



Penerapan *K-Means Clustering* Untuk Mengelompokkan Performa Siswa Dalam Pelajaran Bahasa Indonesia

Application Of K-Means Clustering For Grouping Student Performance In Indonesian Language Lessons

Trovina Tige Kati¹, Reynaldi Thimotius Abineno², Arini Aha Pekuwali³

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Kristen Wira Wacana Sumba

Jl.R.Suprato No.35, Prailiu, Kec.Kota Waingapu, Kabupaten Sumba Timur, Nusa Tenggara Timur

Corresponding author: trovinakati2001@gmail.com

ABSTRACT

*The development of a country depends greatly on education. Waingapu State Vocational School 1 in East Sumba Regency is facing problem in management and utilization of value data students, which results in a lack of personalization learning. Study This aim for grouping student class XII based mark knowledge and skills with use method *K-Means Clustering*. Data used originate from report semester 1 and 2 grades of the year 2021/2022 class XII teachings. Analysis results show that student can grouped to in a number of group based on similar values, which makes things easier introduction needs and potential student. Performance $k=2$ with value 0.858, cluster 0 has 19 students and cluster 1 has 47 students, with testing *Davies-Bouldin Index*. Pattern found in study This is student with mark high knowledge and skills that is student with type Female gender in Cluster 0. Research results This showing that use algorithm *k-means* can help grouping student with high and low values to in more groups small.*

Keywords: *K-Means clustering, RapidMiner, Literacy, Indonesian.*

ABSTRAK

Pembangunan suatu negara sangat bergantung pada pendidikan. SMK Negeri 1 Waingapu di Kabupaten Sumba Timur menghadapi masalah dalam pengelolaan dan pemanfaatan data nilai siswa, yang berimbas pada kurangnya personalisasi pembelajaran. Penelitian ini bertujuan untuk mengelompokkan siswa kelas XII berdasarkan nilai pengetahuan dan keterampilan dengan menggunakan metode *K-Means Clustering*. Data yang digunakan berasal dari laporan nilai semester 1 dan 2 pada tahun ajaran 2021/2022 kelas XII. Hasil analisis menunjukkan bahwa siswa dapat dikelompokkan ke dalam beberapa kelompok berdasarkan nilai yang mirip, yang memudahkan pengenalan kebutuhan dan potensi siswa. Performa $k=2$ dengan nilai 0.858, *cluster 0* terdapat 19 siswa dan *cluster 1* terdapat 47 siswa, dengan pengujian *Davies-Bouldin Index*. Pola yang ditemukan dalam penelitian ini merupakan siswa dengan nilai pengetahuan dan keterampilan yang tinggi yaitu siswa dengan jenis kelamin Perempuan pada *Cluster 0*. Hasil penelitian ini menunjukkan bahwa penggunaan algoritma *k-means* dapat membantu mengelompokkan siswa dengan nilai yang tinggi dan yang rendah ke dalam kelompok yang lebih kecil.

Kata kunci: *K-Means clustering, RapidMiner, Literasi, Bahasa Indonesia.*

PENDAHULUAN

Salah satu bidang yang berperan penting dalam pembangunan suatu negara merupakan pendidikan. Tingkat keberhasilan dan kegagalan siswa yang tinggi mencerminkan kualitas standar pendidikan. Pendidikan adalah keseluruhan pengalaman belajar yang dialami setiap orang sepanjang hidupnya, yang tidak memiliki batas waktu dan berlangsung sepanjang usia, mulai dari usia dini, anak-anak, remaja, hingga dewasa. Demikian pula sebagai pengalaman belajar, pendidikan terjadi disemua lingkungan seperti lingkungan yang diciptakan oleh manusia seperti sekolah (Noor, 2018). Bahasa Indonesia adalah salah satu mata pelajaran yang penting untuk perkembangan akademik siswa.

Siswa yang senang membaca adalah salah satu contoh siswa yang berwawasan luas. Keberhasilan pendidikan tidak ditentukan oleh jumlah siswa dengan nilai tertinggi pada mata pelajaran tertentu, tetapi oleh jumlah siswa yang senang membaca atau memiliki kemampuan literasi yang baik (Utami & Yanti, 2022). Siswa yang memiliki kemampuan membaca dan menulis yang baik memiliki kemampuan untuk memahami dan menginterpretasikan informasi secara efektif. Kemampuan literasi juga merupakan dasar bagi keberhasilan dalam bidang studi lainnya, karena kemampuan ini mendukung pemahaman mereka tentang materi di berbagai disiplin ilmu.

Salah satu sekolah kejuruan di Waingapu, Sumba Timur adalah SMKN 1 Waingapu. Data internal sekolah menunjukkan variasi yang cukup signifikan dalam literasi siswa pada pelajaran Bahasa Indonesia. Sementara beberapa siswa menunjukkan kinerja yang luar biasa, yang lain menghadapi kesulitan dalam memahami dan menguasai materi. Sehingga intervensi yang tepat dapat digunakan untuk meningkatkan literasi secara keseluruhan, analisis yang mendalam diperlukan untuk menentukan kelompok siswa berdasarkan kinerja mereka.

Mengatasi masalah ini, penggunaan metode data mining seperti *k-means clustering* menjadi penting. Metode ini memungkinkan sekolah untuk mengelompokkan siswa berdasarkan prestasi akademik siswa dilihat dari nilai pengetahuan dan keterampilan. Dengan demikian, lebih mudah untuk menemukan kebutuhan khusus siswa dalam setiap kelompok dan membuat strategi pembelajaran yang lebih baik. Tujuan penelitian ini adalah untuk menerapkan *k-means clustering* pada data nilai siswa di SMK Negeri 1 Waingapu.

MATERI DAN METODE

Data Mining

Data mining adalah salah satu cara untuk melakukan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstrak dan mengidentifikasi pengetahuan dan informasi berguna dari *database* besar (Rizmayanti et al., 2021).

Data mining adalah metode untuk menemukan pengetahuan tersembunyi dan hubungan yang tidak terduga antara data. Teknik ini banyak digunakan diberbagai bidang seperti pemasaran, pengembangan produk, deteksi penipuan dan pendidikan. *Data mining* juga dikenal sebagai *Knowledge Discovery in Database (KDD)* adalah teknik menemukan informasi berguna yang sebelumnya tidak diketahui dalam kumpulan data yang sangat besar (Naibaho & Zahra, 2023).

Clustering

Data mining menggunakan *clustering* untuk membedakan kumpulan data menjadi beberapa kelompok kumpulan data dengan dilihat dari kemiripan data. Dalam *data mining*, *clustering* melibatkan kumpulan objek atau data dalam kelompok, setiap kelompok memiliki

data yang hampir sama atau dengan kata lain memiliki kemiripan data dan dapat dibedakan dari objek di kelompok lain. (Saputra & Nataliani, 2021).

Clustering digunakan untuk menyatukan data yang serupa dalam satu kelompok dan mengurangi jarak antar kelompok. Ini menunjukkan bahwa objek-objek dalam kelompok tidak sama dengan objek-objek kelompok atau *cluster* lain, tetapi sangat mirip satu sama lain (Nurul Alifia et al., 2024).

Karena data yang belum diolah sangat besar dan sulit untuk dianalisis atau dipelajari oleh karena itu pengelompokan (*clustering*) dilakukan. Tujuan pengelompokan data untuk meningkatkan pemahaman tentang data dan menganalisis kualitas dari data, algoritma *k-means* merupakan metode *cluster* yang paling umum digunakan dalam penelitian ilmiah dan penelitian lainnya (Melpa Metisen & Latipa Sari, 2015).

Penilaian

Proses pengumpulan data untuk tujuan mengukur hasil belajar siswa dikenal sebagai penilaian. Pendidikan harus melakukan penilaian untuk menentukan tingkat pengetahuan siswa mereka. Penilaian pencapaian kompetensi keterampilan dilakukan untuk mengetahui seberapa baik siswa mencapai SKL, KL, dan KD khususnya dalam dimensi keterampilan, penilaian ini mencakup sikap, keterampilan, dan pengetahuan yang digunakan secara proporsional untuk menentukan posisi relatif setiap siswa (Kurniati et al., 2019).

Dalam aktivitas seperti menggunakan, mengurai, merangkai, memodifikasi, dan membuat, dimensi keterampilan dinilai. Namun, keterampilan dalam kehidupan sehari-hari mencakup hal-hal seperti menulis, menghitung, membaca dan mengarang (Kurniati et al., 2019). Penilaian pengetahuan yang terdiri dari ujian tertulis, penjelasan singkat, penjadohan, dan pilihan ganda adalah salah satu bentuk penilaian yang paling umum dilakukan oleh guru.

Metode K-Means

k-means merupakan algoritma *clustering* yang sangat sederhana yang membagi dataset menjadi beberapa bagian. *Algoritma* ini berbasis jarak data dan memiliki pusat *cluster* terdekat atau *centroid*, *algoritma* ini juga mudah pada saat digunakan dan relatif cepat, mudah dipilih dan banyak digunakan (Sibuea & Safta, 2017).

Davies-Bouldin Index

Pahlevi Kurniawan, (2023) menyatakan salah satu cara untuk menilai kelompok dalam metode pengelompokan adalah *Davies-Bouldin Index (DBI)*, yang didasarkan pada nilai kohesi dan separasi. Kohesi adalah jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti, sedangkan separasi adalah jarak antar *centroid* dari *cluster*nya.

RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka. *RapidMiner* menawarkan analisis *data mining*, *text mining*, dan prediksi, berbagai teknik deskriptif dan prediksi membantu pengguna membuat keputusan yang lebih baik (Ranjawali et al., 2023). Lebih dari lima ratus operator data mining tersedia di *RapidMiner*, termasuk operator untuk *input*, *output*, data *preprocessing*, dan visualisasi, *RapidMiner* adalah program analisis data yang berdiri sendiri dan juga mesin *data mining* yang dapat diintegrasikan pada produknya sendiri dan *RapidMiner* dapat dijalankan di semua sistem operasi karena dibuat menggunakan java (Sibuea & Safta, 2017).

Gambar 1 dibawah ini merupakan alur penelitian yang akan dilakukan.



Gambar 1 Alur Penelitian

Pada tahapan pertama dilakukan pengumpulan data, pada tahap ini dilakukan pengumpulan data siswa yang diambil langsung dari SMKN 1 Waingapu melalui tahapan wawancara dengan data yang didapatkan yaitu data nilai bahasa indonesia kelas 12 semester 1 dan semester 2, untuk data siswa yang sama pada 2 kelas di jurusan akuntansi dengan jumlah 66 siswa, dengan kategori penilaian pengetahuan dan keterampilan.

Pada tahap kedua akan dilakukan *Pre-processing* Data, pada tahap ini dilakukan pembersihan data atau normalisasi data. Dengan cara data yang tidak utuh atau hilang akan dihapus dari data siswa tersebut, dan pemilihan atribut yang akan digunakan sebagai *dataset*.

Tahap ketiga yaitu pengolahan data, *dataset* yang telah dibuat akan diolah menggunakan aplikasi *RapidMiner* agar menghindari perhitungan yang salah. Selanjutnya hasil kelompokan akan diperiksa dengan menggunakan tabel perbandingan nilai (*aggregate value*) yaitu nilai *min-max*, nilai rata-rata dan standar deviasi. Tujuan dari analisis ini adalah untuk menemukan pola, hubungan dan pengetahuan yang akan digunakan untuk melihat tingkat literasi dalam pelajaran Bahasa Indonesia.

Visualisasi, hasil analisis yang diperoleh dari pengelompokan sebelumnya akan divisualisasikan, sehingga mempermudah peneliti dalam melakukan visualisasi dengan tabel dan grafik menggunakan.

Evaluasi, langkah terakhir dari penelitian ini adalah mengevaluasi metode *clustering* menggunakan *davies-bouldin index* untuk mengukur seberapa efektif teknik tersebut digunakan. Dilihat dari 2 parameter yaitu kohesi atau separasi, untuk menentukan tingkat akurasi model yang dibangun.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Pengumpulan data dilakukan di SMK Negeri 1 Waingapu terkait data nilai siswa kelas XII Akuntansi A dan Akuntansi B, jenis data yang diperoleh berupa format *excel* yang didapatkan dari SMK Negeri 1 Waingapu. Penelitian ini menggunakan data pada tahun ajaran 2021/2022 semester 1 (ganjil) dan semester 2 (genap) dengan kategori penilaian pengetahuan dan keterampilan, adapun jumlah data siswa yang digunakan sebanyak 66 siswa. Pengolahan data dilakukan dengan dataset yang sudah disiapkan.

Pre-processing Data

Preprocessing data sangat penting karena setelah pengumpulan data, perlu dilakukan pengecekan kembali dan dilakukan inialisasi data. Tabel berikut menunjukkan *dataset* yang telah disiapkan untuk pengolahan *clustering* dengan aplikasi *RapidMiner*.

Tabel 1 *Dataset*

Nama Siswa	JK	Nilai Bahasa Indonesia			
		Semester 1		Semester 2	
		P	K	P	K
Siswa 1	0	79	77	80	77
Siswa 2	1	81	77	76	77
Siswa 3	1	78	77	86	77

Siswa 4	0	77	77	86	77
Siswa 5	1	77	77	83	77
....
Siswa 65	0	75	75	76	76
Siswa 66	0	75	75	76	75

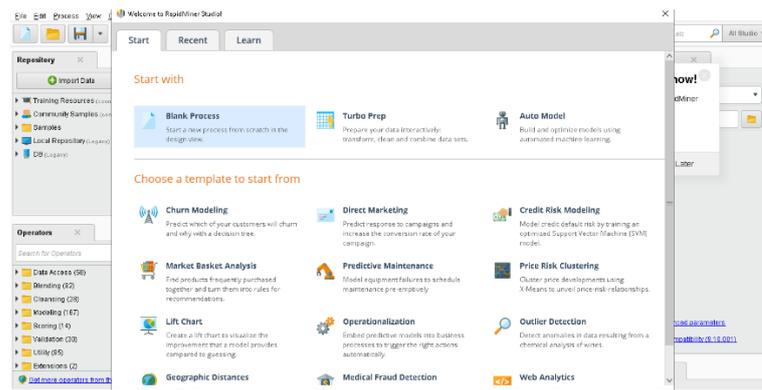
Berdasarkan *dataset* yang telah dibuat, maka data yang digunakan adalah data dengan atribut nama siswa, Jenis kelamin, nilai pengetahuan dan nilai keterampilan pada semester 1 dan nilai pengetahuan dan keterampilan pada semester 2. Dengan jumlah data siswa sebanyak 66 siswa.

Pengolahan Data Dengan *RapidMiner*

Untuk memudahkan data yang terjadi selama mengelompokkan data siswa SMK Negeri 1 Waingapu, aplikasi *RapidMiner* digunakan untuk melakukan proses *k-means clustering*. Parameter seperti jumlah anggota *cluster*, nilai *centroid*, dan daftar data yang dibuat untuk setiap *cluster* digunakan untuk memvalidasi data yang akan diolah. Beberapa langkah harus dilakukan sebelum data dapat diolah:

Tampilan Utama *RapidMiner*

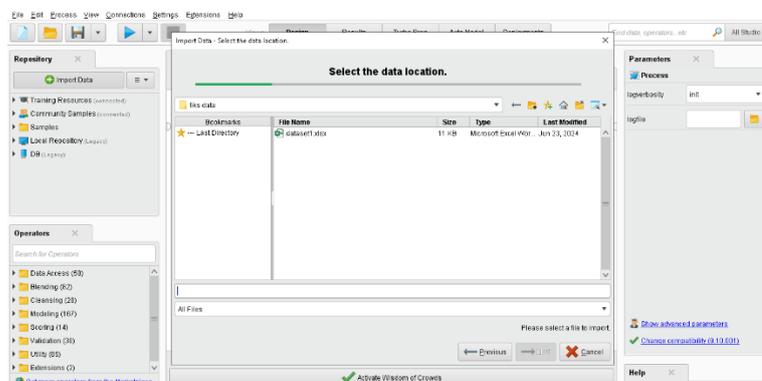
Tampilan utama *RapidMiner* memiliki tampilan awal yang mencakup berbagai fitur menu. Salah satunya yaitu *New Process*, yang merupakan langkah pertama dalam menggunakan aplikasi *RapidMiner*.



Gambar 2 Tampilan awal *RapidMiner*

Tampilan *Import Data*

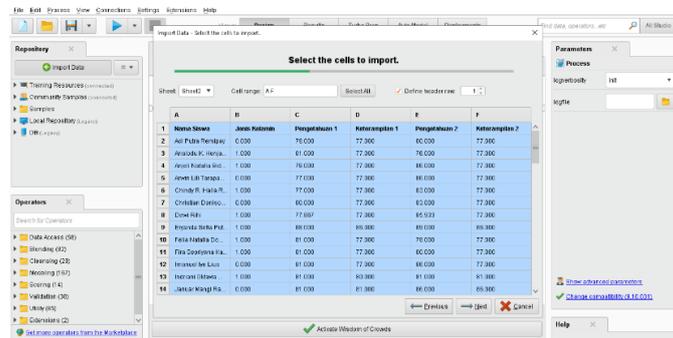
Langkah kedua mengimport data pada bagian menu *repository* dengan mengklik “*Add Data*” untuk memasukkan data yang akan diolah.



Gambar 3 Tampilan *Import Data*

Tampilan *Select The Cells To Import*

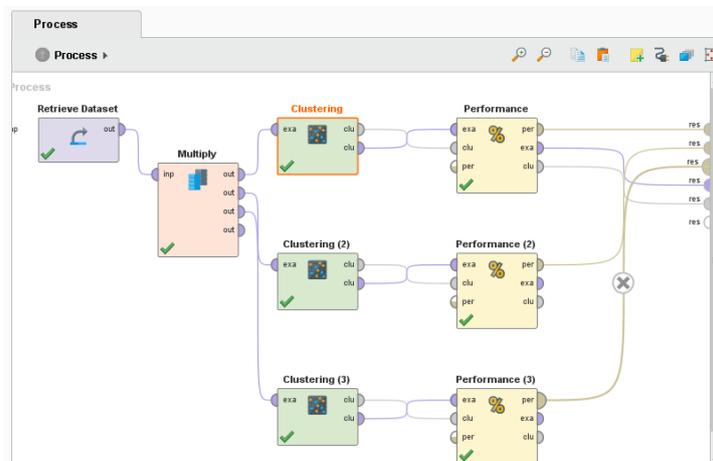
Tampilan *RapidMiner* petunjuk tentang bagaimana cara mengimpor data baru untuk diproses lanjut pada tahap selanjutnya, dengan format data yang dipakai dalam bentuk *Excel*. Tahapan ini terlihat dalam gambar dibawah ini.



Gambar 4 *Selection Cells To Import*

Tampilan *Clustering Rapidminer*

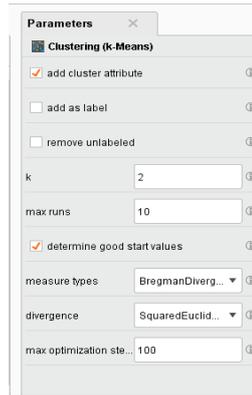
Dalam proses yang di rancang untuk melakukan pengelompokan atau *clustering*, algoritma *k-means* akan digunakan. *K-means* adalah operasi *clustering* proses akan dilakukan untuk mendapatkan hasil *clustering* setelah pembuatan model selesai. Model proses optimal untuk algoritma *k-means* ditunjukkan dalam gambar 5. Gambar dibawah ini menunjukkan operator seperti *read Excel*, yang memungkinkan *dataset* dimasukkan ke dalamnya, *multiply* yang dapat menjalankan lebih dari satu operator, algoritma *k-means* dan *performance* digunakan untuk mengevaluasi kinerja *k-means*.



Gambar 5 Desain Model *clustering*

Menentukan jumlah *Cluster*

Selanjutnya, menu parameter *clustering k-means* dapat digunakan untuk melakukan pengaturan algoritma *k-means* dan menentukan jumlah *cluster* yang akan digunakan. Tampilan gambar penentuan jumlah parameter *cluster* dibawah ini.



Gambar 6 Parameter *Clustering*

Nilai K yang diinginkan harus dimasukkan pada gambar 6. Nilai K yang dimasukkan untuk menguji performa siswa dilakukan 3 kali *clustering* dengan nilai $k = 2, 3, \text{ dan } 4$ dalam iterasi yang dilakukan sebanyak 10 kali. Tujuan iterasi dilakukan yaitu untuk menemukan *cluster* dengan nilai *DBI* yang mendekati 0 atau yang paling kurang, maka dianggap sebagai *cluster* terbaik. Setelah iterasi selesai, proses *clustering* dilakukan untuk mendapatkan hasil dari algoritma *k-means* yang digunakan oleh aplikasi *RapidMiner*. Setelah beberapa iterasi dilakukan diketahui bahwa jumlah *cluster* (k) yang baik adalah 2. Nilai k ini digunakan oleh algoritma *k-means* untuk mendapatkan hasil akhir.

Tabel Hasil *Clustering*

Analisis *cluster k-means* dilakukan untuk memisahkan siswa kedalam kelompok-kelompok berdasarkan variabel-variabel terkait nilai siswa seperti nilai semester 1 dan nilai semester 2 pada kelas XII dengan kategori penilaian pengetahuan dan keterampilan. Misalnya *cluster* tertentu memiliki nilai pada kategori pengetahuan dan keterampilan yang rendah, menunjukkan adanya masalah dalam performa belajar siswa pada pelajaran Bahasa Indonesia dikelompok tersebut. Dengan mengelompokkan nilai siswa berdasarkan kelompok, dapat membantu mengidentifikasi faktor-faktor penyebab penurunan nilai siswa. Setelah dilakukan proses *clustering*, data tersebut dikelompokkan menjadi 2 *cluster* yang diberi nama *cluster 0* dan *cluster 1*. Hasil *cluster* pada setiap nilai siswa ditampilkan pada bentuk Tabel dibawah ini.

Tabel 2 Hasil *Clustering*

Id	Nama Siswa	cluster	Jenis Kelamin	Nilai Bahasa Indonesia			
				P1	K1	P2	K2
1	Siswa 1	cluster_1	0	79	80	77	77
2	Siswa 2	cluster_1	1	81	76	77	77
3	Siswa 3	cluster_1	1	78	86	77	77
....
65	Siswa 65	cluster_0	1	88	86	89	86
66	Siswa 66	cluster_0	0	81	81	86	86

Menunjukkan hasil perhitungan menggunakan aplikasi *RapidMiner*, yang terbagi menjadi 2 *cluster*, didapatkan 2 *cluster* karena disesuaikan dengan performa terbaik yang diperoleh oleh masing-masing siswa. *Cluster 0* terdiri dari 19 data siswa dengan nilai yang diperoleh tinggi, sedangkan *cluster 1* terdiri dari 47 data siswa dengan nilai yang diperoleh rendah. Pada proses *clustering*, data dikelompokkan kedalam beberapa *cluster* berdasarkan kemiripan tertentu. Setelah *cluster* terbentuk, selanjutnya yaitu perhitungan *aggregation Value*, seperti nilai rata-rata pengetahuan dan keterampilan. Perhitungan ini membantu dalam memahami karakteristik umum dari setiap *cluster*.

Tabel 3 *Aggregate Value*

MAX				
Cluster	P1	K1	P2	K2
0	88	88	93	89
1	85	83	86	84
MIN				
Cluster	P1	K1	P2	K2
0	81	78	77	76
1	75	75	75	75
RATA-RATA				
Cluster	P1	K1	P2	K2
0	85	83	83	83
1	77	77	78	77
STANDAR DEVIASI				
Cluster	P1	K1	P2	K2
0	2.73	2.52	3.91	3.36
1	3.05	2.43	3.30	2.05

Nilai total *aggregate value* pada *cluster 0* dan 1 dapat dilihat pada tabel diatas. Pada tabel diatas terdapat nilai P1 dan K1 yang merupakan nilai pengetahuan dan keterampilan pada semester 1 dan nilai P2 dan K2 merupakan nilai pengetahuan dan keterampilan pada semester 2, dikelas XII. Berdasarkan nilai *aggregate value* diatas nilai tertinggi (*Max*) terdapat pada *cluster 0*, dilihat pada variabel P1 = 88, K1 = 88, P2 = 93, dan K2 = 89. Nilai terendah (*Min*) terdapat pada *cluster 1*, dilihat pada variabel P1 = 75, K1 = 75, P2 = 75, dan K2 = 75. Nilai rata-rata (*mean*) dari *cluster 0* yaitu P1 = 85, K1 = 83, P2 = 83, dan K2 = 83. Sedangkan nilai rata-rata yang paling rendah terdapat pada *cluster 1* yaitu P1 = 77, K1 = 77, P2 = 78, dan K2 = 77. Dari data diatas dapat dilihat nilai pada pengetahuan dan keterampilan pada *cluster 0* lebih baik dibandingkan dengan *cluster 1*, dengan nilai rata-rata 83. *Cluster 0*

lebih baik dikarenakan nilai tertinggi terdapat pada *cluster* 0 dilihat dari standar deviasi yang menunjukkan rentan nilai yang bervariasi dibandingkan *cluster* 1.

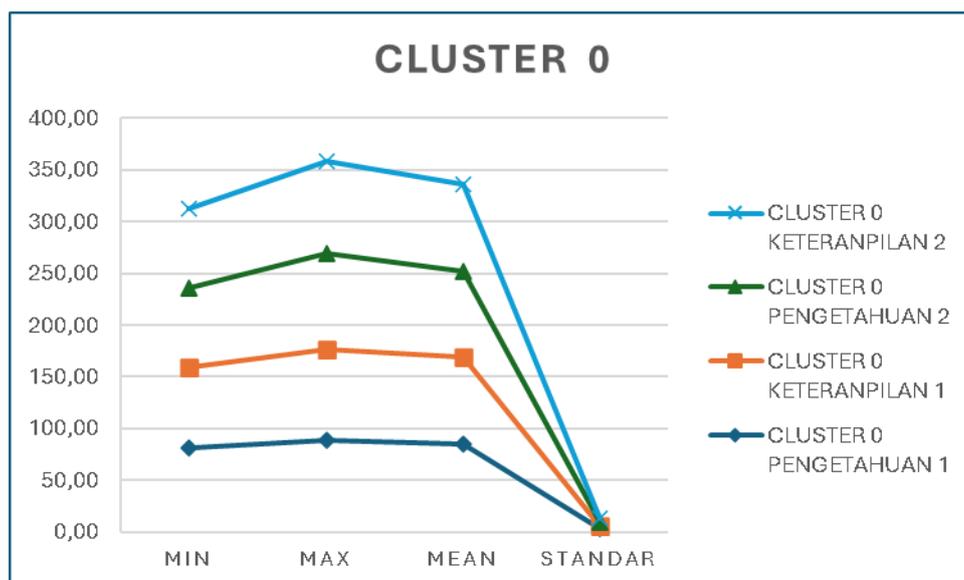
Visualisasi Hasil *Clustering*

Perangkat lunak *RapidMiner* menghasilkan berbagai *output* sebagai hasil dari pengujian data. Ada garis yang menunjukkan jumlah kelompok *cluster* untuk nilai siswa di SMK Negeri 1 Waingapu dalam gambar visualisasi. Gambar dibawah ini menunjukkan sumbu x, sumbu y, dan warna kustom (*cluster*), yang terlihat seperti gambar dibawah ini



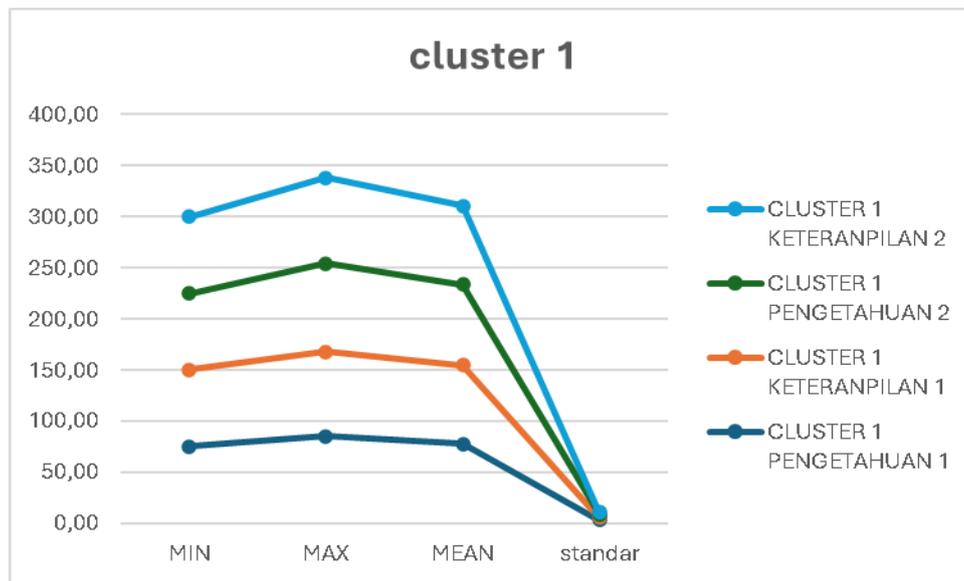
Gambar 7 *Scatter/Bubble* Jenis kelamin (*x*), *Cluster* (*y*)

Gambar 7 merupakan gambar visualisasi *Scatter* pada *RapidMiner* dengan jenis kelamin sebagai variabel (*x*), jenis kelamin 0 menandakan laki-laki sedangkan 1 menandakan perempuan dan *cluster* sebagai variabel (*y*), *cluster* 1 berwarna biru sedangkan *cluster* 0 berwarna hijau. Sehingga dapat dilihat bahwa nilai paling tinggi berada pada *cluster* 0 berdasarkan jenis kelamin 1 yaitu perempuan. Begitu juga pada gambar 8 dan 9 *cluster* 0 dan *cluster* 1.



Gambar 8 Grafik *Excel* *cluster* 0

Gambar 8 menunjukkan gambar grafik pada cluster 0 yang divisualisasikan menggunakan excel. Dapat dilihat pada gambar diatas terdapat nilai pengetahuan dan keterampilan yang cukup bervariasi. Dimana pada *cluster 0* terdapat nilai terendah (*Min*) P1=81, K1=78, P2=77, dan K2=76. Nilai tertinggi (*Max*) P1=88, K1=88, P2=93, dan K2=89. Nilai rata-rata (*Mean*) P1=85, K1=83, P2=83, dan K2=83. Dengan standar deviasi pada P1=2.73, K1=2.52, P2=3.91, dan K2=3.36. Dari standar deviasi diatas dapat dilihat bahwa nilai pada P2 dan K2 lebih bervariasi di dibandingkan nilai pada P1 dan K1.



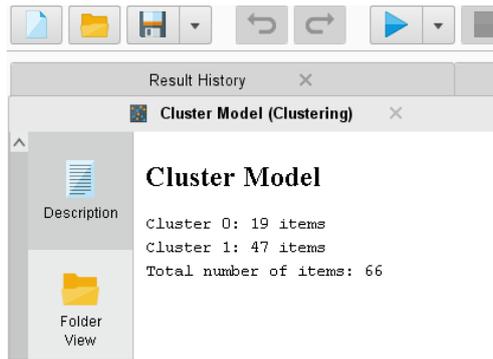
Gambar 9 Grafik Excel Cluster 1

Gambar 9 merupakan hasil visualisasi menggunakan grafik excel pada cluster 1. Dimana pada cluster 1 nilai terendah (*Min*) P1=75, K1=75, P2=75, Dan K2=75. Nilai tertinggi (*Max*) P1=85, K2=83, P2=86, dan K2=84. Nilai rata-rata (*Mean*) P1=77, K1=77, P2=78, dan K2=77. Dengan standar deviasi P1=3.05, K1=2.43, P2=3.30, dan K2=2.05. Dari standar deviasi pada *cluster 1* dapat dilihat P1 dan P2 memiliki nilai yang lebih bervariasi dibandingkan dengan nilai pada K1 dan K2.

Gambar 7, gambar 8 dan gambar 9 menunjukkan hasil akhir dari proses visualisasi pengelompokan data *cluster* yang dilakukan dengan menggunakan *RapidMiner* dan *excel*. Berdasarkan hasil pengelompokan, dapat disimpulkan bahwa 19 siswa dalam *cluster 0* diberi tanda hijau, dan 47 siswa dalam *cluster 1* diberi tanda biru. Kemudian, hasil *clustering* divisualisasikan menggunakan *scatter/Bubble* dan *grafik* pada *excel*, seperti yang ditunjukkan pada gambar diatas. Dari gambar visualisasi diatas dapat disimpulkan bahwa nilai Bahasa Indonesia dengan kategori penilaian pengetahuan dan keterampilan, nilai lebih tinggi terdapat pada jenis kelamin 1 yaitu perempuan pada *cluster 0*.

Cluster Model

Dalam model *cluster (clustering)*, terdapat beberapa tampilan yang menampilkan hasil *cluster*, salah satunya adalah tampilan teks (*text view*) yang menunjukkan hasil pengelompokan berdasarkan *cluster* dan jumlah anggota dalam setiap *cluster*. *cluster 0* memiliki 19 siswa, sedangkan *cluster 1* memiliki 47 siswa. Tampilan *cluster model* dapat dilihat pada gambar dibawah ini.



Gambar 9 Cluster Model



Gambar 10 Folder View

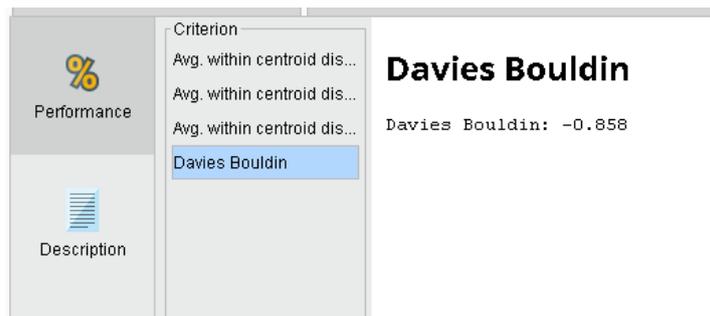
pada gambar 10 terdapat gambar folder view yang terdapat *cluster 0* dan *cluster 1* dan setiap *cluster 0* dan *cluster 1* menampilkan id siswa, setelah dilakukan pengelompokan nilai siswa berdasarkan performa siswa dengan menggunakan aplikasi *RapidMiner*.

Evaluasi Hasil Clustering

Dengan menggunakan metode *k-means clustering* dan didukung oleh perhitungan *Davies-Bouldin Index (DBI)*, dilakukan tiga iterasi percobaan untuk menemukan *cluster* terbaik berdasarkan nilai *DBI* terbaik, seperti yang ditunjukkan dalam tabel 4 dibawah ini merupakan tabel nilai $K=2$ yang terpilih.

Tabel 4 Nilai *Davies-Bouldin Index (DBI)*

K	Avg. Within cenroid distance	DBI	Terpilih
2	32.096	0.858	K=2
3	21.886	0.932	
4	15.889	0.935	



Gambar 11 Nilai *Davies-Bouldin Index* Iterasi Ke-2

Performa *Davies-Bouldin* dapat dilihat pada gambar 11. Untuk mendapatkan hasil performa, *RapidMiner* digunakan untuk menguji perhitungan *Davies-Bouldin Index (DBI)*. Hasil ditemukan ketika $k=2$, dengan nilai *DBI* sebesar 0.858. Hasil yang didapatkan masih terbilang tinggi dikarenakan data nilai yang digunakan kurang bervariasi dan memiliki cukup banyak nilai ganda, dapat dilihat pada tabel 3 *aggregate value* standar deviasi.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, hasil pengujian data menggunakan *tools RapidMiner* pada pengelompokan performa data nilai siswa di SMK Negeri 1 Waingapu tahun ajaran 2021/2022 menggunakan metode *k-means clustering* dengan variabel jenis kelamin, nilai pengetahuan dan keterampilan. Sehingga menghasilkan 2 *cluster*, dimana *cluster* dengan nilai rendah pada *cluster* 1 yang terdiri dari 47 siswa, dengan nilai rata-rata (*mean*) $P1=77$, $K1=77$, $P2=78$, $K2=77$ dan *cluster* 0 adalah kelompok dengan tingkat nilai tinggi, terdiri dari 19 siswa, dengan rata-rata (*mean*) $P1=85$, $K1=83$, $P2=83$, $K2=83$ dilihat dari tabel *aggregate value*. Berdasarkan pengujian menggunakan *Davies-Bouldin Index (DBI)*, perhitungan dengan *tools RapidMiner* sehingga menghasilkan nilai evaluasi performa $K=2$ memiliki nilai 0.858. Sehingga dari hasil *clustering* ini dapat mempermudah pihak sekolah dalam mengidentifikasi kebutuhan dan potensi masing-masing siswa. Dengan demikian diharapkan bahwa penelitian ini akan memberikan informasi tambahan tentang bagaimana performa siswa dikelompokkan dalam Pelajaran Bahasa Indonesia sehingga dapat meningkatkan kualitas pembelajaran di SMK Negeri 1 Waingapu.

SARAN

Berdasarkan hasil penelitian, perlu dilakukan pengembangan *k-means clustering* lebih lanjut dengan menambahkan jumlah variabel data yang digunakan dengan variasi nilai yang lebih banyak untuk mendapatkan akurasi yang lebih baik lagi seperti nilai Pelajaran Bahasa Indonesia dan banyak kelas yang digunakan. Penting juga untuk mempertimbangkan pendekatan yang lebih rinci dari penelitian berikutnya. Hal ini dapat dilakukan dengan menggunakan data yang dibagi berdasarkan kabupaten atau kecamatan/kota.

DAFTAR PUSTAKA

- Kurniati, N., Khaliq, A., & Bulan, A. (2019). Penilaian Sikap , Pengetahuan dan Keterampilan Dalam Pembelajaran Bahasa Inggris yang Berorientasi Kurikulum 2013. *Seminar Nasional Taman Siswa Bima Tahun 2019, 2013*, 309–316.
- Melpa Metisen, B., & Latipa Sari, H. (2015). Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2), 110–118.
- Naibaho, A. F. A., & Zahra, A. (2023). Prediksi Kelulusan Siswa Sekolah Menengah Pertama Menggunakan Machine Learning. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3). <https://doi.org/10.23960/jitet.v11i3.3056>
- Noor, T. (2018). RUMUSAN TUJUAN PENDIDIKAN NASIONAL Pasal 3 UNDANG-UNDANG SISTEM PENDIDIKAN NASIONAL NO 20 TAHUN 2003. *Wahana Karya Ilmiah Pendidikan*, 2(01), 123–144.
- Nurul Alifia, A., Fahrudi Setiawan, A., & Rudhistiar, D. (2024). Penerapan Algoritma K-Means Clustering Dalam Peringatan Dini Resiko Kegagalan Siswa Pada Mata Pelajaran Bahasa Indonesia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1174–1181. <https://doi.org/10.36040/jati.v8i2.9075>
- Pahlevi Kurniawan, R. (2023). Penerapan Algoritme K-Means Clustering Untuk Mengelompokkan Siswa Berdasarkan Nilai Akademik Di Smp Negeri 207 Ssn. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 2(2), 530–538.
- Ranjawali, R., Talakua, A. C., & Abineno, R. T. (2023). *CLUSTERING STUNTING PADA BALITA DENGAN METODE K- MEANS DI PUSKESMAS KANATANG*. 80–92.
- Rizmayanti, A. I., Hidayati, N., Nugraha, F. S., & Gata, W. (2021). Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decision Tree (Studi Kasus Smk Multicomp Depok). *Swabumi*, 9(1), 9–18. <https://doi.org/10.31294/swabumi.v9i1.8363>
- Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of Information Systems and Informatics*, 3(3), 424–439. <https://doi.org/10.51519/journalisi.v3i3.164>
- Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurteksi*, 4(1), 85–92. <https://doi.org/10.33330/jurteksi.v4i1.28>
- Utami, N. P., & Yanti, P. G. (2022). Pengaruh Program Literasi terhadap Hasil Belajar Bahasa Indonesia Siswa Sekolah Dasar. *Jurnal Basicedu*, 6(5), 8388–8394. <https://doi.org/10.31004/basicedu.v6i5.3825>