *KOPERTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer*, vol. 3, no. 2, 2024, pp. 14–18.

- [2] Anggraeni, Nurliana, and Lisda Fitriana Masitoh. "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Siswa Baru pada SMK Letris Indonesia 1 Berbasis Web." *Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat*, vol. 7, no. 1, 2022, pp. 359–368.
- [3] Arikunto, Suharsimi. *Prosedur Penelitian: Suatu Pendekatan Praktik*. Rineka Cipta, 2009.
- [4] Azmi, Nurul, et al. "Implementasi Metode K-Means sebagai Upaya Penentuan Lokasi Promosi Penerimaan Siswa Baru." *Building of Informatics, Technology and Science (BITS)*, vol. 2, no. 3, 2022, pp. 649–660.
- [5] Black, Paul, Christine Harrison, et al. *Assessment for Learning: Putting It into Practice*. McGraw-Hill, 2003.
- [6] De Ville, Baryy. *Microsoft Data Mining: Integrated Business Intelligence for e-Commerce and Knowledge Management*. Butterworth-Heinemann, 2001.
- [7] Denny, Jollyta, et al. *Konsep Data Mining dan Penerapan*. Deepublish, 2020.
- [8] Elda, Yusma, et al. "Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means." *Jurnal Informasi dan Teknologi*, vol. 3, no. 2, 2021.
- [9] Fowler, Martin. *UML Distilled: A Brief Guide to the Standard Object Modeling Language*. 3rd ed., Pearson Education, 2004.
- [10] Han, Jiawei, et al. *Data Mining: Concepts and Techniques*. 3rd ed., Morgan Kaufmann, 2012.
- [11] Hendrastuty, Nirwana. "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa." *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 2, no. 1, 2024, pp. 46–56.
- [12] Jollyta, Denny, William Ramadhan, and Muhammad Zarlis. *Konsep Data Mining dan Penerapan*. Deepublish, 2020.
- [13] Lutz, Mark. *Learning Python*. 5th ed., O'Reilly Media, 2013.
- [14] Marisa, Fitri, et al. *Algoritma Populer dalam Intelligent Sistem Beserta Contoh Kasus*. Deepublish, 2023.
- [15] McLeod, Raymond, and George Schell. *Sistem Informasi Manajemen*. 10th ed., Salemba Empat, 2008.
- [16] Missa, Dionisius, et al. "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Penghasilan Orang Tua Siswa." *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 2, no. 1, 2021.
- [17] Pemerintah Kota. *Perwali No. 13 Tahun 2024 tentang Pedoman Penerimaan Peserta Didik Baru*. Pemerintah Kota, 2024.
- [18] Pramudya, Aditya, et al. "Pengelompokan Hasil Belajar Siswa SDN Tunas Jaya dengan Algoritma K-Means." *JATI (Jurnal Mahasiswa Teknik Informatika)*, Dec. 2023.
- [19] Pressman, Roger, and Bruce Maxim. *Software Engineering: A Practitioner's Approach*. McGraw-Hill, 2019.
- [20] Riani, Adella Putri, et al. "Penerapan K-Means Clustering dalam Pengelompokan Hasil Belajar Peserta Didik dengan Metode Elbow." *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 4, no. 2, 2023, pp. 164–172.
- [21] Riadi, and Muchlisin. www.kajianpustaka.com/2022/02/sistem-pendukung-keputusan-spk.html. Diakses pada tanggal 18 Aug. 2024.
- [22] Rosa, A. S., and M. Shalahuddin. *Rekayasa Perangkat Lunak: Terstruktur dan Berorientasi Objek*. Informatika Bandung, 2013.
- [23] Safiq, Rosad. *Bahasa Pemrograman Java Tingkat Dasar*. Nas Media Pustaka, 2022.
- [24] Sugiyono. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta, 2013.
- [25] Wali, Muhammad, et al. *Pengantar 15 Bahasa Pemrograman Terbaik di Masa Depan (Referensi & Coding)*. Sonpedia Publishing, 2023.
- [26] Wu, Junjie. *Advances in K-Means Clustering: A Data Mining Thinking*. Springer-Verlag, 2012.