



PENERAPAN ALGORITMA *K-MEANS CLUSTERING* NILAI KEMAMPUAN SISWA DALAM PELAJARAN MATEMATIKA

*Application Of The K-Means Clustering Algorithm Of Student Abilities In Mathematics
Lessons*

Ally Anilshi¹, Reynaldi Thimotius Abineno², Arini Aha Pekuwali³

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Kristen Wira Wacana
Sumba, Jln. W. Soeprapto No 35 Waingapu, Sumba Timur, Nusa Tenggara Timur

Email:ellysitumorangnamora@gmail.com

ABSTRACT

*Education is an important aspect in a country's development, which aims to produce quality human resources. The indicator of success in the educational process is student academic achievement. Mathematical numeracy literacy refers to a person's ability to understand, use, and think critically about mathematical concepts in everyday life. Data mining is the process of extracting useful knowledge or interesting patterns from a large or complex data set. How to carry out clustering to determine the level of numeracy literacy using the *K-MEANS* method utilizing student achievement scores and skills at SMK Negeri 2 Waingapu. Carrying out clustering to determine the level of numeracy literacy using the *K-means* method utilizing student achievement scores and skills at SMK Negeri 2 Waingapu. Clustering to determine the level of numeracy literacy using the *K-means* method utilizes the achievement and skill scores of students at SMK Negeri 2 Waingapu.*

Keywords: *Clustering, data mining, K-means, SMK N 2 Waingapu.*

ABSTRAK

Pendidikan merupakan aspek penting dalam pembangunan suatu negara, yang bertujuan untuk menghasilkan sumber daya manusia yang berkualitas. Indikator keberhasilan dalam proses pendidikan adalah prestasi akademik siswa. Literasi numerasi matematika mengacu pada kemampuan seseorang untuk memahami, menggunakan, dan berpikir secara kritis tentang konsep matematika dalam kehidupan sehari-hari. Data mining adalah proses ekstraksi pengetahuan yang berguna atau pola yang menarik dari sebuah kumpulan data besar atau kompleks. Bagaimana melakukan *clustering* untuk menentukan tingkat literasi numerasi dengan metode *K-MEANS* memanfaatkan nilai prestasi dan keterampilan siswa di SMK Negeri 2 Waingapu. Melakukan *clustering* untuk menentukan tingkat literasi numerasi dengan metode *K-means* memanfaatkan nilai prestasi dan keterampilan siswa di SMK Negeri 2 Waingapu. Melakukan *clustering* untuk menentukan tingkat literasi numerasi menggunakan metode *K-means* memanfaatkan nilai prestasi dan keterampilan siswa di SMK Negeri 2 Waingapu.

Kata kunci: *Clustering, data mining, K-means, SMK N 2 Waingapu.*

PENDAHULUAN

Pendidikan merupakan aspek penting dalam pembangunan suatu negara, yang bertujuan untuk menghasilkan sumber daya manusia yang berkualitas. Indikator keberhasilan dalam proses pendidikan adalah prestasi akademik siswa. Faktor yang memengaruhi prestasi belajar siswa adalah kemampuan siswa dalam belajar dan memahami setiap materi pelajaran yang diajarkan di sekolah. Kemampuan literasi numerasi matematika juga menjadi hal yang sangat penting dalam dunia pendidikan. Pembelajaran ini sudah ada di setiap jenjang mulai dari SD, SMP, dan SMA/SMK. Siswa pada SMK Negeri 2 Waingapu berjumlah 806 siswa. Saat ini SMK Negeri 2 Waingapu memiliki akreditasi B. Dengan banyaknya siswa pada SMK Negeri 2 Waingapu membutuhkan data mining untuk mengetahui nilai kemampuan siswa pada mata pelajaran matematika.

Data mining adalah proses ekstraksi pengetahuan untuk menggali sebuah data dan informasi yang besar, yang sebelumnya tidak diketahui untuk diketahui dan dipahami untuk database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang penting sekali. Tujuan utama dari data mining adalah untuk menemukan pola yang berulang, tren, atau informasi yang tersembunyi di dalam data yang mungkin tidak langsung terlihat atau dapat dikenali dengan metode analisis konvensional. Banyak algoritma untuk *clustering* sebuah data salah satunya algoritma *K-means* dan pada penelitian ini menggunakan algoritma *K-means*. *K-means* adalah salah satu algoritma dari beberapa algoritma untuk analisis *clustering* untuk mengklasifikasi data ke dalam kategori yang berbeda secara otomatis. Tujuan algoritma adalah untuk mempartisi himpunan data ke dalam beberapa kelompok yang berbeda berdasarkan kesamaan antar data. Dengan ini melakukan *clustering* untuk menentukan tingkat literasi numerasi dengan metode *K-means* memanfaatkan nilai prestasi dan keterampilan siswa di SMK Negeri 2 Waingapu.

MATERI DAN METODE

Data Mining

Data mining adalah pengambilan keputusan sebuah informasi dengan adanya pola yang terdapat pada sekumpulan data yang ada di database dengan jumlah yang cukup besar. Dari teori di atas dijabarkan oleh ahli di atas dapat ditarik garis besar, data mining adalah suatu pencarian dan Analisa pada database untuk ditemukan suatu pola dengan tujuan mengekstrak informasi pengetahuan yang akurat dan potensial serta dipahami dan berguna untuk pengambilan keputusan.

K-means

K-means ialah algoritma yang membagi data menjadi beberapa grup tidak selaras. Data yang mempunyai kesamaan kemiripan dikelompokkan menjadi satu cluster dan data yang mempunyai perbedaan dikelompokkan ke grup yang tidak selaras.

Adapun langkah-langkah dari algoritma *K-means* dijelaskan sebagai berikut:

1. Menentukan berapa jumlah cluster (k) yang diinginkan pada *dataset*.
1. Tentukan nilai pusat (centroid). Penentuan centroid pada tahap awal dilakukan secara acak, sedangkan pada tahap iterasi digunakan persamaan yang dapat dilihat pada rumus 1 berikut ini:

$$V_{ij} = \frac{1}{N_i} = \sum_{k=0}^{N_i} X_{kj}$$

Rumus 1 tentukan nilai centroid

Keterangan:

V_{ij} = Centroid rata-rata pada cluster ke – i untuk variabel ke – j

N_i = Jumlah anggota cluster ke – i

i, k = Indeks dari cluster

j = Indeks variabel

X_{kj} = Nilai data ke – k variabel ke – j untuk cluster tersebut

2. Menghitung jarak terpendek tiap data dengan *Centroid*. Gunakan rumus *Euclidean distance* (d) untuk menghitung jarak terpendek dengan *Centroid* tersebut. Adapun rumus K-Means tersebut adalah:

$$de = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Rumus 2. Menghitung jarak terpendek

Keterangan:

de = Euclidean distance

i = Banyaknya objek

(x, y) = Koordinat objek

(s, t) = Koordinat centroid

3. Kelompokkan objek berdasarkan jarak ke *centroid* terpendek.
4. Ulangi langkah ke-3 hingga ke-4, lakukan iterasi hingga *centroid* bernilai optimal.
Dalam perhitungan *K-means* jika ada data dengan perbedaan besaran angka yang cukup jauh maka data perlu dinormalisasi. Adapun rumus yang digunakan adalah:

$$X = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

Rumus 3. Normalisasi data

Keterangan:

X_i = Bobot atribut ke i

X_{min} = Bobot minimum atribut i

X_{max} = Bobot maksimum atribut i

Normalisasi data ini bertujuan agar nilai dari masing-masing data berada dalam rentang yang sama yakni rentang 0 sampai 1.

Clustering

Clustering adalah suatu metode dalam data mining yang berguna dalam menganalisis data agar lebih akurat dalam memecahkan masalah pembagian kelompok data atau penggolompokan Kumpulan data menjadi beberapa himpunan bagian . Dan *clustering* menurut ialah suatu metode pengelompokan data. *Clustering* artinya suatu metode pengelompokan

data kedalam beberapa cluster atau kelompok dimana antar data didalam satu cluster serta taraf kemiripan yang minimum terhadap antar cluster. Salah satu Teknik *clustering* ialah algoritma *K-means*.

Davies Bouldin Index

Davies Bouldin Index (DBI) adalah salah satu metode evaluasi yang memiliki skema evaluasi cluster internal dimana dilihat dari baik atau tidaknya hasil dari cluster dilihat dari kuantitas dan kedekatan antar data hasil cluster .

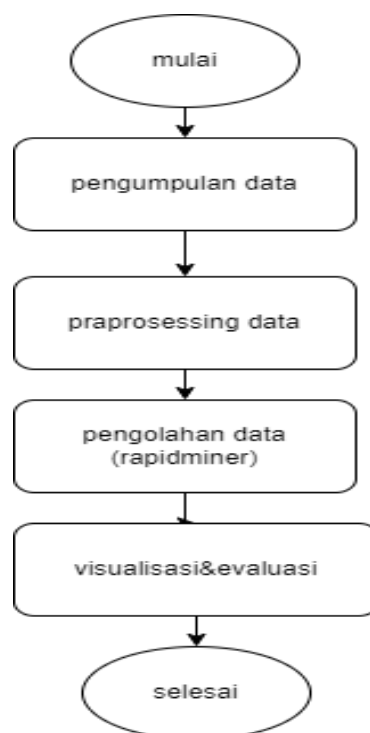
Langkah-langkah hitungan *Davies Bouldin Index* (DBI):

- 1) Menghitung nilai *Sum of Square Within-Cluster (SSW)* merupakan salah satu cara untuk mengetahui kohesi dalam sebuah cluster ke-i. Kohesi diartikan sebagai jumlah dari kedekatan atau kasamaan data terhadap pusat cluster dari sebuah cluster yang diikuti.
- 1) Menghitung nilai *Sum of Square Between-Cluster (SSB)*. *SSB* dihitung untuk mengetahui separasi atau jarak antar *cluster*.
- 2) Melakukan perhitungan Rasio dari nilai *SSW* dan nilai *SSB*.
- 3) Menghitung nilai *Davies Bouldin Index* (DBI) dari nilai rasio yang diperoleh sebelumnya dengan menggunakan persamaan berikut.

Rapidminer

Rapidminer merupakan aplikasi yang digunakan pada penelitian yang berhubungan dengan prediksi, peramalan, dan analisis karena banyak library pada berbagai macam bahasa dan rapidminer ialah platform perangkat lunak. Dan rapidminer ialah sebuah perusahaan platform aplikasi perangkat lunak atau *software* data ilmu pengetahuan yang dikembangkan oleh perusahaan yang sama dengan nama sebuah aplikasi yaitu rapidminer, yang terpadu untuk *machinelearning, deep learning, text mining, predictive analytics* .

Berikut merupakan kerangka berpikir yang akan menggambarkan tahapan-tahapan yang akan dilakukan.



Gambar 1. tahapan penelitian

Tahap pertama perlu dilakukan metode pengumpulan data yang akan digunakan dalam penelitian ini dengan menggunakan metode K-means. Penelitian akan berfokus pada hasil-hasil wawancara dan observasi, laporan, jurnal, berita maupun hasil penelitian serupa yang telah dihasilkan sebelumnya. Pengumpulan data dilakukan dengan cara melakukan survei atau mengunjungi secara langsung di lokasi objek penelitian yaitu di SMK Negeri 2 Waingapu. Terdapat jenis data yang diperoleh yaitu berupa format data excel. Penelitian ini menggunakan data pada bulan Mei 2024 sebanyak 806 siswa yang diperoleh dari SMK Negeri 2 Waingapu. Data yang digunakan dalam penelitian dicatat dalam aplikasi bernama Rapidminer.

Pada tahap ini perlu mempersiapkan dataset, karena setelah melakukan pengumpulan data, perlu dilakukan pengecekan kembali data-data yang sudah terkumpul. Kemudian melakukan pengolahan data atau melakukan analisis kluster lebih cepat untuk meminimalisir terjadinya kesalahan data. *Dataset* dapat terdiri dari berbagai jenis data, seperti data numerik (angka), data kategorikal (kelas), atau data teks, serta membersihkan tabel-tabel kosong lainnya. Setelah *dataset* sudah disiapkan maka akan lebih mempermudah dalam pengolahan atau *cluster*.

Pada tahap ini, *clustering* dilakukan dengan menggunakan metode *clustering* yang telah dipilih yaitu *K-Means*. Perhitungan menggunakan aplikasi *RapidMiner* untuk analisis data stunting, dari *Dataset* yang sudah disiapkan.

Hasil *clustering* yang diperoleh kemudian diinterpretasikan dengan cara menganalisis karakteristik atau ciri-ciri yang dimiliki oleh setiap kelompok data yang terbentuk, hasil tersebut disajikan dalam bentuk Tabel, dan Grafik. Hal ini dapat membantu dalam mengidentifikasi faktor-faktor yang berpotensi mempengaruhi kondisi stunting pada balita.

Metode evaluasi yang digunakan untuk mengukur kualitas setiap kluster dalam pengelompokan data adalah *Davies-Bouldin Index*. Metode ini digunakan untuk mengevaluasi performa dan mengukur akurasi dalam menggunakan metode *K-Means*.

HASIL DAN PEMBAHASAN

Pengumpulan data

Pengumpulan data telah dilakukan di SMK Negeri 2 Waingapu terkait data kenaikan kelas pada siswa, terdapat jenis data yang diperoleh yaitu berupa format dan excel yang didapat dari SMK Negeri 2 Waingapu. Penelitian ini menggunakan data pada Mei 2024 sebanyak 806 siswa yang diperoleh dari SMK Negeri 2 Waingapu. Pengolahan data akan dilakukan dengan dataset yang sudah disiapkan. Data yang digunakan dalam penelitian dicatat dalam aplikasi bernama Rapidminer.

Mempersiapkan Dataset

Mempersiapkan *dataset* sangat penting, karena setelah melakukan pengumpulan data, perlu dilakukan pengecekan kembali data-data yang sudah terkumpul. Berikut *dataset* yang sudah disiapkan untuk pengolahan *clustering* menggunakan aplikasi *Rapidminer*.

Tabel 1. Dataset

Nama siswa	Nilai pengetahuan	Nilai keterampilan	Nilai pengetahuan	Nilai keterampilan	Uas
------------	-------------------	--------------------	-------------------	--------------------	-----

No		uan	lan	n	n	
1	Abraham J. T. Pandang	74	84	74	84	74
2	Alponsius Katauhi Ndapa Ula	70	79	70	79	68
3	ANJAS UMBU NODAR	69	80	69	80	68
4	Arnis Ndawa Ndula Anding	72	81	72	81	70
5	ASMIN LANDU PRAING	72	81	72	81	69
6	DEFRISTO MANJA MBUMBAR	74	84	74	84	72
65	JEKSON TAY KAPURU	73	94	73	82	75
66	NIKOLAUS D.P. SAMB	72	82	72	94	75

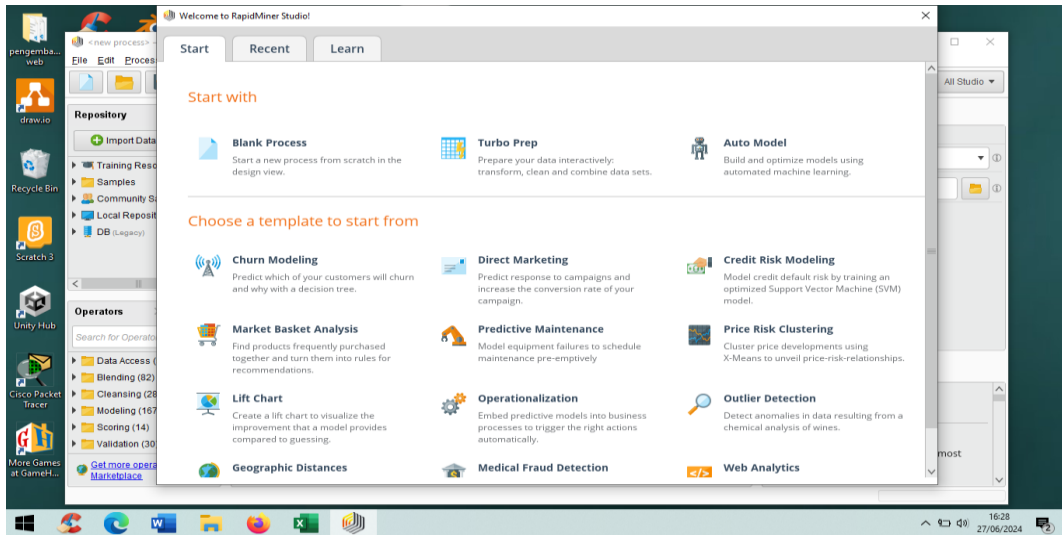
Berdasarkan dataset yang diperoleh dari hasil pengumpulan data, maka data yang digunakan adalah data dengan ciri-ciri nama siswa, nilai pengetahuan, nilai keterampilan, nilai UAS. Nama siswa merupakan siswa yang dikategorikan berdasarkan nilai, Sekolah merupakan atribut yang berisi 1 sekolah, nilai kenaikan bila diukur berisikan dari kelas 10 sampai kelas 12.

Clustering Menggunakan Perangkat Lunak RapidMiner.

Data yang diperoleh digunakan sebagai input untuk membangun model algoritma KMeans *Clustering* menggunakan perangkat lunak *RapidMiner*.

Tampilan Menu Utama *RapidMiner*

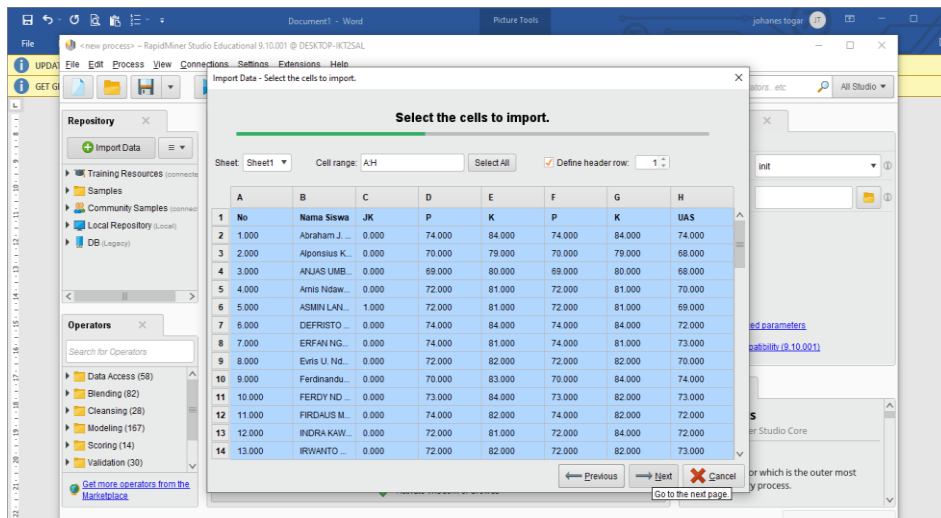
Menu utama *RapidMiner* memiliki tampilan awal yang mencakup berbagai komponen dan fitur. Salah satunya adalah "New Process", yang merupakan langkah pertama dalam menggunakan *Rapidminer*



Gambar 2. Tampilan Menu Utama RapidMiner Versi 9.9

Tampilan *Select The Cells To Import*

Sistem memberikan instruksi tentang bagaimana cara mengimpor data baru yang akan diolah selanjutnya, dengan format data saat ini dalam bentuk *Excel*. Tahapan ini terlihat dalam gambar berikut ini.

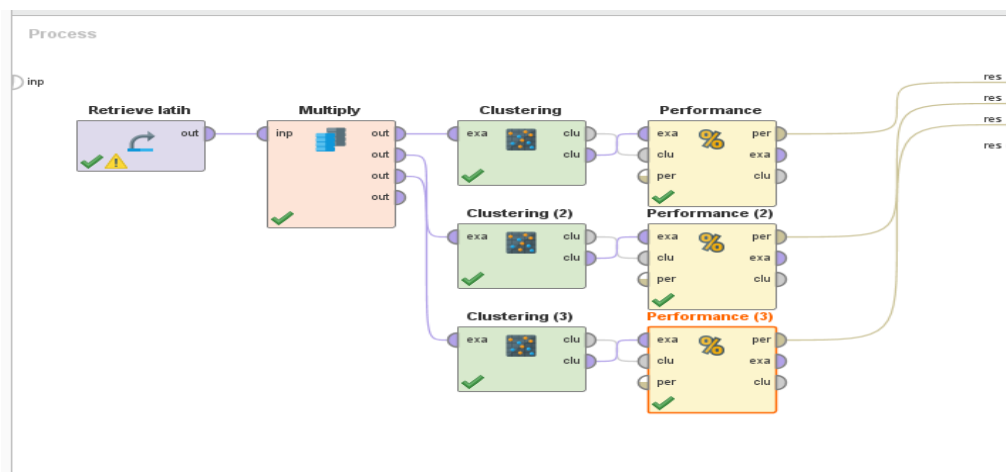


Gambar 3. Tampilan *Select The Cells To Import*

Tampilan *Clustering RapidMiner*

Algoritma *K-Means* akan diterapkan dalam proses yang telah dibentuk untuk melakukan klusterisasi atau pengelompokan. Operator klusterisasi yang digunakan adalah *KMeans*. Setelah pembuatan model selesai, proses akan dijalankan untuk mendapatkan hasil klusterisasi. Gambar

4 menunjukkan model proses terbaik yang menggunakan algoritma *KMeans*.

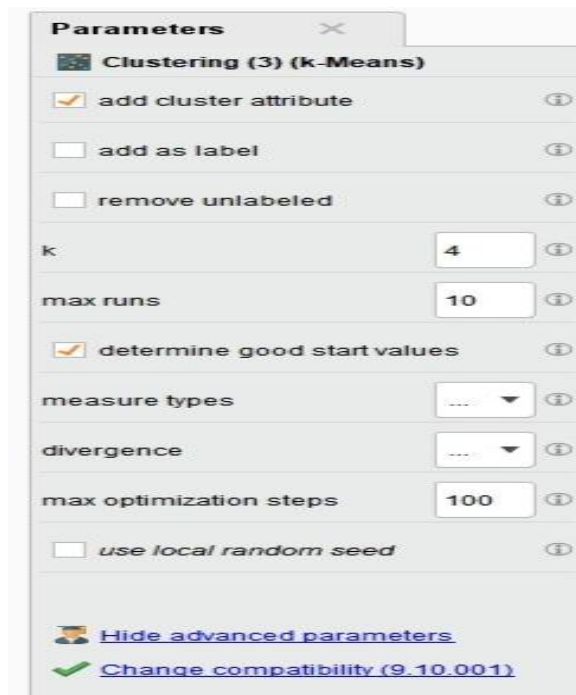


Gambar 4. Tampilan *Clustering RapidMiner*

Dari gambar diatas, menunjukkan beberapa operator seperti *Read Excel* dimana *dataset* akan dimasukkan dalam *Rad Excel*, *Multiplay* yang berfungsi untuk menjalankan satu atau lebih operator, operator *cluster* yaitu algoritma *k-means*, kemudian *performance* untuk evaluasi kinerja dari *k-means*.

Penentuan jumlah kluster

Langkah selanjutnya adalah melakukan pengaturan *algoritma K-Means* melalui menu "*Parameter Clustering K-Means*", berikut tampilan penentuan jumlah parameter kluster.



Gambar 5. Hasil jumlah *clustering*

Berdasarkan gambar 5, perlu memasukan nilai k yang diinginkan dalam literasi yang dilakukan sebanyak 10 kali, nilai k yang dimasukan adalah k=2,3,4. **Langkah-1: Pilih angka K untuk menentukan jumlah cluster.** Langkah-2: Pilih titik K atau centroid secara acak. Langkah-3: Tetapkan setiap titik data ke centroid terdekat, yang akan membentuk cluster K yang telah ditentukan. Langkah-4: Hitung varians dan tempatkan centroid baru dari setiap cluster. Tujuan literasi tersebut adalah untuk menemukan kluster dengan nilai DBI yang mendekati 0 atau yang paling kurang, maka akan dianggap sebagai kluster terbaik. Setelah literasi selesai, proses klusterisasi yang sudah dilakukan untuk mendapatkan hasil dari algoritma K -Means yang diimplementasikan melalui aplikasi Rapidminer. Setelah melalui beberapa literasi, yang ditemukan bahwa nilai optimal untuk jumlah kluster(k) adalah 2. Hasil klusterisasi akhir diperoleh menggunakan algoritma K-Means berdasarkan nilai K yang telah ditentukan tersebut.

Analisis kluster K-Means dilakukan untuk memisahkan siswa ke dalam kelompok-kelompok berdasarkan variabel-variabel pengetahuan dan keterampilan terkait nilai matematika seperti nilai pengetahuan, nilai keterampilan, nilai dan lain-lain. Analisis kluster dapat membantu mengidentifikasi nilai keterampilan nilai pengetahuan nilai ujian sekolah lainnya. Misalnya, kluster tertentu memiliki nilai pengetahuan dan nilai keterampilan nilai ujian sekolah tersebut.

Tabel 2. Anggota hasil kluster 0

AGREGAT	P1	K1	P2	K2	US
E					
MAX	78	84	81	76	75
MIN	72	78	72	70	67

RATA- RATA	73,63158	81,47368	74,73684	72,63158	72
SD	1,535163	1,645479	2,376788	1,60591	2,624669

Tabel 2 ini menunjukkan nilai agregat untuk *cluster* 0. Untuk *cluster* ini, p1 minimum adalah 72, sedangkan maksimum adalah 78. Rata-rata dari p1 di *cluster* ini adalah 73,63158 dan SD dari p1 adalah 1,535163, k1 minimum adalah 78, sedangkan maksimum adalah 84. Rata-rata dari p1 di *cluster* ini adalah 81,47368 dan SD dari k1 adalah 1,645479, p2 minimum adalah 72, sedangkan maksimum adalah 81. Rata-rata dari p2 di *cluster* ini adalah 74,73684 dan SD dari p2 adalah 2,376788, k2 minimum adalah 70, sedangkan maksimum adalah 76. Rata-rata dari k2 di *cluster* ini adalah 72,63158 dan SD dari k2 adalah 1,60591, US minimum adalah 67, sedangkan maksimum adalah 75. Rata-rata dari US di *cluster* ini adalah 72 dan SD dari US adalah 2,624669.

Tabel 1. *clustering* 1

AGREGATE	P1	K1	P2	K2	US
MAX	78	84	83	87	79
MIN	74	84	72	84	74
RATA- RATA	72,85714	79,71429	74,80952	82,2381	71,61905
SD	1,904881	3,848933	3,10836	1,972429	2,854403

Tabel 3 ini menunjukkan nilai agregat untuk *cluster* 1. Untuk *cluster* ini, p1 minimum adalah 74, sedangkan maksimum adalah 78. Rata-rata dari p1 di *cluster* ini adalah 72,85714 dan SD dari p1 adalah 1,904881, k1 minimum adalah 84, sedangkan maksimum adalah 84. Rata-rata dari p1 di *cluster* ini adalah 79,71429 dan SD dari k1 adalah 3,848933, p2 minimum adalah 72, sedangkan maksimum adalah 83. Rata-rata dari p2 di *cluster* ini adalah 74,80952 dan SD dari p2 adalah 3,10836, k2 minimum adalah 84, sedangkan maksimum adalah 87. Rata-rata dari k2 di *cluster* ini adalah 82,2381 dan SD dari k2 adalah 1,972429, US minimum adalah 74, sedangkan maksimum adalah 79. Rata-rata dari US di *cluster* ini adalah 71,61905 dan SD dari US adalah 2,854403.

Tabel 4. *clustering* 2

AGREGAT E	P1	K1	P2	K2	US
MAX	76	77	88	80	74
MIN	70	70	78	70	69
RATA- RATA	73,2	73,3	82,1	73	72,1
SD	2,20101	2,263233	2,806738	2,828427	1,72884

Tabel 4 ini menunjukkan nilai agregat untuk *cluster* 2. Untuk *cluster* ini, p1 minimum adalah 70 sedangkan maksimum adalah 76. Rata-rata dari p1 di *cluster* ini adalah 73,2 dan SD dari p1 adalah 2,20101, k1 minimum adalah 70, sedangkan maksimum adalah 77. Rata-rata dari p1 di *cluster* ini adalah 73,3 dan SD dari k1 adalah 2,263233, p2 minimum adalah 78,

sedangkan maksimum adalah 88. Rata-rata dari p2 di *cluster* ini adalah 82,1 dan SD dari p2 adalah 2,806738, k2 minimum adalah 70, sedangkan maksimum adalah 80. Rata-rata dari k2 di *cluster* ini adalah 73 dan SD dari k2 adalah 2,828427, US minimum adalah 69, sedangkan maksimum adalah 74. Rata-rata dari US di *cluster* ini adalah 71,61905 dan SD dari US adalah 1,72884.

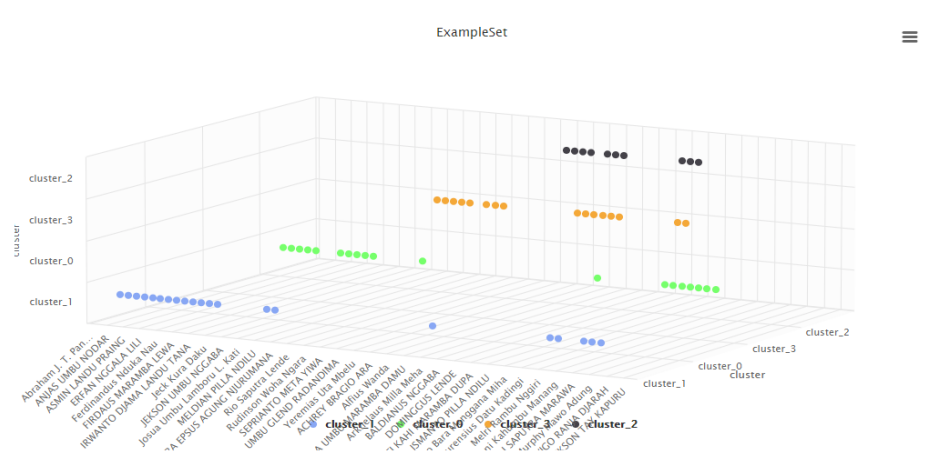
Tabel 2. *clustering 3*

AGREGATE	P1	K1	P2	K2	US
MAX	75	77	76	81	78
MIN	71	71	71	72	68
RATA-RATA	72,9375	72,875	73,125	74,75	71,875
SD	1,062623	1,408309	1,204159	2,645751	3,074085

Tabel 5 ini menunjukkan nilai agregat untuk *cluster 3*. Untuk *cluster* ini, p1 minimum adalah 71 sedangkan maksimum adalah 75 Rata-rata dari p1 di *cluster* ini adalah 72,9375 dan SD dari p1 adalah 1,062623, k1 minimum adalah 71, sedangkan maksimum adalah 77. Rata-rata dari p1 di *cluster* ini adalah 72,876 dan SD dari k1 adalah 1,408309, p2 minimum adalah 71, sedangkan maksimum adalah 76. Rata-rata dari p2 di *cluster* ini adalah 73,125 dan SD dari p2 adalah 1.204159, k2 minimum adalah 72, sedangkan maksimum adalah 81. Rata-rata dari k2 di *cluster* ini adalah 74,75 dan SD dari k2 adalah 2,828427, US minimum adalah 68, sedangkan maksimum adalah 78. Rata-rata dari US di *cluster* ini adalah 71,875 dan SD dari US adalah 3,074085.

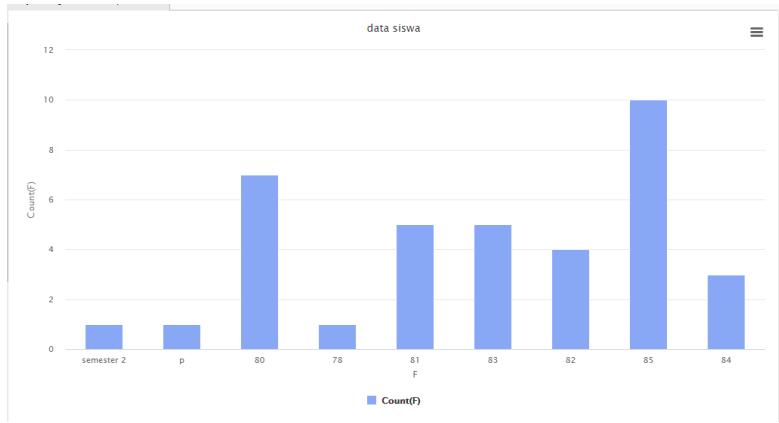
Visualisasi

Hasil analisis data dengan perangkat lunak Rapidminer menghasilkan beberapa output. Pada gambar visualisasi, terlihat penyebaran data yang menunjukkan jumlah cluster data nilai siswa kelas XII TKR di SMK Negeri 2 Waingapu. Visualisasi ini menampilkan sumbu x, dan sumbu y, warna khusus untuk cluster.



Gambar 6. Visualisasi nama dan cluster

Gambar 6 visualisasi nama dan cluster dari nama siswa sebagai variabel x dan kelompok cluster sebagai variabel y. Sehingga dibagian cluster 1 jumlah terbanyak, siswa yang memiliki nilai terendah.



Gambar 7. Bar plot

Gambar 7 merupakan Bar plot dari jumlah siswa variabel x dan nilai siswa sebagai variabel y. Sehingga siswa di bagian nilai 85 paling tinggi dan paling rendah di bagian nilai 75.



Gambar 8. folder view

Gambar 8 merupakan folder view data dari bagian-bagian kluster, dengan setiap anggota dari 4 cluster menampilkan id variabel. Setelah pengelompokan data nilai siswa menggunakan *Rapidminer*.

Evaluasi Hasil *Clustering*

Proses ini melibatkan tiga kali percobaan literasi untuk mendapatkan nilai DBI terbaik.

Tabel 3. Tabel *k-means*

k-means		
K	Avg. within centroid distance	DBI

2	45.821	1.381
3	35.893	1.258
4	26.376	0.994

Davies Bouldin

Davies Bouldin: -0.994

Gambar 9. DBI

Gambar 9 menampilkan informasi mengenai nilai DBI dan performa klusterisasi. Data ini diperoleh melalui pengujian dengan perhitungan *Davies-Bouldin Index* (DBI) menggunakan *RapidMiner*. Nilai DBI terbaik terdapat pada penggunaan nilai DBI sebesar 0,994 yang mendekati 0. Ini menunjukkan bahwa semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), semakin baik klusterisasi yang terbentuk. Hasil ini diperoleh dengan metode *K-Means* dalam klusterisasi yang dilakukan menggunakan *RapidMiner*.

Pengujian hasil klusterisasi dilakukan dengan menggunakan perhitungan *Davies-Bouldin Index* dengan jumlah *cluster* sebanyak 4. Pada dataset, nilai *Davies-Bouldin Index* yang diperoleh adalah 0,994. Berdasarkan hasil pengujian tersebut, dapat disimpulkan bahwa pada data nilai semester 1 tahun ajaran 2023/2024 terdapat kluster yang sesuai. Nilai pengujian sebesar 0,994 menunjukkan kekuatan *cluster* yang baik.

KESIMPULAN

Berdasarkan penelitian yang dilakukan di SMK Negeri 2 Waingapu dengan menggunakan metode *K-Means clustering*, dapat disimpulkan bahwa penerapan data mining ini efektif dalam mengelompokkan siswa berdasarkan nilai pengetahuan, keterampilan, dan sikap. Hasil *clustering* menunjukkan bahwa siswa dapat dibagi ke dalam 2 kelompok yang memiliki karakteristik nilai yang serupa, sehingga mempermudah sekolah dalam mengidentifikasi kebutuhan dan potensi masing-masing siswa. Selain itu, dari hasil analisis dengan *Davies-Bouldin Index* (DBI), nilai DBI terbaik adalah 0,994 dengan jumlah cluster sebanyak dua, yang menunjukkan kekuatan cluster yang baik. Penggunaan *K-Means clustering* dalam penelitian ini berhasil memberikan informasi yang mendalam mengenai pola prestasi siswa, yang dapat digunakan oleh sekolah untuk lebih memahami kinerja akademik siswa dan merancang strategi pembelajaran yang lebih personal dan efektif. Dari tabel-tabel aggregate value yang dihasilkan, dapat dilihat bahwa cluster 2 terdiri dari siswa dengan nilai pengetahuan dan keterampilan yang relatif seimbang di sekitar angka 72, sementara cluster 1 memiliki nilai pengetahuan dan keterampilan yang lebih tinggi dengan rata-rata sekitar 72. Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan tambahan dalam pengelompokan data nilai siswa untuk meningkatkan kualitas pembelajaran di SMK Negeri 2 Waingapu.

DAFTAR PUSTAKA

- Ahmad, I., Samsugi, S., & Irawan, Y. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Teknoinfo*, 16(1), 46. <http://portaldata.org/index.php/portaldata/article/view/107>
- Dacwanda, D. O., & Nataliani, Y. (2021). Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *Aiti*, 18(2), 125–138. <https://doi.org/10.24246/aiti.v18i2.125-138>
- Hutagalung, J. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(1), 606–620. <https://doi.org/10.35957/jatisi.v9i1.1516>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Nofitri, R., & Irawati, N. (2019). Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 5(2), 199–204. <https://doi.org/10.33330/jurteksiv5i2.365>
- Permadi, F. T., Nilogiri, A., & Rosyidah, U. A. (2023). Algoritma Fuzzy C-Means dan Metode Davies Bouldin Index (DBI) untuk Mengolompokkan Kabupaten/Kota di Jawa Timur Berdasarkan Sanitasi Layak, Air Minum Layak, dan Rumah Layak Huni Fuzzy C-Means Algorithm and Davies Bouldin Index (DBI) Method for Grouping Re. *Jurnal Smart Teknologi*, 4(4), 2774–1702. <http://jurnal.unmuhjember.ac.id/index.php/JST>
- Primanda, R. P., Alwi, A., & Mustikasari, D. (2021). DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri). *Komputek*, 5(1), 88. <https://doi.org/10.24269/jkt.v5i1.686>
- Ramadhani, R. D., & AK, D. J. (2017). Evaluasi K-Means dan K-Medoids pada Dataset Kecil. *Seminar Nasional Informatika Dan Aplikasinya, September*, 20–24.
- Winarta, A., & Kurniawan, W. J. (2021). Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python. *JTIK (Jurnal Teknik Informatika Kaputama)*, 5(1), 113–119. <https://doi.org/10.59697/jtik.v5i1.593>
- Ahmad, I., Samsugi, S., & Irawan, Y. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Teknoinfo*, 16(1), 46. <http://portaldata.org/index.php/portaldata/article/view/107>
- Dacwanda, D. O., & Nataliani, Y. (2021). Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *Aiti*, 18(2), 125–138. <https://doi.org/10.24246/aiti.v18i2.125-138>
- Nofitri, R., & Irawati, N. (2019). Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 5(2), 199–204. <https://doi.org/10.33330/jurteksiv5i2.365>
- Permadi, F. T., Nilogiri, A., & Rosyidah, U. A. (2023). Algoritma Fuzzy C-Means dan Metode Davies Bouldin Index (DBI) untuk Mengolompokkan Kabupaten/Kota di Jawa Timur Berdasarkan Sanitasi Layak, Air Minum Layak, dan Rumah Layak Huni Fuzzy C-Means Algorithm and Davies Bouldin Index (DBI) Method for Grouping Re. *Jurnal Smart Teknologi*, 4(4), 2774–1702. <http://jurnal.unmuhjember.ac.id/index.php/JST>
- Primanda, R. P., Alwi, A., & Mustikasari, D. (2021). DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri). *Komputek*, 5(1), 88. <https://doi.org/10.24269/jkt.v5i1.686>
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika*

Budidarma, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>
Winarta, A., & Kurniawan, W. J. (2021). Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python. JTIK (Jurnal Teknik Informatika Kaputama), 5(1), 113–119. <https://doi.org/10.59697/jtik.v5i1.593>