



Pengelompokan Performa Siswa Dalam Pelajaran Matematika Dengan Algoritma *K-means* Di SMP Negeri 4 Mauluru

Grouping Student Performance in Mathematics Lessons Using the K-means Algorithm at SMP Negeri 4 Mauluru

Suranto Tamu Boku¹, Reynaldi Thimotius Abineno², 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: surantotamuboku@gmail.com

ABSTRACT

Numeracy literacy is a crucial aspect of mathematics education that has a major impact on students' ability to understand and apply mathematical concepts in everyday life. This research uses the K-Means algorithm, an efficient clustering method for grouping data based on similar attributes. The data used in this research includes students' knowledge and skills scores in mathematics subjects. The aim of this research is to assess the numeracy literacy level of students at SMP Negeri 4 Mauluru using the K-Means algorithm. Analysis and evaluation of the clustering results was carried out to gain a better understanding of the student's numeracy literacy level. Based on the evaluation results of the Davies Bouldin Index calculated using the RapidMiner application, it produces optimal performance, namely $K=2$ with a value of 0.862, where this value is high and not close to 0, this shows that student value data does not vary significantly. Even though the DBI is high, the clustering results can still be considered good because the low standard deviation value shows consistency in student performance. Cluster 0 is a cluster that has a high score, namely 14 students, and Cluster 1 is a group consisting of 46 students who have the lowest score.

Keywords: *K-means algorithm, Data mining, Clustering, literacy, Education*

ABSTRAK

Literasi numerasi adalah aspek krusial dalam pendidikan matematika yang memiliki dampak besar terhadap kemampuan siswa dalam memahami dan mengaplikasikan konsep matematika dalam kehidupan sehari-hari. Penelitian ini menggunakan algoritma *K-Means*, metode *clustering* yang efisien untuk mengelompokkan data berdasarkan kesamaan atribut. Data yang digunakan dalam penelitian ini mencakup nilai pengetahuan dan keterampilan siswa dalam mata pelajaran matematika. Tujuan penelitian ini adalah untuk menilai tingkat literasi numerasi siswa di SMP Negeri 4 Mauluru menggunakan algoritma *K-Means*. Analisis dan evaluasi hasil *clustering* dilakukan untuk memperoleh pemahaman yang lebih baik tentang tingkat literasi numerasi siswa tersebut. Berdasarkan hasil evaluasi *Davies Bouldin Index* yang dihitung menggunakan aplikasi *RapidMiner*, menghasilkan performa yang optimal yaitu $K=2$ dengan nilai 0.862 dimana nilai tersebut tinggi dan tidak mendekati 0, hal ini menunjukkan bahwa data nilai siswa tidak bervariasi secara signifikan. Meskipun DBI tinggi, hasil klusterisasi masih dapat dianggap baik karena nilai standar deviasi yang rendah menunjukkan konsistensi dalam performa siswa. *Cluster 0* merupakan *cluster* yang memiliki nilai tinggi yaitu 14 siswa, dan *Cluster 1* adalah kelompok yang terdiri dari 46 siswa yang memiliki nilai terendah.

Kata kunci: *Algoritma K-means, Data mining, Clustering, Literasi, Pendidikan.*

PENDAHULUAN

Pendidikan merupakan faktor krusial yang menentukan performa siswa dalam pembelajaran matematika. Pendidikan yang berkualitas memberikan dasar yang kuat bagi siswa untuk mengembangkan keterampilan matematika. Selain itu, literasi numerasi juga berperan penting dalam pembelajaran matematika. Literasi numerasi memungkinkan siswa memahami dan menggunakan materi matematika dengan baik, sehingga dapat meningkatkan kemampuan mereka dalam menyelesaikan masalah matematika dalam kehidupan sehari-hari.

Setiap anak harus memiliki keterampilan dasar literasi numerasi. Pemahaman matematika memberi mereka kemampuan untuk menggunakan simbol dan persamaan matematika dalam mengungkapkan masalah dalam kehidupan sehari-hari. Selain itu, kinerja numerik yang tidak memadai dapat dikaitkan langsung dengan risiko ekonomi yang lebih tinggi. Studi menunjukkan bahwa bakat matematika pada usia 42 tahun memiliki korelasi yang signifikan dengan status sosial dan ekonomi individu mereka. Orang dewasa dengan tingkat literasi numerasi rendah cenderung memiliki peluang kerja yang terbatas, gaji yang rendah, dan berisiko mengalami kemiskinan. Oleh karena itu, literasi numerasi memiliki peran penting dalam menentukan tingkat pendapatan seseorang.

Tantangan terkait kemampuan matematika tetap menjadi masalah yang menghambat negara-negara berpendapatan rendah dan menengah dalam mencapai tujuan pembangunan mereka. Data kemampuan matematika siswa Indonesia pada ujian Program *for International Student Assessment* (PISA) menunjukkan tren yang tidak stabil. Meskipun nilai rata-rata kemampuan matematika pada PISA adalah 500, Indonesia memperoleh skor yang rendah. Hal ini menunjukkan pentingnya untuk melakukan upaya dalam meningkatkan kemampuan matematika siswa.

SMP Negeri 4 Mauliru menjadi fokus penelitian ini. Pada studi kasus ini, analisis akan dilakukan terhadap performa siswa dalam pembelajaran matematika tersebut. Salah satu permasalahan di SMP Negeri 4 Mauliru adalah adanya perbedaan yang cukup signifikan dalam literasi numerasi siswa dalam pelajaran matematika. Beberapa siswa menunjukkan kinerja yang sangat baik, sementara yang lain menghadapi kesulitan dalam memahami dan menguasai materi pembelajaran matematika tersebut. Untuk meningkatkan literasi numerasi secara keseluruhan, diperlukan analisis yang mendalam untuk mengidentifikasi kelompok siswa berdasarkan kinerja mereka.

Analisis *data mining* adalah metode yang berguna untuk mengungkap informasi tersembunyi dalam data. Dalam konteks ini, *data mining* digunakan untuk mengidentifikasi pola dan tren dalam performa siswa dalam pembelajaran matematika di SMP Negeri 4 Mauliru. Dengan menggunakan teknik *data mining*, siswa dapat dikelompokkan berdasarkan performa mereka, dengan demikian memberikan wawasan yang lebih dalam tentang faktor-faktor yang mempengaruhi kemampuan matematika mereka.

Metode *K-Means* adalah salah satu teknik *clustering* yang digunakan untuk mengelompokkan data menjadi beberapa kelompok berdasarkan kesamaan atribut. Dalam konteks ini, *K-Means* bisa diterapkan untuk mengelompokkan siswa berdasarkan performa mereka dalam matematika. Metode ini dipilih karena keunggulannya dalam mengelompokkan data secara otomatis dan efisien.

Dengan adanya penelitian tentang pengelompokan performa siswa dalam pembelajaran matematika menggunakan algoritma K-Means di SMP Negeri 4 Mauliru bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi performa siswa dalam

matematika serta memberikan pemahaman yang lebih mendalam tentang kinerja siswa di sekolah tersebut.

MATERI DAN METODE

Data Mining

Data mining adalah proses mengidentifikasi pola atau informasi menarik dalam data melalui berbagai teknik atau metode tertentu. Terdapat berbagai jenis teknik, metode, atau algoritma yang digunakan dalam *data mining*. Salah satu metode klasifikasi dalam *data mining* adalah *K-Means*. *K-Means* merupakan teknik *clustering* yang bersifat partisional, bukan hierarkis. *Clustering partisional* membagi data menjadi beberapa kelompok secara bersamaan, sedangkan *clustering hirarkis* membangun *cluster* secara bertahap. *K-Means* bukanlah metode hirarkis langkah demi langkah, karena tidak ada penambahan elemen secara bertahap. *K-Means* beroperasi dengan cara mengelompokkan data berdasarkan jarak ke *centroid* (titik pusat) *cluster*. (Sopyan, Lesmana, en Juliane 2022)

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan (*artificial intelligence*), dan *machine learning* untuk mengekstrak serta mengidentifikasi informasi yang berguna dan pengetahuan terkait dari berbagai basis data besar. (Br Sembiring, Winata, en Kusnasari 2022)

Clustering

Clustering adalah proses mengelompokkan objek-objek yang mirip ke dalam kelompok-kelompok yang berbeda. Secara ringkas, *clustering* adalah proses membagi data menjadi *subset* yang lebih kecil, di mana setiap subset memiliki kesamaan dan makna yang penting. Dalam satu *cluster*, objek-objek dalam satu *cluster* memiliki kesamaan satu sama lain dan berbeda dari objek-objek di *cluster* lainnya. Algoritma *clustering* dapat diklasifikasikan menjadi dua jenis: hirarkis dan partisi. Algoritma hirarkis menemukan *cluster* secara bertahap menentukan *cluster* satu per satu. Sementara itu, algoritma partisional menentukan semua kelompok sekaligus. *Clustering* juga dapat diartikan sebagai proses mengelompokkan dan memisahkan pola data menjadi beberapa set data, sehingga pola-pola yang serupa dikelompokkan bersama dalam satu *cluster*, sementara pola-pola yang berbeda dikelompokkan dalam *cluster* yang terpisah. *Clustering* memiliki peran penting dalam kehidupan sehari-hari, mengingat kita dikelilingi oleh banyak data yang perlu dianalisis untuk memenuhi kebutuhan kita. Salah satu cara penting untuk mengolah data adalah dengan mengelompokkannya ke dalam kategori atau *cluster*. *Clustering* banyak digunakan dalam berbagai bidang, seperti analisis data statistik, pembelajaran mesin, *data mining*, pengenalan pola, analisis citra, dan bioinformatika. Secara umum, teknik *clustering* dapat dibagi menjadi tiga kategori: partisi, hierarkis, dan berbasis lokalitas.

Algoritma K-means

Algoritma *K-Means* termasuk dalam kategori algoritma partisional karena bergantung pada jumlah *cluster* awal yang telah ditentukan dan nilai awal *centroid* yang telah ditentukan. *K-Means* menggunakan iterasi untuk menemukan *cluster data*. Input yang diperlukan adalah jumlah *cluster* yang diinginkan pada awalnya, dan *outputnya* adalah lokasi akhir *centroid*. Algoritma *K-Means* secara acak memilih titik awal untuk *centroid*, dan proses iteratifnya dipengaruhi oleh titik awal tersebut untuk mencapai *centroid cluster*. Dalam pengembangan algoritma, penting untuk menentukan *centroid cluster* dari data awal yang padat untuk mendapatkan hasil yang lebih optimal. (Ningrum, Maulindar, en Farida 2023).

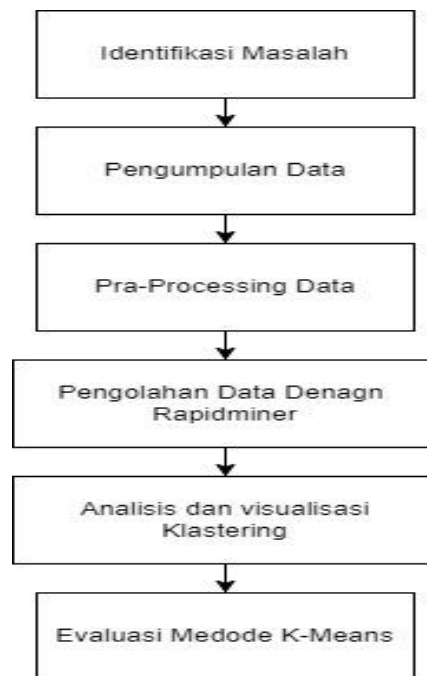
RapidMiner

RapidMiner menyediakan berbagai prosedur untuk *data mining* dan *machine learning*, termasuk ETL (ekstraksi, transformasi, *loading*), *pra-pemrosesan data*, visualisasi, pemodelan, dan evaluasi. Proses *data mining* di *RapidMiner* terdiri dari operator-operator yang bisa disusun secara bertingkat, dijelaskan menggunakan XML, dan dibuat melalui antarmuka grafis pengguna (GUI). Hasil akhirnya ditampilkan menggunakan bahasa pemrograman Java.(Anwar et al. 2022)

Davies Bouldin Indeks

Indeks Davies Bouldin (DBI) adalah metrik yang digunakan untuk menilai hasil dari algoritma *clustering*. Metrik ini pertama kali diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Dalam evaluasi *clustering*, *cluster* dianggap optimal jika memiliki nilai DBI yang minimal.(Sopyan et al. 2022)

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



Gambar 1. Alur Penelitian

Penelitian ini dimulai dengan mengidentifikasi masalah yang muncul di SMP Negeri 4 Mauluru. Tujuan dari penelitian ini adalah untuk mengidentifikasi berbagai masalah terkait data nilai pembelajaran matematika siswa di SMP Negeri 4 Mauluru. Studi Literatur: Tahap studi literatur dilakukan untuk mencari landasan teori dari berbagai sumber seperti buku, internet, *e-book*, dan jurnal penelitian sejenis, guna memperkuat argumentasi dan mencari informasi yang relevan.

Tahap ini dilakukan melalui pengamatan langsung terhadap objek yang diteliti. Data yang terkumpul mencakup data nilai pengetahuan dan keterampilan matematika siswa di SMP Negeri 4 Mauluru. Tahap awal dalam penelitian ini melibatkan pengumpulan data menggunakan metode penelitian kualitatif. Penelitian akan berfokus pada hasil-hasil wawancara dan observasi, laporan, jurnal, berita maupun hasil penelitian serupa yang telah dihasilkan pada sebelumnya. Pengumpulan data dilakukan dengan cara melakukan survei atau

mengunjungi secara langsung di lokasi objek penelitian yaitu di SMP Negeri 4 Mauiru. Terdapat jenis data yang diperoleh yaitu berupa format data excel. Penelitian ini menggunakan data kelas IX A dan IX B tahun 2021/2022 yang diperoleh dari SMP Negeri 4 Waingapu.

Persiapan dalam *data mining* melibatkan serangkaian kegiatan yang dilakukan sebelum memproses data. Tujuannya adalah untuk meningkatkan efisiensi waktu dan pekerjaan. Tahapan ini mencakup beberapa proses penting untuk mempersiapkan data dari bentuk mentah menjadi bentuk yang siap digunakan. Proses ini termasuk: Pembersihan Data: Menghapus data yang tidak konsisten dan *noise*, Seleksi Data: Memilih data yang relevan sesuai dengan tujuan analisis, dan Transformasi Data: Mengubah data ke bentuk yang cocok untuk analisis *data mining*.

Pada tahap ini, klasifikasi menggunakan *algoritma K-Means* dilakukan untuk menganalisis performa siswa dalam pembelajaran matematika. Pengolahan data dilakukan menggunakan aplikasi *RapidMiner*. Data yang telah diproses akan diolah menggunakan *RapidMiner*, sebuah perangkat lunak yang digunakan untuk analisis data. Dalam tahap ini, data dimasukkan ke *RapidMiner* dan algoritma yang tepat diterapkan untuk mengelompokkan siswa berdasarkan performa mereka.

Setelah pengolahan data, hasil klustering dianalisis dan divisualisasikan. Analisis ini bertujuan untuk memahami karakteristik dari setiap kelompok siswa yang terbentuk. Visualisasi klustering dapat berupa grafik atau diagram yang menunjukkan bagaimana siswa dikelompokkan berdasarkan performa mereka.

Tahap terakhir adalah evaluasi algoritma *K-Means* yang digunakan untuk klustering. Evaluasi ini dilakukan untuk menilai seberapa baik algoritma *K-Means* dalam mengelompokkan siswa. Kriteria evaluasi bisa meliputi seberapa baik klaster yang terbentuk memisahkan siswa berdasarkan performa mereka dan seberapa konsisten hasil klustering jika diterapkan pada data yang berbeda. Setelah proses analisis *clustering* menggunakan *K-Means* selesai, evaluasi *cluster* dilakukan menggunakan *Davies-Bouldin Index (DBI)*. Tahapan evaluasi DBI dimulai dengan menghitung nilai *Sum of Square Within-cluster (SSW)*, *Sum of Square Between Cluster (SSB)*, dan menghitung rasio antara keduanya untuk mendapatkan nilai DBI.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Tahap awal melibatkan pengumpulan informasi melalui penelitian kualitatif. Penelitian ini terutama difokuskan pada hasil wawancara dan temuan penting dari penelitian sebelumnya yang relevan. Data dikumpulkan melalui wawancara di SMP Negeri 4 Mauluru, dengan informasi yang terkumpul mencakup nilai pengetahuan dan keterampilan dalam pelajaran matematika. Penelitian ini didasarkan pada data tahun 2021/2022 yang terdiri dari 60 data siswa, dengan fokus pada data nilai kelas IX A dan IX B yang dikumpulkan dari SMP Negeri 4 Mauluru.

Persiapan Dataset

Setelah pengumpulan data, dataset disiapkan dengan baik dan diperiksa ulang untuk memastikan keakuratannya. Langkah berikutnya adalah melakukan pengolahan data atau analisis. klasifikasi dengan lebih efisien karena kesalahan data telah dikurangi.

Tabel 1. Dataset

No	Nama Siswa	JK	Nilai Matematika Tahun 2021/2022			
			Semester 1		Semester 2	
			P	K	P	K
1.	A****	0	84	85	83	85
2.	Y****	0	82	82	74	83
3.	K****	0	74	75	81	80
4.	R****	1	78	79	73	80
5.	A****	0	72	75	76	83
....
59.	Y****	1	70	74	75	78
60.	L****	1	78	81	80	85

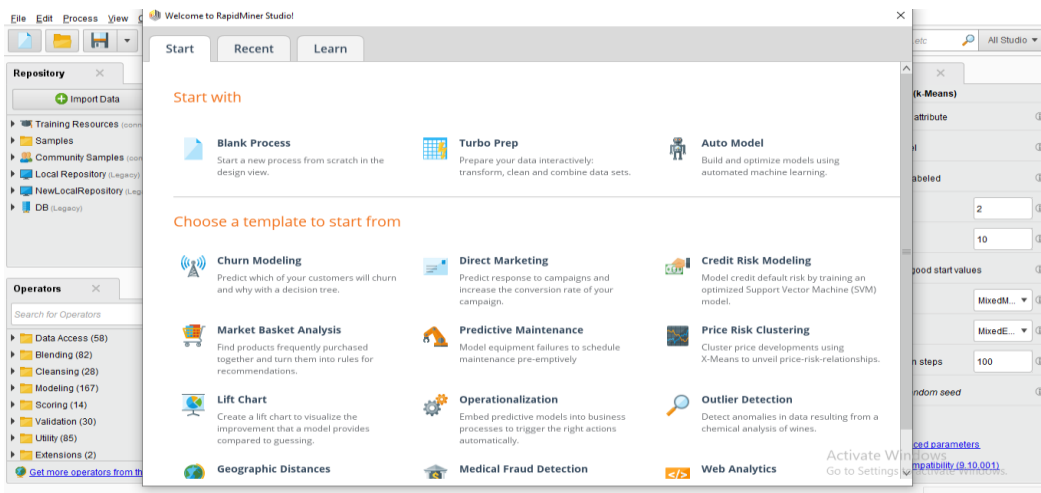
Dari dataset yang diperoleh dari pengumpulan data, data yang digunakan mencakup atribut nama siswa, jenis kelamin, serta nilai pengetahuan dan keterampilan

Clustering Menggunakan Perangkat Lunak RapidMiner

Data yang diperoleh digunakan sebagai input untuk mengembangkan model *algorithm* K-Means Clustering menggunakan perangkat lunak RapidMiner.

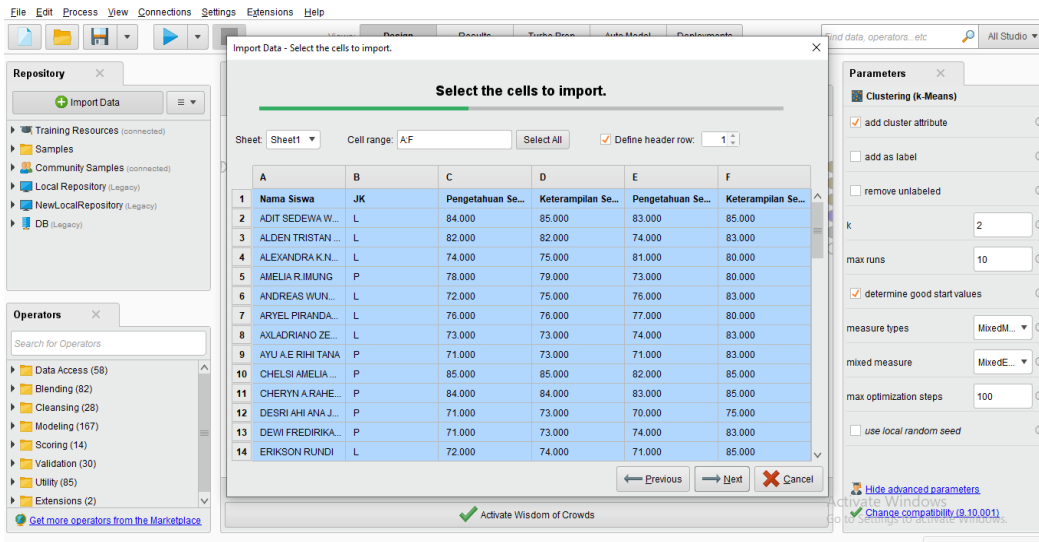
Tampilan Menu Utama RapidMiner

Tampilan awal dari menu utama RapidMiner mencakup berbagai komponen dan fitur, termasuk fitur "New Process" yang merupakan langkah pertama dalam menggunakan RapidMiner.



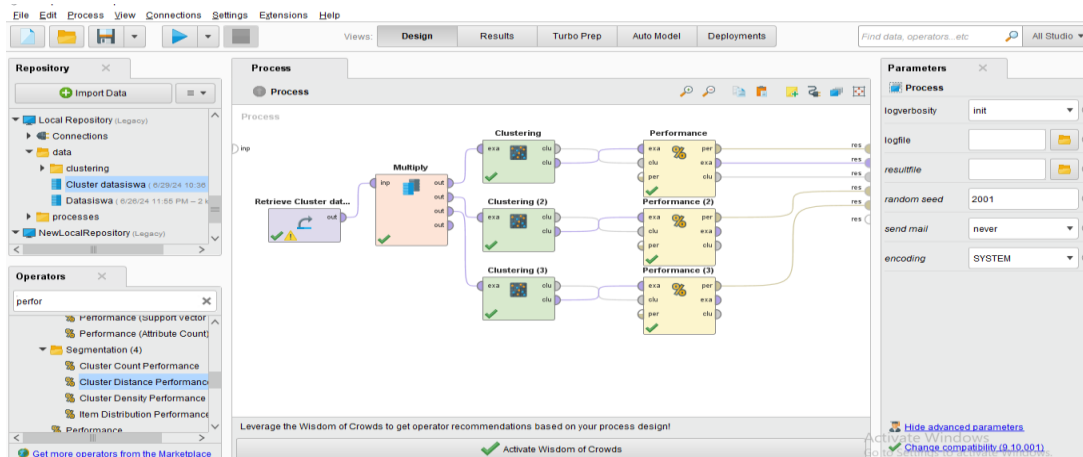
Gambar 1. Tampilan Menuawal RapidMiner

Gambar 1 tampilan awal dari menu utama *RapidMiner* mencakup berbagai komponen dan fitur, termasuk fitur "New Process" yang merupakan langkah pertama dalam menggunakan *RapidMiner*.



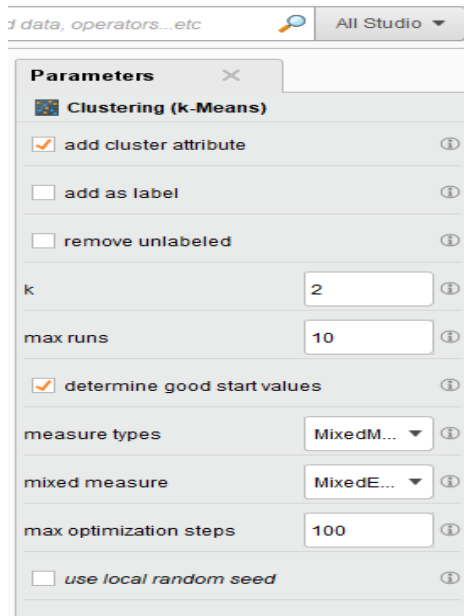
Gambar 2. Tampilan *Select The Cells To Import*

Gambar 2 adalah sistem memberikan petunjuk tentang cara mengimpor data baru yang akan diolah selanjutnya dalam format *Excel*.



Gambar 3. Tampilan *Clustering Rapidminer*

Gambar 3 terlihat beberapa operator seperti "retrieve" yang digunakan untuk memuat dataset yang telah disimpan sebelumnya di repositori *RapidMiner*, "set role" operator yang digunakan untuk mengubah peran atribut dalam *dataset*. Peran atribut menentukan penggunaan atribut tersebut dalam proses analisis, seperti fungsi atribut sebagai pengenal unik untuk setiap baris data. Operator "Nominal to Numerical" digunakan untuk mengonversi atribut dengan tipe data nominal menjadi tipe data numerik, sementara operator "Multiply" berfungsi untuk menjalankan satu atau lebih operator. Operator "Cluster" menggunakan algoritma *k-means*, dan terdapat operator "Performance" untuk mengevaluasi kinerja dari *k-means*.



Gambar 4. Penentuan Jumlah Cluster

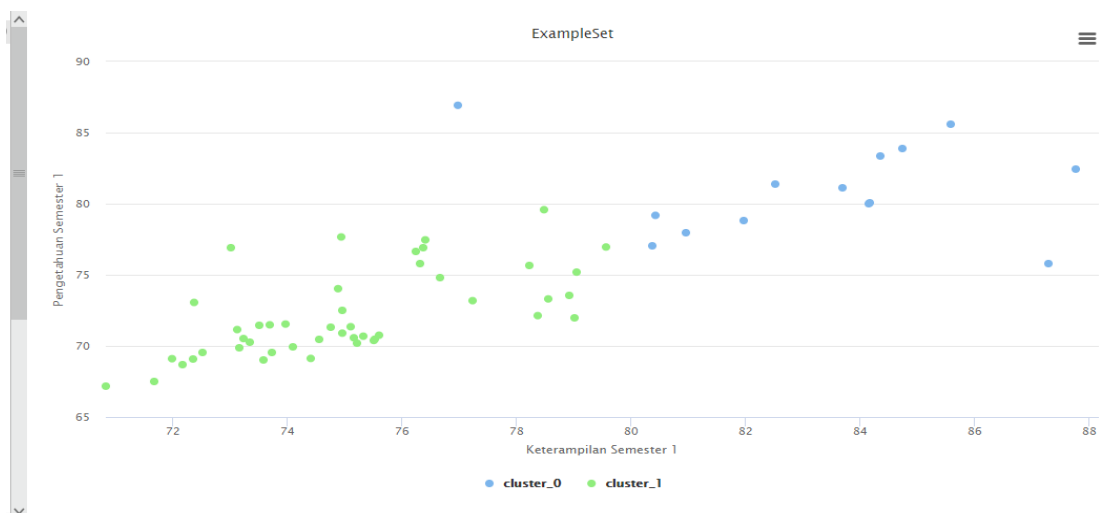
Gambar 4 merupakan langkah pertama adalah memasukkan nilai K yang diinginkan. Iterasi dilakukan sebanyak 10 kali dengan nilai K=2. Tujuan dari iterasi ini adalah untuk menemukan kluster dengan nilai DBI yang mendekati 0 atau paling rendah, yang dianggap sebagai kluster optimal. Setelah iterasi selesai, proses *clustering* dilakukan untuk mendapatkan hasil dari algoritma *K-Means* yang diterapkan menggunakan *RapidMiner*. Dari hasil iterasi, nilai k terbaik adalah k=2, dan hasil klusterisasi diperoleh menggunakan algoritma *K-Means*.

Tabel 2. Aggregate Value Clustering

MIN				
Cluster	P1	K1	P2	K2
0	71	73	71	79
1	68	71	70	73
MAX				
0	77	80	81	85
1	71	76	76	80
RATA-RATA				
0	74.233	76.933	74.883	81.083
1	69,773	73.636	73.182	77.409
STANDAR DEVIASI				

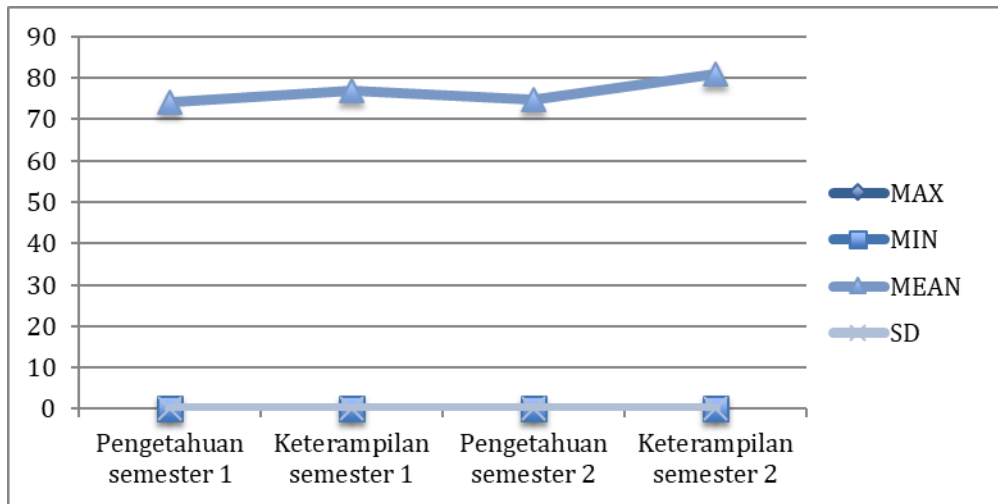
0	3,162	2,983	5,820	2,303
1	3,631	2,299	2,336	3,227

Setiap *cluster* memiliki karakteristik yang berbeda berdasarkan variabel pengetahuan semester 1, keterampilan semester 1, pengetahuan semester 2 dan keterampilan semester 2. Berdasarkan hasil Agregat Value diatas, nilai terendah (Min) yang ditemukan pada Cluster 0, terdapat pada variabel pengetahuan semester 1 dengan nilai minimum 71, untuk variabel keterampilan semester 1 adalah 73 dan 79 untuk nilai minimum pada variabel pengetahuan semester 2. Nilai tertinggi (Max) yang terdapat pada Cluster 1, yaitu dalam variabel pengetahuan semester 1 adalah 71, untuk variabel keterampilan semester 1 dengan nilai 76, sedangkan variabel pengetahuan semester 2 adalah 76 dan 80 untuk nilai max pada variabel pengetahuan semester 2. Nilai rata-rata (Mean) dari Cluster 0, pada variabel pengetahuan semester 1 adalah 74.233, sedangkan nilai mean dari variabel pengetahuan semester 1 dengan nilai 73,671, sedangkan untuk nilai rata-rata pada variabel pengetahuan semester 2 adalah 5,820 dan untuk nilai rata-rata dari variabel keterampilan semester 2 adalah 2,303. Nilai standar deviasi dari cluster 1 pada variabel pengetahuan semester 1 adalah 3,631, sedangkan nilai standar deviasi yang paling rendah dari variabel keterampilan semester 1 dengan nilai 2,299, untuk nilai standar deviasi pada variabel pengetahuan semester 2 adalah 2,336 dan untuk nilai rata-rata dari variabel keterampilan semester 2 adalah 3,227.



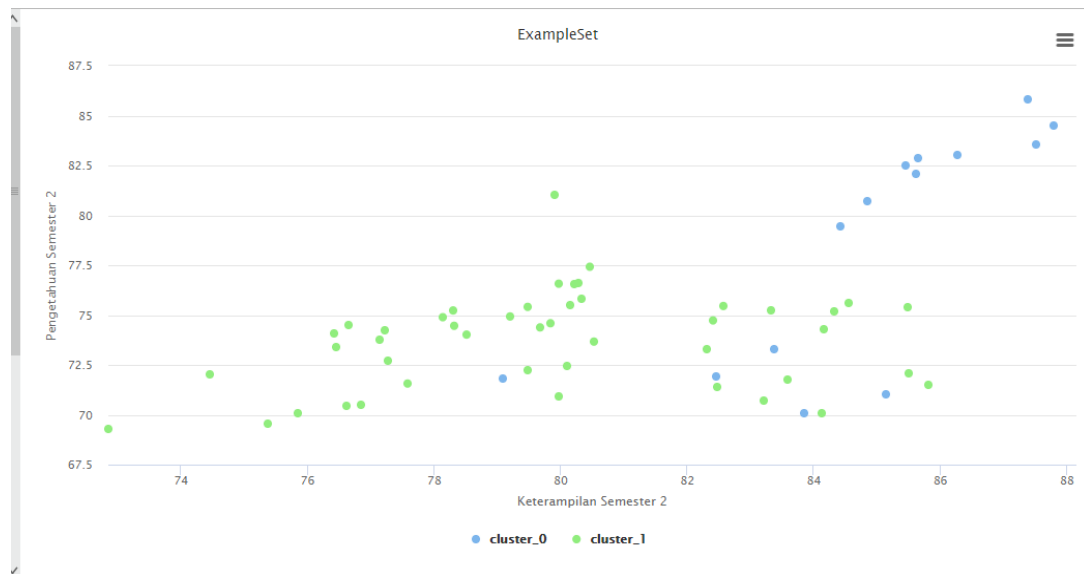
Gambar 5 Scatter Plot Nilai Pengetahuan semester 1 dan Nilai Keterampilan semester 1

Gambar 5 Menunjukkan tampilan dari aplikasi *rapidminer* yang digunakan untuk melakukan pengelompokan data *cluster* nilai pengetahuan semester 1 dan nilai keterampilan semester 1. Berdasarkan hasil pengujian, dapat ditarik kesimpulan bahwa terdapat 14 siswa dalam *cluster* 0 yang ditandai oleh warna biru, *cluster* 0 memiliki nilai di sumbu x dari rentang nilai 65-88 dan sumbu y dari 65-88 sedangkan *cluster* 1 ditandai dengan warna hijau, memiliki nilai berdasarkan sumbu x dari 65-80 dan sumbu y dari 65-80.



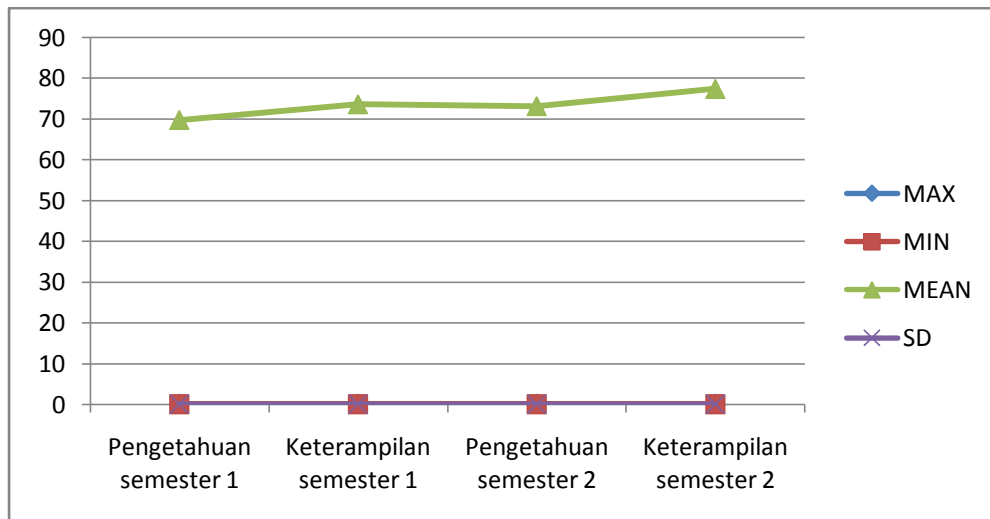
Gambar 6. Grafik nilai pengetahuan semester 1 dan keterampilan semester 1

Gambar 5 dan Gambar 6 menampilkan hasil akhir dari klustering data menggunakan aplikasi *Rapidminer* dan *Excel* untuk mengelompokkan data dalam kluster. Berdasarkan hasil pengujian, disimpulkan bahwa 14 siswa termasuk dalam kluster 0 yang ditandai dengan warna biru. Visualisasi hasil klustering dilakukan dengan menggunakan grafik dan *Scatter Plot*, mirip dengan contoh yang terlihat pada gambar.



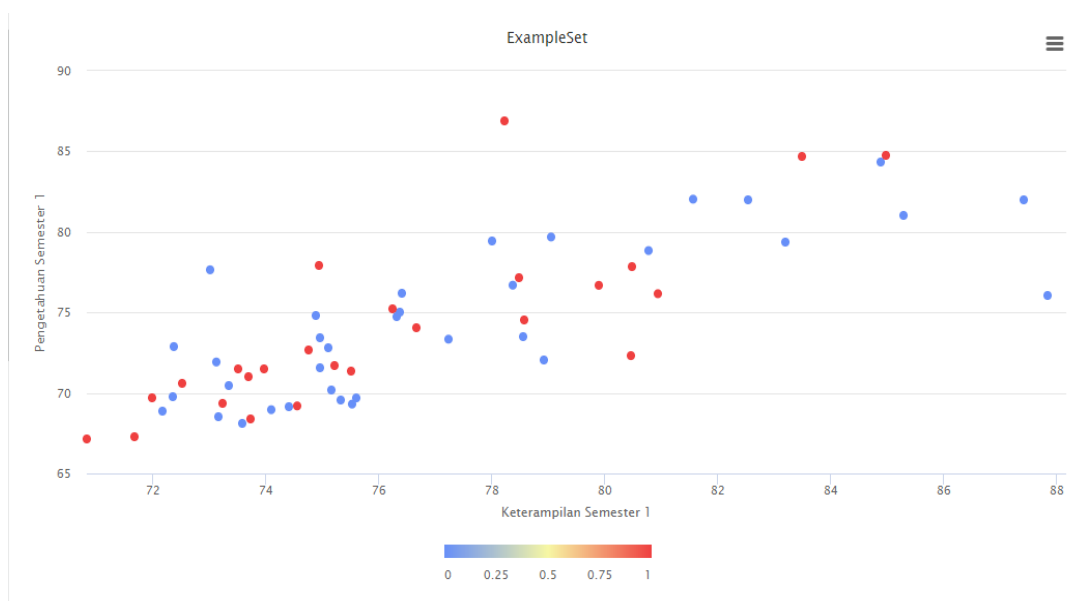
Gambar 7 Scatter Plot Nilai Pengetahuan semester 2 dan Nilai Keterampilan semester 2

Gambar 7 Menunjukkan tampilan dari aplikasi *rapidminer* yang digunakan untuk melakukan pengelompokan data *cluster* nilai pengetahuan semester 2 dan nilai keterampilan semester 2. Berdasarkan hasil pengujian, dapat disimpulkan bahwa terdapat 14 siswa dalam *cluster0* yang ditandai oleh warna biru, *cluster 0* memiliki nilai di sumbu x dari rentang nilai 65-88 dan sumbu y dari 65-88 sedangkan *cluster 1* ditandai dengan warna hijau, memiliki nilai berdasarkan sumbu x dari 65-80 dan sumbu y dari 65-80.



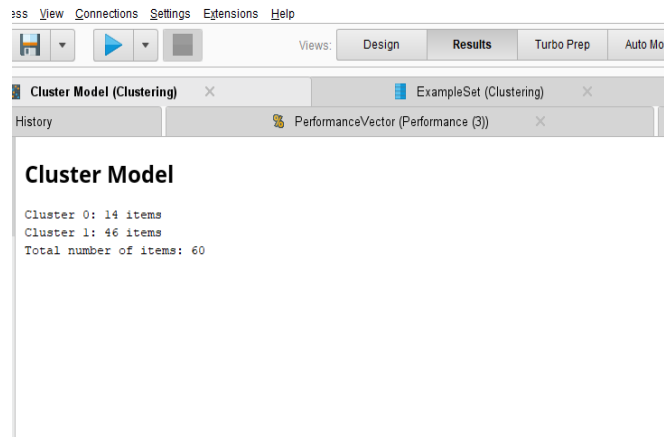
Gambar 8. Grafik nilai pengetahuan semester 2 dan keterampilan semester 2

Gambar 7 dan Gambar 8 menunjukkan hasil akhir dari klustering data menggunakan aplikasi *Rapidminer* dan *Excel* untuk mengelompokkan data dalam kluster. Berdasarkan hasil pengujian, terdapat 46 siswa dalam kluster 1 yang ditandai dengan warna hijau. Visualisasi hasil klustering dilakukan dengan menggunakan grafik dan *Scatter Plot*, serupa dengan contoh yang terlihat pada gambar.



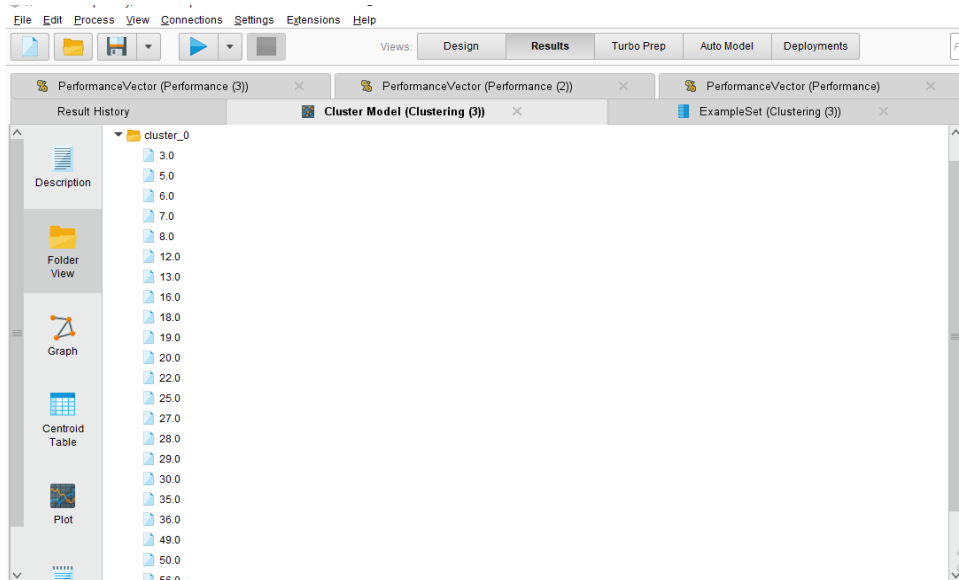
Gambar 9. Scatter Plot Berdasarkan Jenis Kelamin

Gambar 9 menunjukkan tampilan aplikasi *rapidminer* yang digunakan untuk melakukan pengelompokan nilai pengetahuan semester 1 dan nilai keterampilan semester 1 berdasarkan jenis kelamin. Dimana warna biru ditandai sebagai siswa laki-laki dan warna merah ditandai sebagai siswa perempuan. Dalam *cluster 0* dapat disimpulkan bahwa siswa laki-laki lebih banyak yang memiliki nilai tertinggi dibandingkan dengan siswa perempuan.



Gambar 10. Tampilan *Cluster Model*

Gambar 10 merupakan tampilan model kluster yang menampilkan hasil kluster, termasuk tampilan teks yang menunjukkan hasil pengelompokan berdasarkan kluster dan jumlah anggota di setiap kluster. Kluster 0 memiliki 14 item, kluster 1 memiliki 46 item dari jumlah total 60 item.



Gambar 11. Tampilan *folder view*

Gambar 11 tampilan *Folder view* menampilkan data secara rinci dari bagian-bagian kluster, dengan setiap anggota dari dua kluster menunjukkan id variabel. Setelah proses pengelompokan data nilai siswa menggunakan *RapidMiner*.

Evaluasi Hasil *Clustering*

Berdasarkan Tabel 3 di bawah, teknik klustering *k-means* digunakan bersama dengan perhitungan *Indeks Davies-Bouldin (DBI)* untuk menentukan kluster yang optimal. Proses ini melibatkan tiga iterasi percobaan untuk mencapai nilai DBI terbaik.

Tabel 3. Perhitungan *DaviesBouldin Index*

K-means		
K	Avg. within centroid distance	DBITerpilih

2	34.981	0.862	K=2
3	21.450	0.899	
4	15.763	0.891	



Gambar 12. *DaviesBouldin*

Gambar 12 menggambarkan informasi mengenai nilai DBI dan kinerja klusterisasi yang diperoleh dari pengujian menggunakan perhitungan *Davies-Bouldin Index* (DBI) dengan *RapidMiner*. Ditemukan bahwa nilai DBI terbaik terjadi saat $K=2$, dengan nilai DBI sekitar 0,862, yang mendekati 0. Hasil ini menunjukkan bahwa semakin rendah nilai DBI yang diperoleh (yang non-negatif dan ≥ 0), semakin baik klusterisasi yang terjadi. Hasil ini dicapai melalui metode *K-Means* dalam proses klusterisasi menggunakan *RapidMiner*. *Davies-Bouldin Index* dengan tiga kluster. Nilai *Davies-Bouldin Index* yang diperoleh dari dataset adalah 0,862. Berdasarkan hasil pengujian ini, dapat disimpulkan bahwa terdapat kluster yang cocok dalam data nilai semester 1 tahun pelajaran 2021/2022. Nilai pengujian sebesar 0,862 menunjukkan keberhasilan klusterisasi yang baik.

KESIMPULAN

Berdasarkan penelitian tersebut penggunaan algoritma *K-Means* berhasil dalam mengidentifikasi tingkat literasi numerasi siswa di SMP Negeri 4 Mauliru berdasarkan nilai pengetahuan dan keterampilan matematika. Analisis *clustering* berhasil mengelompokkan siswa menjadi dua kelompok (*cluster 0* dan *cluster 1*), *cluster 0* 14 siswa dan *cluster 1* 46 siswa. *Cluster 0* menunjukkan tingkat literasi numerasi yang lebih tinggi dibandingkan dengan *cluster 1*. Analisis dan evaluasi hasil *clustering* menunjukkan bahwa terdapat perbedaan yang signifikan dalam performa siswa dalam pembelajaran matematika di SMP Negeri 4 Mauliru. *Cluster 0*, dengan jumlah siswa sedikit adalah 14 siswa, memiliki nilai rata-rata yang lebih tinggi dan standar deviasi yang rendah, menunjukkan bahwa data nilai siswa tidak bervariasi secara signifikan di *cluster* ini. Hal ini menandakan bahwa siswa dalam *cluster 0* memiliki pemahaman yang lebih baik terhadap konsep matematika dan mampu menyelesaikan soal dengan lebih baik. Sebaliknya, *Cluster 1* terdiri dari 46 siswa yang memiliki tingkat literasi numerasi lebih rendah dibandingkan dengan siswa di *Cluster 0*. Dalam *cluster* ini, nilai rata-rata siswa lebih rendah. Hal ini menandakan bahwa siswa dalam *Cluster 1* mungkin menghadapi kesulitan dalam memahami konsep matematika dan menyelesaikan soal-soal matematika dengan baik. Visualisasi hasil *clustering* memberikan gambaran yang lebih jelas tentang karakteristik setiap kelompok siswa. *Scatter plot* menunjukkan bahwa siswa laki-laki cenderung memiliki nilai lebih tinggi dibandingkan dengan siswa perempuan, terutama pada *cluster 0*. Evaluasi metode *K-Means* menggunakan *Davies-Bouldin Index* (DBI) menunjukkan bahwa algoritma ini efektif dalam mengelompokkan siswa berdasarkan performa mereka. Meskipun nilai DBI sebesar 0,862

tinggi dan tidak mendekati 0, namun hal ini dapat dijelaskan oleh rendahnya nilai standar deviasi dalam data, menunjukkan bahwa data nilai siswa tidak bervariasi secara signifikan. Meskipun DBI tinggi, hasil klusterisasi masih dapat dianggap baik karena nilai standar deviasi yang rendah menunjukkan konsistensi dalam performa siswa.

SARAN

Berdasarkan hasil penelitian ini, terdapat beberapa saran untuk pengembangan lebih lanjut dari proyek tugas akhir. Pertama, pertimbangkan untuk melakukan modifikasi atau penggabungan algoritma *k-means* dengan algoritma lain untuk meningkatkan kinerja klusterisasi. Selain itu, disarankan untuk menambah jumlah data yang digunakan karena *k-means* biasanya memberikan hasil yang lebih baik dengan data yang lebih banyak. Untuk melengkapi hal tersebut, tambahkan variasi data yang dapat mencakup berbagai aspek yang relevan dengan tujuan klusterisasi. Dengan mempertimbangkan variasi data ini, dapat membantu memperkaya analisis dan memastikan keberagaman informasi yang digunakan dalam proyek tugas akhir.

DAFTAR PUSTAKA

- Anwar, S., Suprpti, T., Dwilestari, G., Ali, I., Studi Rekayasa Perangkat Lunak Jln Perjuangan No, P., & Kesambi Kota Cirebon, B. (2022). PENGELOMPOKAN HASIL BELAJAR SISWA DENGAN METODE CLUSTERING K-MEANS Program Studi Sistem Informasi Jln Perjuangan No 10B Kesambi Kota Cirebon 4). *Jurnal Sistem Informasi dan Teknologi Informasi*, 4(2), 60–72.
- Asmana, A., Arie Wijaya, Y., & Martanto, M. (2022). Clustering Data Calon Siswa Baru Menggunakan Metode K-Means Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 552–559.
- Br Sembiring, S. N., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 1(1), 31.
- 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.
- Febrinita, F., Puspitasari, W., & Zaman, W. (2023). Klusterisasi Hasil Belajar Matematika dengan Algoritma K-Means Clustering. *Generation Journal*, 7(2), 116–125.
- 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.
- Ranjawali, R., Talakua, A. C., & Abineno, R. T. (2023). *CLUSTERING STUNTING PADA BALITA DENGAN METODE K- MEANS DI PUSKESMAS KANATANG*. 80–92.
- Sopyan, Y., Lesmana, A. D., & Juliane, C. (2022). Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1464–1470.

Teknik, I., Peminatan, P., Siswa, J., & Algoritma, M. (2021). *Indra Gunawan 3) , Iin Parlina 4) , Irawan 5) 1) 3) 4) 5) Teknik Informatika. 2(2), 1–5.*