



Universitas Kristen Wira Wacana Sumba

Fakultas Sains dan Teknologi

SATI: Sustainable Agricultural Technology Innovation

Homepage: <https://ojs.unkriswina.ac.id/index.php/semnas-FST>

3rd Nasional Seminar on Sustainable Agricultural Technology Innovation

2 Agustus 2024/ Pages: 523-537

PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PERFORMA SISWA PADA PEMBELAJARAN BAHASA INDONESIA(Studi Kasus: SD INPRES WAINGAPU 3)

Aplication Of The K-Means Clustering Algorithm For Grouping Student Performance In Indonesia Language Learning (Case Study: SD INPRES WAINGAPU 3)

Natalia Bili¹, Reynaldi Thimotius Abineno² dan Arini Aha Pekuwali³

^{1,2,3}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:

lhya389@gmail.com¹, reynaldi@unkriswina.ac.id², arini.pekuwali@unkriswina.ac.id³

ABSTRACT

In an effort to improve the language literacy of students at SD Inpres Waingapu 3, this research applies the K-Means Clustering algorithm to group student performance based on their knowledge and skills scores in the Indonesian Language subject. The research question addressed is how to perform clustering to improve language literacy using the K-Means method that utilizes students' knowledge and skills scores. This research aims to group students into several clusters based on their performance, so that more targeted learning strategies can be designed and implemented according to the needs of each group. The research results show that using the DBI method, the optimal number of clusters is determined to be three. The clustering process conducted with RapidMiner resulted in three groups of students with different performance characteristics, which can be used as a basis to design more effective and specific educational interventions.

Keywords: *K-Means Clustering, Indonesian Language Learning, Value*

ABSTRAK

Dalam upaya meningkatkan literasi bahasa siswa di SD Inpres Waingapu 3, penelitian ini menerapkan algoritma *K-Means Clustering* untuk mengelompokkan performa siswa berdasarkan nilai pengetahuan dan keterampilan mereka dalam mata pelajaran Bahasa Indonesia. Rumusan masalah yang diangkat adalah bagaimana melakukan clustering untuk meningkatkan literasi bahasa dengan menggunakan metode *K-Means* yang memanfaatkan nilai pengetahuan dan keterampilan siswa. Penelitian ini bertujuan untuk mengelompokkan siswa pada beberapa *cluster* berdasarkan performa mereka sehingga strategi pembelajaran yang lebih tepat sasaran dapat dirancang dan diterapkan sesuai kebutuhan setiap kelompok. Hasil penelitian menunjukkan bahwa dengan menggunakan metode DBI, jumlah *cluster* optimal ditentukan sebanyak tiga *cluster*. Proses *clustering* yang dilakukan dengan RapidMiner menghasilkan tiga kelompok siswa dengan karakteristik performa yang berbeda-beda, yang dapat dijadikan dasar untuk merancang intervensi pendidikan yang lebih efektif dan spesifik.

Kata Kunci : *K-Means Clustering, Pembelajaran Bahasa Indonesia, Nilai*

PENDAHULUAN

Pendidikan merupakan aspek fundamental yang sangat menentukan masa depan suatu masyarakat dan bangsa. Untuk mewujudkan kualitas pendidikan yang optimal, sangat penting bagi kita untuk memahami faktor-faktor yang dapat menyebabkan kegagalan siswa dalam mata pelajaran tertentu. Dengan mengidentifikasi akar penyebab dari permasalahan ini, kita dapat merancang dan menerapkan intervensi yang tepat sasaran untuk membantu siswa mencapai keberhasilan akademik (Alifia, A. N., et al. 2024).

Pendidikan merupakan fondasi utama dalam mengembangkan sumber daya manusia yang unggul. Kualitas pendidikan dasar, khususnya di sekolah dasar, menjadi perhatian utama karena masa inilah yang menjadi dasar bagi pembentukan karakter dan kemampuan akademis siswa. SD Inpres Waingapu 3, sebagai salah satu institusi pendidikan dasar di Kabupaten Sumba Timur, memiliki tantangan dalam memastikan setiap siswa mencapai kompetensi yang diharapkan, khususnya dalam mata pelajaran Bahasa Indonesia. Peningkatan kemampuan literasi Bahasa Indonesia pada siswa SD sangat krusial karena kemampuan ini tidak hanya mempengaruhi prestasi akademik dalam mata pelajaran tersebut, tetapi juga dalam pelajaran lain yang menggunakan Bahasa Indonesia sebagai media pengantar.

Literasi Bahasa Indonesia adalah kemampuan yang sangat berguna untuk siswa sekolah dasar. Kemampuan membaca, menulis, mendengarkan, dan berbicara dalam berkomunikasi yang baik dan benar menjadi kunci sukses di pendidikan lanjutan dan kehidupan sehari-hari. Berdasarkan wawancara dengan guru-guru di SD Inpres Waingapu 3, ditemukan bahwa tingkat literasi Bahasa Indonesia siswa bervariasi secara signifikan. Beberapa siswa menunjukkan kemampuan yang baik, sementara yang lain mengalami kesulitan yang cukup serius. Keragaman ini menunjukkan perlunya pendekatan yang lebih terstruktur dan tepat dalam menangani masalah literasi.

Hasil wawancara dengan para guru mengungkapkan bahwa tingkat literasi Bahasa Indonesia di SD Inpres Waingapu 3 cenderung rendah. Guru-guru mengidentifikasi beberapa faktor penyebab rendahnya literasi ini, termasuk keterbatasan akses terhadap bahan bacaan berkualitas, kurangnya dukungan belajar di rumah, dan metode pembelajaran yang belum sepenuhnya efektif. Data ini menunjukkan perlunya metode evaluasi yang lebih mendalam dan alat bantu yang dapat membantu dalam mengidentifikasi dan mengelompokkan siswa berdasarkan kemampuan literasi mereka.

Algoritma K-Means Clustering merupakan metode yang sangat efektif dalam mengelompokkan data ke dalam kelompok-kelompok dengan karakteristik yang serupa. Dalam penelitian ini, algoritma K-Means Clustering diterapkan pada data kemajuan siswa dalam mata pelajaran Bahasa Indonesia. Data tersebut akan dianalisis untuk mengidentifikasi pola dan ciri-ciri yang mengindikasikan risiko kegagalan siswa. Hasil dari penerapan metode ini akan memberikan dasar bagi pihak sekolah dan pendidik dalam mengambil langkah-langkah intervensi yang sesuai sebagai tindakan peringatan dini. Dengan menggunakan algoritma clustering ini, data siswa dapat dikelompokkan secara efektif berdasarkan kesamaan profil, sehingga memungkinkan identifikasi dini terhadap siswa-siswa yang membutuhkan perhatian khusus. Informasi ini selanjutnya dapat dimanfaatkan untuk merancang program-program pendukung yang tepat guna memastikan keberhasilan siswa (Yannuansa, N., et al. 2021).

Algoritma K-Means Clustering dapat menjadi penyelesaian yang baik untuk mengatasi permasalahan ini. *K-Means* merupakan metode pengelompokan (*clustering*) yang memungkinkan pengelompokan data berdasarkan karakteristik yang serupa. Dalam konteks SD Inpres Waingapu 3, *K-Means* dapat digunakan untuk mengelompokkan siswa berdasarkan

performa mereka dalam literasi Bahasa Indonesia. Penggunaan metode ini dapat membantu guru untuk lebih mudah mengidentifikasi kelompok siswa yang memerlukan perhatian khusus, merancang intervensi yang tepat, dan memantau perkembangan siswa secara lebih terstruktur dan berkelanjutan. Dengan demikian, penerapan *K-Means Clustering* diharapkan dapat meningkatkan efektivitas pembelajaran dan kualitas pendidikan di sekolah tersebut.

MATERI DAN METODE

Nilai

Nilai merupakan realitas yang tersembunyi dibalik realitas lain. Realitas lain merupakan perwujudan dari nilai, seperti halnya suatu benda dapat menjadi wadah atau pembawa warna merah atau warna lain. Nilai adalah kualitas yang keberadaannya tidak tergantung pada wadahnya. Satu objek atau satu tindakan saja sudah cukup untuk menangkap nilai yang terkandung didalamnya. Semua pengalaman yang berhubungan dengan baik dan buruk mengasumsikan adanya pengetahuan dan pemahaman dasar tentang konsep baik dan buruk (Alfariz, F. 2020).

Clustering Data

Menurut Anggraeni dan Yudatama (2023) clustering merupakan metode untuk mengelompokkan data yang digunakan untuk mengidentifikasi kelompok-kelompok (cluster) berdasarkan kesamaan karakteristik diantara unsur-unsur yang lebih kecil. Pengelompokan ini dilakukan berdasarkan adanya kemiripan diantara elemen-elemen data, meski kemiripan tersebut tidak bersifat universal. Oleh karena itu, peneliti atau penganalisis harus mendefinisikan ukuran kesamaan yang akan digunakan terlebih dahulu (Faujia et al., 2022).

Clustering adalah proses membagi satu set objek data kedalam kelompok-kelompok atau cluster yang berbeda. Metode ini sangat berguna untuk menentukan kelompok-kelompok yang tidak diketahui sebelumnya dalam suatu data (Fadilah et al., 2022).

Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) adalah sebuah metrik yang dipakai untuk mengevaluasi hasil dari algoritma clustering, khususnya untuk menentukan kualitas pemisahan antara *cluster-cluster* yang dihasilkan. DBI mengukur rata-rata kesamaan antara setiap cluster dengan *cluster* yang paling mirip, dengan kesamaan didefinisikan sebagai rasio antara jarak dalam-*cluster* (*intra-cluster*) dan jarak antar-*cluster* (*inter-cluster*). Nilai DBI yang lebih rendah membuktikan bahwa *cluster* yang dihasilkan lebih baik dan terpisah satu sama lain, yang artinya hasil *clustering* lebih baik. Secara formal, *DBI* dihitung sebagai rata-rata dari semua rasio kesamaan untuk setiap *cluster*, di mana kesamaan untuk masing-masing *cluster* diambil sebagai nilai maksimum dari rasio tersebut dengan setiap *cluster* lain.

Data Mining

Pola-pola yang ditemukan melalui data mining dikenali oleh perangkat tertentu, yang dapat memberikan analisis data yang bermanfaat. Hasil analisis ini dapat dipelajari lebih lanjut, dan dapat digunakan untuk mendukung pengambilan keputusan lainnya (Sari, 2021).

Data mining juga disebut sebagai knowledge discovery atau pattern recognition. Discovery pengetahuan atau pengenalan pola adalah istilah lain untuk data mining. Istilah-istilah ini memiliki arti yang berbeda. Karena tujuan utama dari penggalian data adalah untuk mendapatkan pengetahuan yang tidak terlihat atau tersembunyi di balik tumpukan data, istilah "penemuan pengetahuan baru" atau penemuan pengetahuan baru lebih tepat digunakan. Karena pengetahuan yang akan diperoleh melalui pengolahan data terdiri dari pola-pola yang mungkin perlu dipelajari lebih lanjut dari kumpulan data yang sedang dianalisis, istilah "pengenalan pola" atau pengenalan pola juga tepat digunakan. Oleh karena itu,

penemuan pengetahuan baru dan identifikasi pola-pola yang tidak terlihat secara langsung dari data yang ada dimungkinkan oleh proses data mining. Istilah-istilah ini menggambarkan tujuan dan prosesnya.(Nainggolan dan Purba, 2019).

Data mining adalah kombinasi dari berbagai disiplin ilmu, seperti basis data, kecerdasan buatan, statistika, dan pembelajaran mesin, yang memungkinkan pengembangan teknik dan metode yang efisien untuk menganalisis jumlah besar data dan mengubahnya menjadi informasi dan pengetahuan yang bermanfaat bagi organisasi. Oleh karena itu, data mining menarik perhatian industri karena ketersediaan data yang melimpah dan kebutuhan untuk mengekstraknya(Yulia, 2021).

Metode K-Means

Metode *K-Means* adalah sebuah *algoritma* yang bersifat *unsupervised learning*. *K-Means* sangat berguna dalam menggabungkan data pada data *cluster*. Algoritma ini dapat menerima data tanpa ada label kategori (Matdoan et al., 2022), menyatakan bahwa *K-Means Clustering Algoritma* juga merupakan metode non-hierarchy. Metode *Clustering Algoritma* merupakan pengelompokan dari beberapa data ke dalam kelompok yang memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda dengan data yang ada di kelompok lain. *Cluster Sampling* adalah teknik pengambilan sampel di mana unit-unit populasi dipilih secara acak dari kelompok yang sudah ada yang disebut *cluster*, jadi *Clustering* atau klasterisasi adalah salah satu masalah yang menggunakan teknik *unsupervised learning* (Yulia, 2021)

RapidMiner

RapidMiner adalah perangkat lunak analisis data yang digunakan untuk mengambil informasi dari data. *RapidMiner* dapat digunakan untuk mengumpulkan data yang diambil dari sumber-sumber tertentu, termasuk database, file Excel, dan data streaming. Setelah data dikumpulkan, *RapidMiner* dapat digunakan untuk menganalisis data dengan menggunakan berbagai teknik analisis, termasuk *clustering*, klasifikasi, *regresi*, dan lain sebagainya (Wibowo et al., 2022). Cara kerja *RapidMiner* adalah dengan mengikuti beberapa langkah berikut:

- a. Pengumpulan Data: Data yang dikumpulkan dapat berupa database, file Excel, dan data *streaming*.
- b. *Preprocessing*: Pada tahap ini akan dilakukan pembersihan data yang tidak relevan, mengisi nilai yang hilang, dan memformat data.
- c. Analisis: *RapidMiner* dapat digunakan untuk menganalisis data dengan berbagai teknik analisis, termasuk *clustering*, klasifikasi, *regresi*, dan lain sebagainya.
- d. Visualisasi: Hasil analisis kemudian dapat divisualisasikan dalam bentuk grafik atau tabel untuk memudahkan pemahaman dan interpretasi.
- e. Prediksi: *RapidMiner* dapat dipakai untuk melacak hasil berdasarkan data yang dikelola dan dianalisis sebelumnya.
- f. Evaluasi: Hasil prediksi kemudian dianalisa untuk memastikan kualitas dan akurasi hasil yang diberikan. Dalam keseluruhan proses kerjanya, *RapidMiner* dapat membantu pengguna dalam proses menentukan hasil berdasarkan data yang dianalisis.



Gambar 1 Alur Penelitian

Pada tahap penelitian ini akan dilakukan dengan tahap pengumpulan data, di mana data nilai siswa yang dipakai yaitu nilai pengetahuan dan keterampilan Bahasa Indonesia dari kelas IV sampai VI dikumpulkan. Selanjutnya, dilakukan tahap *preprocessing* data, di mana data yang telah dikumpulkan dibersihkan, dinormalisasi, dan dikonversi ke format yang sesuai untuk digunakan di RapidMiner.

Pada tahap pengolahan data, algoritma K-Means Clustering diterapkan di RapidMiner. Penentuan jumlah cluster optimal dapat menggunakan *Davies-Bouldin Index (DBI)*, di mana hasil cluster dengan nilai terendah dipilih sebagai yang terbaik. Setelah jumlah cluster ditentukan, RapidMiner menginisialisasi centroid awal dan melakukan proses clustering secara iteratif hingga mencapai konvergensi.

Hasil clustering kemudian divisualisasikan menggunakan alat-alat visualisasi di *RapidMiner*, seperti plot scatter dan heatmap, untuk membantu memahami distribusi dan karakteristik setiap *cluster*. Evaluasi kualitas cluster juga dilakukan menggunakan *DBI* untuk memastikan hasil yang optimal.

Tahap akhir adalah interpretasi hasil clustering, di mana karakteristik masing-masing cluster dianalisis untuk memahami pola atau tren yang muncul. Hasil analisis ini digunakan untuk memberikan rekomendasi strategis dalam meningkatkan literasi bahasa siswa di SD Inpres Waingapu 3, seperti intervensi khusus untuk cluster dengan performa rendah dan tantangan tambahan untuk cluster dengan performa tinggi.

HASIL DAN PEMBAHASAN

Mengumpulkan Data

Pengumpulan data telah dilakukan di SD Inpres Waingapu 3 terkait nilai siswa dalam mata pelajaran Bahasa Indonesia. Data diperoleh dalam format Excel dari SD Inpres Waingapu 3, mencakup nilai pengetahuan dan keterampilan siswa dari tahun ajaran 2020 hingga 2022. Penelitian ini menggunakan data dari tiga semester terakhir, dengan total 25 siswa yang terlibat. Data yang sudah terkumpul ini siap diolah menggunakan sebuah metode *K-Means clustering* dan aplikasi *RapidMiner* untuk mengelompokkan siswa berdasarkan performa akademik mereka.

Pra-Pemrosesan Data

Pada tahap *preprocessing* data, dilakukan pembersihan dan pengorganisasian dataset agar mudah dipahami oleh RapidMiner. Dataset nilai siswa yang awalnya dalam format *Excel* diubah menjadi format *CSV* dengan kolom-kolom yang terstruktur dan jelas. Setiap kolom mewakili nilai performa siswa pada berbagai semester dan jenis keterampilan, seperti "Semester1_P" untuk nilai pengetahuan dan "Semester1_K" untuk nilai keterampilan pada semester pertama, dan seterusnya. Selain itu, semua nilai dan informasi tambahan yang tidak relevan telah dihapus atau diperbaiki untuk memastikan keakuratan dan konsistensi data. Langkah ini penting untuk memastikan bahwa data yang dimasukkan ke dalam *RapidMiner* siap untuk dianalisis lebih lanjut menggunakan metode *K-Means clustering*. Berikut adalah *dataset* yang telah di bersihkan yang akan mudah di pahami *rapidminer* :

Tabel 1 *Dataset*

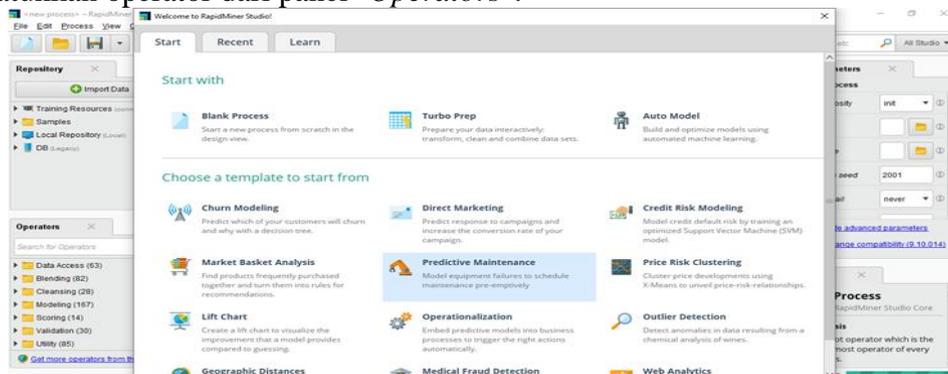
No	Nama	Tahun Ajaran 2020/2021				Tahun Ajaran 2021/2022				Tahun Ajaran 2022/2023			
		Semester 1		Semester 2		Semester 1		Semester 2		Semester 1		Semester 2	
		P	K	P	K	P	K	P	K	P	K	P	K
1.	A**n	85	87	87	91	85	90	87	87	87	87	87	89
2	A***a	85	91	89	93	85	90	85	91	87	90	80	93
3	An***a	83	85	81	85	85	89	87	91	87	93	89	93
4	Re*i	77	79	80	83	75	79	77	79	75	77	80	81
5	Al*****us	80	80	83	85	80	80	85	87	85	89	87	91
N
21	Si***a	80	82	82	83	80	80	83	85	82	83	83	85
22	Li**s	80	80	80	82	77	80	80	85	80	82	82	85
23	P***i	80	80	82	84	80	81	82	84	80	80	80	80
24	U**u	83	85	84	87	85	87	87	89	85	87	86	89
25	A** ***u	80	85	83	85	80	83	80	84	81	83	85	85

Mengolah Data Cluster

Data yang diperoleh digunakan sebagai input untuk membangun model algoritma *K-Means Clustering* menggunakan perangkat lunak *Rapidminer*.

a) Tampilan Utama *Rapid Miner*

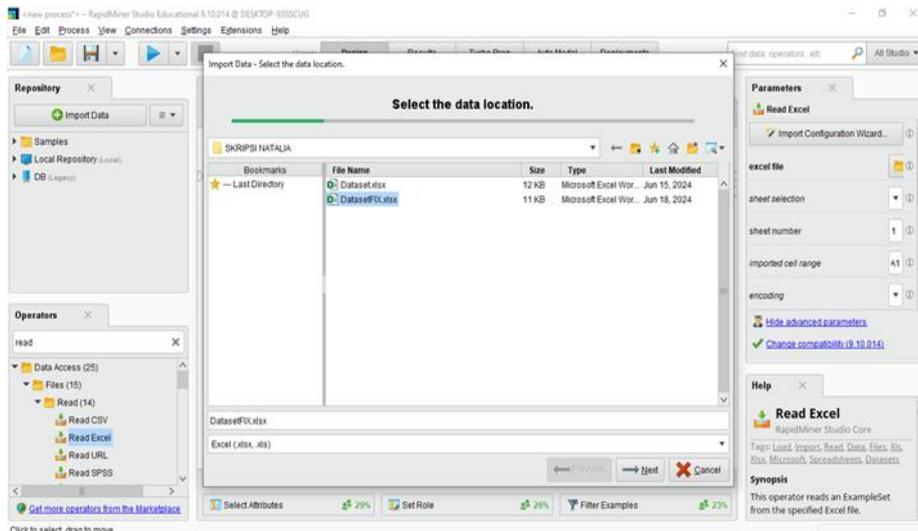
Pada gambar dibawah ini merupakan tampilan utama aplikasi *rapidminer* dalam aplikasi tersebut dapat dilihat beberapa panel utama seperti "*Operators*", "*Repository*", dan "*Design*". Panel "*Operators*" berisi berbagai fungsi yang dapat digunakan dalam analisis data, seperti fungsi untuk pemrosesan data, algoritma *machine learning*, dan alat visualisasi. Panel "*Repository*" memungkinkan Anda untuk mengakses dan mengelola *dataset* dan proyek Anda. Area "*Design*" di tengah adalah tempat Anda membangun proses analisis dengan menyeret dan menjatuhkan operator dari panel "*Operators*".



Gambar 2 Tampilan Utama *Rapidminer*

b) Tampilan Mengimpor Data

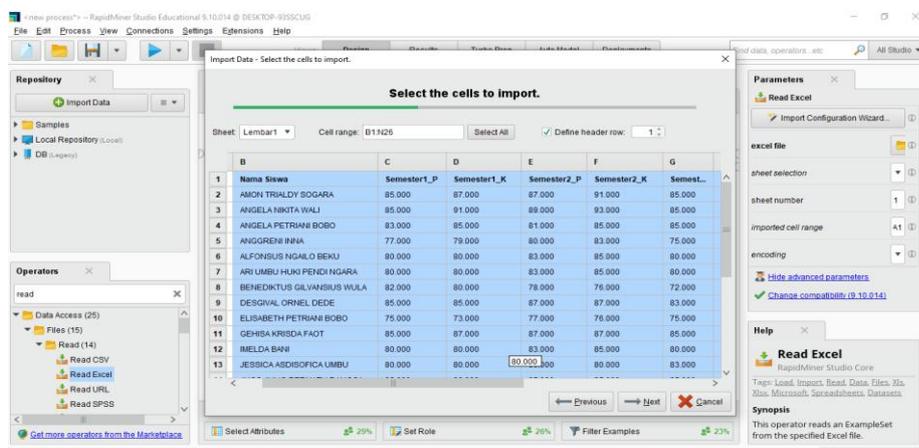
Untuk memulai analisis clustering nilai siswa menggunakan metode K-Means di *RapidMiner*, langkah pertama yang krusial adalah mengimpor dataset ke dalam *RapidMiner*. Dalam kasus ini, *dataset* berisi nilai siswa dari SD Inpres Waingapu 3 pada beberapa semester. *Dataset* ini telah disusun dalam format Excel yang mudah dipahami oleh *RapidMiner*. Berikut adalah gambar tampilan untuk mengimpor file pada *rapidminer*.



Gambar 3 Tampilan Mengimpor Data

c) Tampilan *Select The Cells To Import*

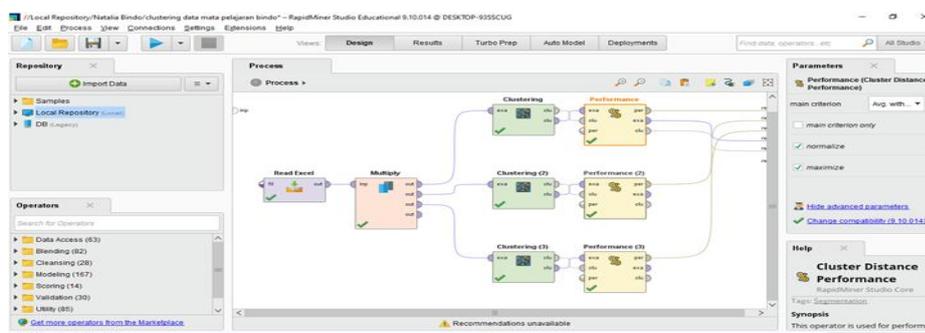
Di *RapidMiner*, tampilan "*Select The Cells To Import*" adalah langkah penting saat mengimpor data dari file *Excel*. Pada tahap ini, Anda akan memilih rentang sel yang berisi data yang ingin digunakan dalam analisis. Saat Anda mengunggah file, *RapidMiner* akan menampilkan pratinjau data dalam bentuk tabel, memungkinkan Anda untuk menentukan kolom dan baris yang relevan untuk diimpor.



Gambar 4 Tampilan *Select The Cells To Import*

d). Tampilan Clustering Rapidminer

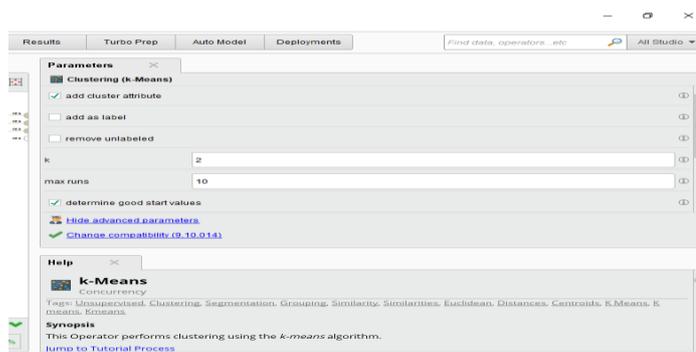
Dari gambar dibawah ini, menunjukkan beberapa operator seperti *Read Excel* dimana *dataset* akan dimasukkan dalam *Read Excel*, *Multiplay* yang berfungsi untuk menjalankan satu atau lebih *operator*; *operator cluster* yaitu *algoritma K-Means*, kemudian *performance* untuk evaluasi kinerja dari *K-Means*.



Gambar 5 Tampilan Clustering Rapidminer

e). Penentuan Jumlah Cluster

Berdasarkan gambar dibawah ini, perlu memasukkan nilai K yang diinginkan. Dalam iterasi yang dilakukan sebanyak 10 kali, nilai K yang dimasukkan adalah $k=2,3,4$. Tujuan iterasi tersebut adalah untuk menemukan kluster dengan nilai DBI yang mendekati 0 atau yang paling kurang, maka dianggap sebagai kluster terbaik. Setelah iterasi selesai, proses klusterisasi dijalankan untuk mendapatkan hasil dari *algoritma K-Means* yang telah diterapkan menggunakan *RapidMiner*. Berdasarkan iterasi yang dilakukan, ditemukan bahwa nilai K terbaik adalah $k=2$, dan hasil klusterisasi diperoleh dengan menggunakan *algoritma K-Means*.



Gambar 6 Penentuan Jumlah Cluster

Interpretasi Hasil

Setelah klusterisasi selesai, hasil kluster pada setiap balita akan ditampilkan dalam bentuk Tabel.

a. Hasil Clustering

	Semester1_P	Semester1_K	Semester2_P	Semester2_K	Semester3_P	Semester3_K	Semester4_P	Semester4_K	Semester5_P	Semester5_K	Semester6_P	Semester6_K	id	cluster
ANGGRINI INNA	77,0	79,0	80,0	83,0	75,0	79,0	77,0	79,0	75,0	77,0	81,0	83,0	4,0	cluster 0
BENEDIKTUS GILVANSIUS WULIA	82,0	80,0	78,0	76,0	72,0	72,0	75,0	75,0	73,0	74,0	77,0	77,0	7,0	cluster 0
ELISABETH PETRIANI BOBO	75,0	73,0	77,0	76,0	75,0	75,0	78,0	75,0	80,0	78,0	76,0	80,0	9,0	cluster 0
NADIRA SHELEN MALO	80,0	81,0	80,0	83,0	80,0	80,0	83,0	83,0	81,0	83,0	83,0	85,0	20,0	cluster 0
SIMPUSIUS REYNARDUS TURU	80,0	80,0	80,0	82,0	77,0	80,0	80,0	85,0	80,0	82,0	82,0	85,0	22,0	cluster 0
MARIA SINTIA JULIA NONO	80,0	80,0	82,0	84,0	80,0	81,0	82,0	84,0	80,0	80,0	80,0	80,0	23,0	cluster 0
AMON TRIADY SOGARJA	85,0	87,0	87,0	91,0	85,0	90,0	87,0	87,0	87,0	87,0	87,0	89,0	1,0	cluster 1
ANGELA NIKITA WAJU	85,0	91,0	89,0	91,0	85,0	90,0	85,0	91,0	87,0	90,0	90,0	99,0	2,0	cluster 1
ANGELA PETRIANI BOBO	83,0	85,0	81,0	85,0	85,0	89,0	87,0	91,0	87,0	93,0	89,0	99,0	3,0	cluster 1
ALFONSUS NGAILO BEKU	80,0	80,0	83,0	85,0	80,0	80,0	85,0	87,0	85,0	89,0	87,0	91,0	4,0	cluster 1
ARI LUMBURUKI PENDI NGARA	80,0	80,0	83,0	85,0	80,0	85,0	87,0	87,0	80,0	80,0	85,0	89,0	6,0	cluster 1
DESIVALORNEL DEDE	85,0	85,0	87,0	87,0	88,0	85,0	85,0	89,0	80,0	81,0	80,0	80,0	8,0	cluster 1
GEHISA KRISDA FACIT	85,0	87,0	87,0	87,0	85,0	85,0	87,0	87,0	85,0	89,0	87,0	89,0	10,0	cluster 1
JEMELDA BANI	80,0	80,0	83,0	85,0	80,0	80,0	85,0	87,0	85,0	85,0	85,0	85,0	11,0	cluster 1
JESSICA ASDISOPICA LUMBURUKI	80,0	80,0	80,0	80,0	80,0	80,0	85,0	85,0	85,0	85,0	87,0	89,0	12,0	cluster 1
JHOS JULIUS BERNATH DANGGIA LUMU	82,0	86,0	85,0	85,0	85,0	85,0	87,0	89,0	87,0	87,0	89,0	89,0	13,0	cluster 1
KRISTIANA DESNAWATI SINTIA NGONGE	83,0	88,0	86,0	88,0	87,0	90,0	89,0	91,0	85,0	87,0	87,0	89,0	14,0	cluster 1
LILIS ASISTA BILU	80,0	85,0	87,0	87,0	80,0	80,0	82,0	80,0	95,0	92,0	85,0	86,0	15,0	cluster 1
MARIA CLARITA RARULAU	80,0	81,0	83,0	85,0	85,0	85,0	85,0	87,0	85,0	85,0	85,0	87,0	16,0	cluster 1
MARIA LANCE SUSANTI DEDE	80,0	80,0	81,0	84,0	80,0	80,0	84,0	87,0	90,0	85,0	80,0	87,0	17,0	cluster 1
MAYBELINDA CLAUDIA NIHA	80,0	82,0	85,0	88,0	82,0	87,0	85,0	89,0	82,0	85,0	85,0	89,0	18,0	cluster 1
MELANIA RIRIN ATE	79,0	81,0	80,0	82,0	80,0	83,0	85,0	85,0	80,0	84,0	85,0	89,0	19,0	cluster 1
SILVIANUS REYNOLDUS WATU	80,0	82,0	82,0	83,0	80,0	80,0	83,0	85,0	82,0	83,0	80,0	85,0	21,0	cluster 1
LUMBU RAUTA LABINI KURI	83,0	85,0	84,0	87,0	85,0	87,0	87,0	89,0	85,0	87,0	86,0	89,0	24,0	cluster 1
VERLY META KONDA MEHA	80,0	85,0	83,0	85,0	80,0	83,0	80,0	84,0	81,0	83,0	85,0	85,0	25,0	cluster 1

Gambar 7 Hasil Clustering

Gambar 7 menunjukkan hasil akhir dari aplikasi *K-Means* yang digunakan untuk melakukan pengelompokan data *cluster*. Berdasarkan hasil pengujian, dapat disimpulkan bahwa terdapat 6 siswa dalam kluster 0, sedangkan cluster 1 terdapat 19 siswa. Pada proses *Clustering*, data dikumpulkan dan digabungkan menjadi beberapa *cluster* menurut kesamaan tertentu. Setelah *cluster-cluster* ini terbentuk, langkah selanjutnya adalah menghitung nilai agregat untuk variabel-variabel penting dalam setiap *cluster*, seperti nilai rata-rata pengetahuan dan keterampilan. Perhitungan ini membantu dalam memahami karakteristik umum dari setiap *cluster*.

Tabel 2 Aggregate Value Cluster 0

	Tahun Ajaran 2020/2021				Tahun Ajaran 2021/2022				Tahun Ajaran 2022/2023			
	Semester 1		Semester 2		Semester 1		Semester 2		Semester 1		Semester 2	
	P	K	P	K	P	K	P	K	P	K	P	K
MIN	75	73	77	76	72	72	75	75	75	73	74	77
MAX	82	81	82	84	80	81	83	85	81	83	83	85
MEAN	79	77	80	80	76	77	79	80	78	78	79	81
DS	3,5	4	2,5	4	4	4,5	4	5	3	5	5	4

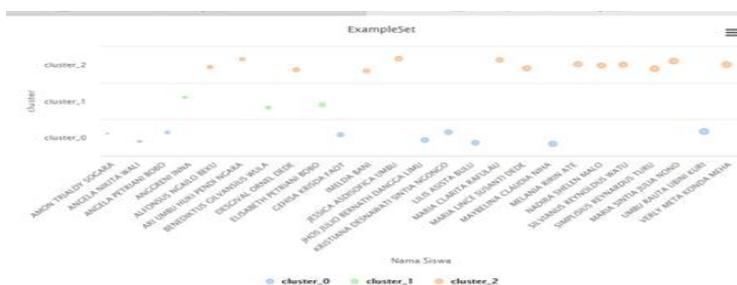
Dari hasil analisis data performa siswa pada setiap semester, terlihat bahwa nilai pengetahuan (P) dan keterampilan (K) mengalami fluktuasi selama enam semester. Nilai terendah (MIN) pada nilai pengetahuan tercatat sebesar 72 pada semester 1 tahun ajaran 2021/2022 dan nilai keterampilan terendah sebesar 72 pada semester 1 tahun ajaran 2021/2022 menunjukkan variasi kinerja yang signifikan di beberapa semester. Nilai tertinggi (MAX) untuk pengetahuan mencapai 83 pada semester 2 tahun ajaran 2022/2023 dan untuk keterampilan mencapai 85 pada semester 2 tahun ajaran 2022/2023, yang menunjukkan adanya peningkatan kinerja maksimal di akhir periode pengamatan. Rata-rata nilai (MEAN) untuk pengetahuan berkisar antara 76 hingga 80, sedangkan untuk keterampilan berkisar antara 77 hingga 81, menunjukkan performa siswa yang cukup konsisten. Standar deviasi yang berkisar antara 2,5 hingga 5 menunjukkan adanya variasi dalam nilai siswa, dengan variasi terbesar terlihat pada keterampilan semester 2 tahun ajaran 2022/2023 (standar deviasi 5). Kesimpulannya, terdapat peningkatan nilai maksimal pada akhir periode dengan variasi nilai yang menunjukkan adanya perbedaan kemampuan antar siswa.

Tabel 3 Aggregate Value Cluster 1

	Tahun Ajaran 2020/2021				Tahun Ajaran 2021/2022				Tahun Ajaran 2022/2023			
	Semester 1		Semester 2		Semester 1		Semester 2		Semester 1		Semester 2	
	P	K	P	K	P	K	P	K	P	K	P	K
MIN	79	80	80	80	80	80	80	80	80	80	80	80
MAX	85	91	89	93	85	90	89	91	95	93	89	93
MEAN	82	86	85	87	83	85	85	86	88	87	85	87
DS	3	5,5	4,5	6,5	2,5	5	4,5	5,5	7,5	6,5	5	6,5

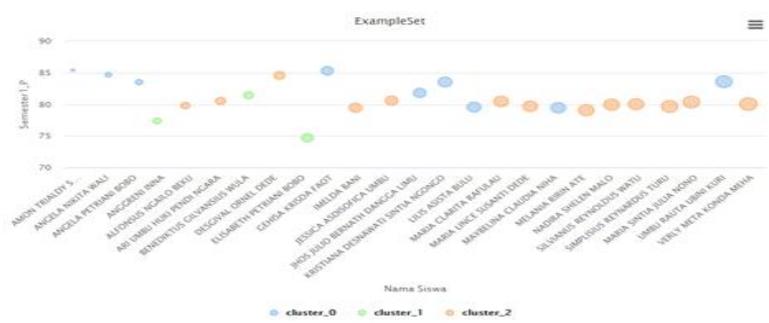
Dari tabel hasil agregat nilai semester siswa, dapat dilihat terdapat variasi yang cukup signifikan dalam performa akademik siswa selama enam semester. Nilai minimum yang diperoleh siswa berkisar antara 79 hingga 80, menunjukkan adanya dasar yang konsisten dalam kemampuan dasar siswa. Nilai maksimum yang diperoleh mencapai hingga 95, mengindikasikan beberapa siswa menunjukkan pencapaian yang sangat tinggi. Rata-rata nilai untuk tiap semester menunjukkan stabilitas, dengan nilai mean berkisar antara 82 hingga 88, yang mencerminkan kinerja keseluruhan siswa yang cukup baik. Namun, standar deviasi yang cukup tinggi, terutama pada Semester2_K (6,5) dan Semester5_K (7,5), menunjukkan adanya perbedaan yang signifikan dalam kemampuan antar siswa dalam kelas yang sama. Hal ini mengindikasikan bahwa meskipun secara umum performa siswa baik, terdapat beberapa siswa yang mungkin memerlukan perhatian tambahan untuk mencapai keseimbangan performa dalam kelas

b). Hasil Visualisasi



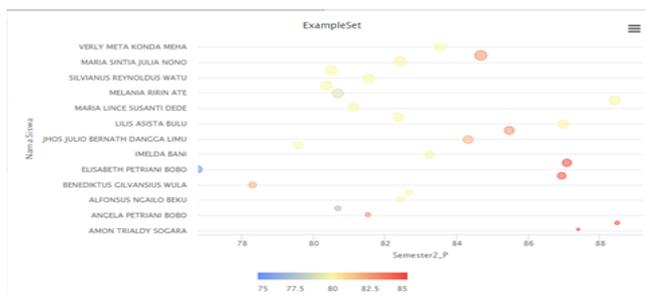
Gambar 4. 1 Hasil Visualisasi Berdasarkan Cluster

Pada gambar diatas merupakan hasil visualiasasi berdasarkan hasil cluster untuk yang berada di sumbu x merupakan nama siswa dan sumbu y merupakan cluster nya. Untuk nilai tertinggi berada di cluster 2 yang berwarna kuning.



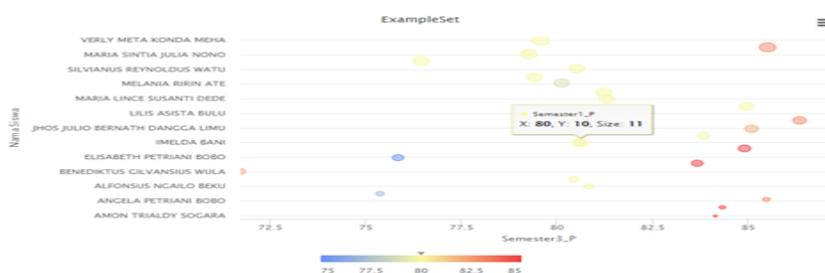
Gambar 8 Hasil Visualisasi Semester 1 tahun ajaran 2020/2021

Pada gambar diatas merupakan hasil visualisasi untuk semester 1, untuk sumbu x berisikan nama siswa dan sumbu y nilai semester serta *cluster 2* menjadi nilai tertinggi.



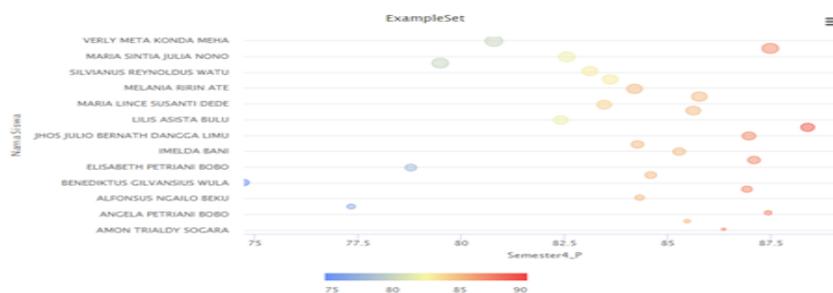
Gambar 9 Hasil Visualisasi Semester 2 tahun ajaran 2020/2021

Pada gambar diatas merupakan hasil visualisasi semester 2 yang pada sumbu x adalah nilai untuk semester 2 dan sumbu y yaitu nama siswa. Nilai tertinggi berada pada garis yang berwarna merah sedangkan nilai terendah berada di garis berwarna biru.



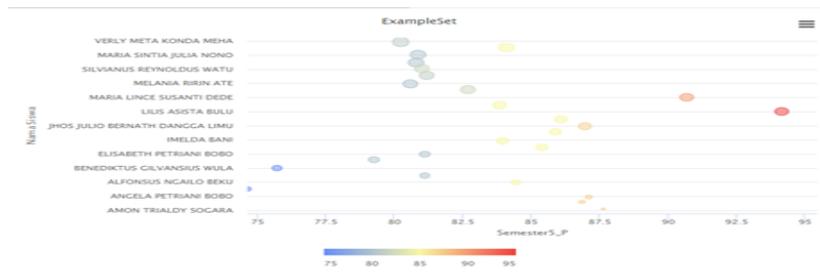
Gambar 10 Hasil Visualisasi Semester 1 tahun ajaran 2021/2022

Pada gambar diatas merupakan hasil visualisasi untuk semester 1 yang nilai nya rendah berwarna biru sedangkan warna merah yang tertinggi dan dapat di lihat pada sumbu x dan y hasil exampleset di atas. Data visualisasi Semester 1 memperlihatkan nilai Bahasa Indonesia siswa di tengah tahun akademik. Grafik ini penting untuk melihat keberlanjutan performa siswa dari semester sebelumnya.



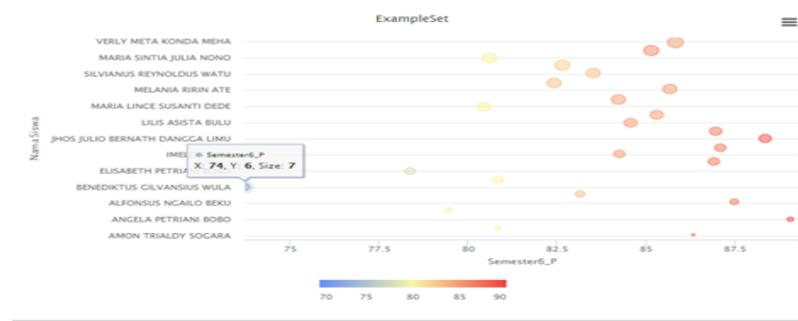
Gambar 11 Hasil Visualisasi Semester 2 tahun ajaran 2021/2022

Pada gambar diatas merupakan hasil visualisasi untuk semester 2 yang nilai nya rendah berwarna biru sedangkan warna merah yang tertinggi dan dapat di lihat pada sumbu x dan y hasil exampleset di atas. Visualisasi Semester 2 menampilkan nilai Bahasa Indonesia siswa setelah melewati separuh tahun akademik.



Gambar 12 Hasil Visualisasi Semester 1 tahun ajaran 2022/2023

Pada gambar diatas merupakan hasil visualisasi untuk semester 1 yang nilai nya rendah berwarna biru sedangkan warna merah yang tertinggi dan dapat di lihat pada sumbu x dan y hasil exampleset di atas. Pada visualisasi Semester 1, kita bisa melihat nilai Bahasa Indonesia siswa mendekati akhir tahun akademik. Grafik ini membantu dalam mengevaluasi konsistensi kinerja siswa selama hampir satu tahun penuh.



Gambar 13 Hasil Visualisasi Semester 2 tahun ajaran 2022/2023

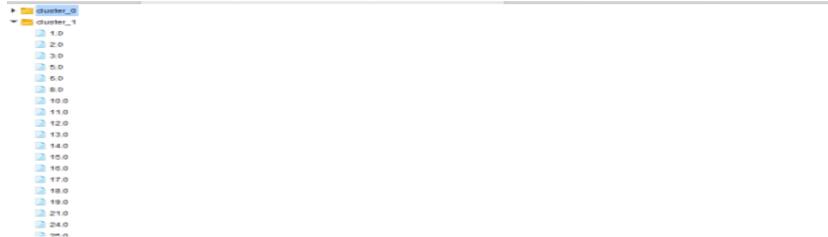
Pada gambar diatas merupakan hasil visualisasi untuk semester 2 yang nilai nya rendah berwarna biru sedangkan warna merah yang tertinggi dan dapat di lihat pada sumbu x dan y hasil exampleset di atas. Visualisasi nilai Bahasa Indonesia pada Semester 2 memberikan gambaran akhir tentang kinerja siswa pada tahun akademik tersebut. Grafik ini memungkinkan kita untuk melihat hasil akhir dari proses belajar mengajar sepanjang tahun. Data ini sangat berguna untuk melakukan evaluasi menyeluruh dan merancang program pembelajaran yang lebih baik di masa depan. Dengan memahami tren dan pola dari semua semester, kita bisa memberikan rekomendasi yang lebih tepat untuk meningkatkan performa siswa di masa mendatang.

c) Cluster Model

Dalam model cluster, terdapat beberapa tampilan yang menampilkan hasil *cluster*, salah satunya adalah tampilan teks yang menunjukkan hasil pengelompokan berdasarkan *cluster* dan jumlah anggota dalam setiap *cluster*. *cluster 0* memiliki 6 items, dan total nya menjadi 25 items.



Gambar 14 Cluster Model



Gambar 15 Folder View

Dari Gambar 15, *Folder view* menunjukkan data secara *komprehensif* dari bagian-bagian *cluster*, dengan setiap anggota dari 2 *cluster* menampilkan id variabel. Setelah menerapkan pengelompokan data nilai bahasa Indonesia menggunakan *RapidMiner*.

Hasil Evaluasi Clustering

Berdasarkan hasil clustering menggunakan metode *K-Means* pada aplikasi *RapidMiner*, performa clustering dievaluasi dengan beberapa metrik. Rata-rata jarak dalam *cluster* terhadap *centroid* adalah 8.370, yang menunjukkan jarak rata-rata setiap titik data ke *centroid* dari *cluster* masing-masing. Secara lebih rinci, rata-rata jarak dalam *cluster* untuk *cluster* 0 adalah 8.650 dan untuk *cluster* 1 adalah 8.281. Evaluasi menggunakan *Davies-Bouldin Index* (DBI) mendapatkan nilai yaitu 0.078 yang berada pada *cluster_3*. Nilai DBI yang rendah ini menunjukkan bahwa clustering yang dilakukan cukup baik, dengan *cluster* yang terbentuk memiliki jarak antar *cluster* yang jelas dan data di dalam *cluster* yang relatif sama. Berdasarkan hasil clustering menggunakan metode *K-Means* pada aplikasi *RapidMiner*, performa clustering dievaluasi dengan beberapa metrik. Rata-rata jarak dalam *cluster* terhadap *centroid* adalah 8.370, yang menunjukkan jarak rata-rata setiap titik data ke *centroid* dari *cluster* masing-masing. Secara lebih rinci, rata-rata jarak dalam *cluster* untuk *cluster* 0 adalah 8.650 dan untuk *cluster* 1 adalah 8.281. kesimpulan akhirnya didapat dari *Davies-Bouldin Index* (DBI) menghasilkan poin paling baik sebesar 0.078 yang berada pada *cluster_2*.

Tabel 4 Perhitungan DBI

<i>K-means</i>			
	<i>Avg. within centroid distance</i>	<i>DBI</i>	<i>Terpilih</i>
	8.370	0.078	<i>K= 2</i>
	5.322	0.081	
	4.477	0.082	

KESIMPULAN

Hasil clustering menggunakan metode *K-Means* pada nilai Bahasa Indonesia siswa SD Inpres Waingapu 3 telah berhasil mengelompokkan siswa ke dalam empat *cluster* berdasarkan performa akademik mereka. Berdasarkan hasil clustering menggunakan metode *K-Means* pada aplikasi *RapidMiner*, performa clustering dievaluasi dengan beberapa metrik penting. Rata-rata jarak dalam *cluster* terhadap *centroid* adalah 8.370, yang menunjukkan jarak rata-

rata setiap titik data ke *centroid* dari *cluster* masing-masing. Secara lebih rinci, rata-rata jarak dalam *cluster* untuk *cluster* 0 adalah 8.650 dan untuk *cluster* 1 adalah 8.281. Evaluasi menggunakan *Davies-Bouldin Index* (DBI) mendapatkan nilai paling baik yaitu 0.078 yang berada pada *cluster_2*. Nilai DBI yang rendah ini menunjukkan bahwa *clustering* yang dilakukan cukup baik, dengan *cluster* yang terbentuk memiliki jarak antar *cluster* yang jelas dan data di dalam *cluster* yang relatif sama. Dari hasil ini, dapat disimpulkan bahwa metode *K-Means* efektif dalam mengelompokkan siswa berdasarkan performa akademik mereka, sehingga dapat membantu dalam memantau kelompok siswa yang memerlukan perhatian khusus atau strategi pembelajaran yang berbeda.

SARAN

Untuk pengembangan lebih lanjut, disarankan untuk membandingkan algoritma *K-Means* dengan algoritma lain. Penambahan jumlah data juga sangat dianjurkan karena metode *K-Means* biasanya bekerja lebih baik dengan dataset yang lebih besar. Selain itu, penelitian mendalam bisa dilakukan dengan melibatkan data dari berbagai sekolah di Sumba Timur agar mendapatkan hasil yang lebih memuaskan.

DAFTAR PUSTAKA

- Alifia, A. N., Setiawan, A. F., & 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.
- Anggraeni, M. R., Yudatama, U., & Maimunah, M. (2023). Clustering Prevalensi Stunting Balita Menggunakan Agglomerative Hierarchical Clustering. *Jurnal Media Informatika Budidarma*, 7(1), 351-359.
- Faujia, R. A., Setianingsih, E. S., & Pratiwi, H. (2022, November). Analisis Klaster K-Means Dan Agglomerative Nesting Pada Indikator Stunting Balita Di Indonesia. In Seminar Nasional Official Statistics (Vol. 2022, No. 1, pp. 1249- 1258).
- Fadilah, A., Pangestu, M. N., Lumbanbatu, S., & Defiyanti, S. (2022). Pengelompokan Kabupaten/Kota Di Indonesia Berdasarkan Faktor Penyebab Stunting Pada Balita Menggunakan Algoritma K-Means. *JIKO (Jurnal Informatika dan Komputer)*, 6(2), 223-230.
- Fatimah, C. (2020). Penggunaan metode praktik dalam meningkatkan keterampilan teknik budi daya tanaman obat. *Al-Azkiya: Jurnal Ilmiah Pendidikan MI/SD*, 5(1), 25-32.
- Matdoan, M. Y., Risdiana, F. Y., & Haumahu, G. (2022). Application of the KMeans Cluster for the Classification of Disadvantaged Districts/Cities in Maluku Province. *JRST (Jurnal Riset Sains dan Teknologi)*, 6(1), 61-64.
- Nainggolan, R., & Purba, E. N. (2019). Perbaikan Performa Cluster K-Means Menggunakan Sum Squared Error (SSE) Pada Analisis Online Customer Review Terhadap Produk Toko Online. *Jurnal Times*, 8(2), 1-8.
- Ningrum, K. K., Maulindar, J., & Farida, A. (2023). Penerapan Algoritma K-Means untuk Pengelompokan Penilaian Akhir Semester di SDN Kadokan 01 Sukoharjo. *INFOTECH journal*, 9(1), 190-197.
- N. Yannuansa, M. Safa'udin, and M. I. Aziz, "Pemanfaatan Algoritma K-Means Clustering dalam Mengolah Pengaruh Hasil Belajar Terhadap Pendapatan Orang Tua Pada Mata Pelajaran Produktif," *J. Tecnoscienza*, vol. 6, no. 1, pp. 43–55, 2021, doi: 10.51158/tecnoscienza.v6i1.530.
- N. Yannuansa, dkk. (2021). Penggunaan Davies-Bouldin Index dalam Evaluasi Hasil Clustering. *Jurnal Teknik Informatika*, 10(2), 100-110.

- Ranjawali, R., Talakua, A. C., & Abineno, R. T. (2023). CLUSTERING STUNTING PADA BALITA DENGAN METODE KMEANS DI PUSKESMAS KANATANG. *Proceeding Sustainable Agricultural Technology Innovation (SATI)*, 2(1), 80-92.
- Sari, P. F., Pardede, A. M., & Maulita, Y. (2021, June). Pengelompokan Populasi Hewan Ternak Menggunakan Metode Clustering (Studi Kasus: Dinas Pertanian dan Ketahanan Pangan Kabupaten Langkat). *In Seminar Nasional Informatika (SENATIKA)* (pp. 37-46).
- Sembiring, S. N. B., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 1(1), 31-40.
- Wibowo, A. P., Darmawan, W., & Amalia, N. (2022). Komparasi Metode Naïve Bayes dan K-Nearest Neighbor terhadap Analisis Sentimen Pengguna Aplikasi Pedulilindungi. *IC-Tech*, 17(1), 18-23.
- Yulia, N., Saragih, R., & Ambarita, I. (2021). Data Mining Pengelompokan Anak Stunting Berdasarkan Usia, Penyebab dan Pekerjaan Orang Tua Dengan Menggunakan Metode Clustering (Studi Kasus: Dinas Kesehatan Kabupaten Langkat). *In Seminar Nasional Informatika (SENATIKA)* (pp. 295-306).