Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Optimalisasi Algoritma K-Means untuk Analisis pengelompokan Data Jurusan Siswa Baru Berbasis Numerical Measure

Muhammad Mahda[1], Rudi Kurniawan[2], Tati Suprapti[3]

[1,2,3] Teknik Informatika, STMIK IKMI Cirebon
Email: [1]muhammadmahda.ikmi11@gmail.com, [2]rudi226ikmi@gmail.com, [3]tatisuprapti112004@gmail.com
Email Penulis Korespondensi: muhammadmahda.ikmi11@gmail.com

Article History:
Received Dec 28th, 2024
Revised Dec 31th, 2024
Accepted Jan 25th, 2025

Abstrak

Dalam analisis pengelompokan data, algoritma K-Means adalah teknik yang umum digunakan. Karena memengaruhi kualitas pengelompokan, sangat penting untuk memilih jumlah cluster K yang tepat. Tujuan penelitian ini adalah untuk mengoptimalkan algoritma K-Means, yang menggunakan *Davies-Bouldin Index* (DBI) untuk menilai dua jenis jarak numerik, yaitu *EuclideanDistance* dan *ManhattanDistance*, untuk pengelompokan data jurusan siswa baru. KDD (*Knowledge Discovery in Database*) adalah pendekatan yang digunakan, yang mencakup proses Data Selection, Preprocessing, Transformasi, Data Mining, dan Evaluasi. Dataset jurusan siswa baru dengan cluster K antara 2 dan 10 digunakan untuk eksperimen. Hasil penelitian menunjukkan bahwa *EuclideanDistance* memiliki pemisahan cluster yang lebih baik daripada *ManhattanDistance*, dengan nilai DBI terendah (0.603) pada K=2. Hasil ini menunjukkan bahwa *Euclidean Distance* lebih efektif dalam mengoptimalkan pengelompokan data. Metode ini dapat diterapkan dalam analisis data pendidikan dan bidang lain.

Kata Kunci : Algoritma K-Means, Clustering, Davies-Bouldin Index (DBI), NumericalMeasures, Knowledge Discovery in Database (KDD).

Abstract

In data clustering analysis, K-Means algorithm is a commonly used technique. Since it affects the quality of clustering, it is very important to choose the right number of K clusters. The purpose of this research is to optimise the K-Means algorithm, which uses the Davies-Bouldin Index (DBI) to assess two types of numerical distances, namely EuclideanDistance and ManhattanDistance, for clustering new student major data. KDD (Knowledge Discovery in Database) is the approach used, which includes the processes of Data Selection, Preprocessing, Transformation, Data Mining, and Evaluation. A new student major dataset with K clusters between 2 and 10 was used for experimentation. The results show that Euclidean Distance has better cluster separation than Manhattan Distance, with the lowest DBI value (0.603) at K=2. These results indicate that Euclidean Distance is more effective in optimising data clustering. This method can be applied in educational data analysis and other fields.

Keyword : *K-Means Algorithm, Clustering, Davies-Bouldin Index (DBI), Numerical Measures, Knowledge Discovery in Database (KDD).*

1. PENDAHULUAN

Peminatan siswa merupakan aspek penting dalam pendidikan yang berkontribusi pada prestasi akademik dan kesuksesan karir di masa depan. Proses pemilihan peminatan melibatkan faktor-faktor kompleks seperti kemampuan, minat, serta kecocokan dengan kebutuhan masa depan. Namun, salah satu tantangan utama adalah bagaimana mengelompokkan siswa berdasarkan minat mereka secara efektif dan efisien. Pendekatan berbasis data mining dengan algoritma clustering K-Means telah terbukti mampu mengidentifikasi pola dalam data yang kompleks[1]. Meski demikian, menentukan jumlah *cluster* yang ideal tetap menjadi tantangan, karena mempengaruhi akurasi pengelompokan. Optimasi parameter algoritma K-Means, yang melibatkan evaluasi dengan *Davies-Bouldin Index (DBI)*, menjadi salah satu solusi yang potensial [2]. Penelitian ini menawarkan pendekatan yang lebih terarah dengan menggunakan dua jenis jarak, yaitu *Euclidean Distance* dan *Manhattan Distance*, untuk meningkatkan akurasi pengelompokan data peminatan siswa [3].

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

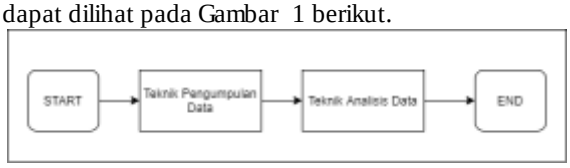
Penelitian terkait sebelumnya menunjukkan bahwa algoritma K-Means telah digunakan secara luas untuk mengelompokkan data pendidikan, mengoptimalkan parameter K-Means untuk menganalisis data peminatan siswa di Sekolah Menengah Kejuruan (SMK)[1]. adapun penelitian Hutagalung dkk., menggunakan K-Means untuk mengidentifikasi siswa kelas unggulan[4]. Sementara itu Nurjannah dkk,. memanfaatkan algoritma ini dalam analisis perkembangan anak[5]. Selain itu, penelitian oleh Maulana dkk,. mengevaluasi pengaruh jenis jarak terhadap hasil clustering, dan Asmana dkk., menyoroti pentingnya algoritma tambahan seperti DBSCAN dan K-Medoids untuk mengatasi data dengan distribusi tidak seragam. Dari penelitian-penelitian ini, terlihat bahwa algoritma K-Means memiliki fleksibilitas yang tinggi, tetapi masih terdapat kesenjangan dalam hal penerapan optimasi parameter untuk memastikan relevansi hasil pengelompokan dengan tujuan pendidikan[6].

Tujuan penelitian ini adalah untuk menerapkan algoritma K-Means dengan optimasi parameter yang mendalam, serta mengevaluasi pengaruhnya terhadap akurasi pengelompokan data peminatan siswa[7]. Dengan menggunakan *Euclidean Distance* dan *Manhattan Distance*, penelitian ini diharapkan dapat memberikan rekomendasi yang lebih akurat bagi pihak sekolah dalam pengambilan keputusan[8]. Selain itu, penelitian ini bertujuan untuk meningkatkan pemahaman tentang bagaimana optimasi parameter dapat memperbaiki hasil pengelompokan dan memberikan kontribusi dalam pengembangan metode data mining yang lebih efektif untuk aplikasi pendidikan[9].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Untuk mengelompokkan data siswa baru di Sekolah Menengah Kejuruan Negeri 6 Kuningan tahun ajaran 2024, menggunakan metode eksperimen dengan pendekatan kuantitatif, serta algoritma K-Means untuk menganalisis data. Adapun tahapan dari penelitian ini dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

2.2 Teknik Pengumpulan Data

Dataset yang digunakan dalam penelitian ini berasal dari data administrasi siswa baru Sekolah Menengah Kejuruan Negeri 6 Kuningan, yang terdiri dari 276 siswa pada tahun 2024 dengan 13 atribut. Data ini dikumpulkan melalui metode pengumpulan data sekunder, yang menjamin kredibilitas dan validitasnya dengan mengambil data dari sumber resmi. Setelah data dikumpulkan, analisis dilakukan dengan mengikuti standar atau metode analisis Knowledge Discovery in Database (KDD). Dataset tersebut dapat dilihat pada Tabel 1 berikut,

Tabel 1. Dataset Penelitian

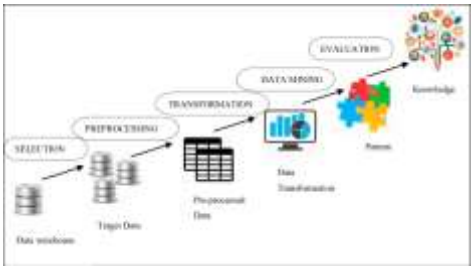
No	Nama Siswa Baru	...	Provinsi	Alamat Lengkap
1	ALICETYA FITRIYANI	…	JAWA BARAT	DUSUN PALEDANG
2	ELISAH	…	JAWA BARAT	DUSUN PAHING
3	ABDUL RAHMAN	…	JAWA BARAT	DUSUN PLAOSAN
…	…	…	…	…
275	NAZWA NAZIKHATUL MAULIDIA	…	JAWA BARAT	DUSUN PLAOSAN
276	ALIFI AKBAR	…	JAWA BARAT	JLN. PASANTRENGANG GIBUG

Adapun dataset lengkap dapat dilihat melalui link https://docs.google.com/spreadsheets/d/1L4FTdgH_DJbCnrsxKSg2vFPsjznY0UjS/edit?usp=drive_link&ouid=107042106278744499984&rtpof=true&sd=true.

2.3 Teknik Analisis Data

Teknik Analisis data dalam penerapan data mining ini menggunakan proses tahapan analisis Knowledge Discovery in Database (KDD) yang terdiri dari Data Selection, Preprocessing, Data Transformation, Data Mining, Pattern Evaluation, dan Knowledge. Adapun Visualisasi dapat dilihat pada Gambar 2 sebagai berikut.

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index



Gambar 2. Tahapan Knowledge Discovery in Database (KDD)[2]

Pada Gambar 2 menunjukkan penjelasan mengenai tahapan Knowledge Discovery In Databases (KDD) sebagai berikut:

a. Data *Selection*
Sebelum pengolahan data, seleksi dilakukan terhadap data administrasi jurusan baru SMKN 6 Kuningan, sesuai dengan kebutuhan *Knowledge Discovery in Database* (KDD).
b. *Preprocessing*
Preprocessing dilakukan untuk memfokuskan data yang akan diproses. Pada tahap ini, data yang tidak konsisten dihilangkan untuk mengurangi kesalahan, diikuti dengan pembersihan data, yaitu menghapus data kosong dan menangani nilai yang hilang, agar tidak menghambat proses pengolahan selanjutnya[3].
c. Data *Transformation*
Pada tahap ini, data nominal diubah menjadi numerik untuk memudahkan pengolahan dengan algoritma K-Means. Proses transformasi dilakukan menggunakan operator Nominal to Numerical dengan parameter a subset, mencakup atribut seperti Alamat Lengkap, Asal Sekolah, Jenis Kelamin, Kabupaten/Kota, Kecamatan, Kelurahan, Nama Siswa Baru, Pilihan 1, Pilihan 2, Provinsi, dan Tempat Lahir.
d. Data Mining
Pada tahap ini, algoritma K-Means digunakan untuk klasterisasi dan diintegrasikan dengan operator Optimize Parameter Grid. Jumlah klaster dihitung, dengan parameter numerik menggunakan *Euclidean Distance* dan *Manhattan Distance*. Evaluasi dilakukan dengan menggunakan Davies-Bouldin Index (DBI) untuk menentukan hasil klaster optimal. Operator *Optimize Parameter Grid* mencakup operator K-Means dan *Cluster Distance Performance*.
e. *Pattern Evaluation*
Pada tahap ini, hasil data mining dievaluasi untuk memastikan keakuratannya. Jika hasilnya tidak konsisten, alternatif dapat diambil, seperti memperbaiki teknik data mining atau menerima hasil yang tidak sesuai harapan.
f. *Knowledge*
Bertujuan untuk mengolah data guna menemukan pola, informasi, dan pengetahuan yang bermanfaat dari kumpulan data besar.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

3.1.1 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari administrasi siswa baru SMKN 6 Kuningan tahun 2024, dengan 276 siswa dan 13 atribut. Penelitian ini fokus pada atribut yang berpengaruh terhadap pengelompokan siswa dan juga perhitungan jarak yang digunakan. Tabel 2 menunjukkan data atribut yang ada pada dataset.

Tabel 2. Atribut dan Type Atribut

No.	Atribut	Type
1	Nama Siswa Baru	Polynominal
2	Alamat Lengkap	Polynominal
3	Asal Sekolah	Polynominal
4	Jenis Kelamin	Polynominal
5	Kabupaten/Kota	Polynominal
6	Kecamatan	Polynominal
7	Kelurahan	Polynominal
8	Pilihan 1	Polynominal
9	Pilihan 2	Polynominal
10	Provinsi	Polynominal
11	RT	Integer
12	RW	Integer

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

3.1.2 Teknik Analisis Data

a. Data *Selection*

Tahap pertama adalah Data Selection, di mana operator Read Excel digunakan untuk mengimpor data dari file Excel (.xlsx atau .xls) ke repository Rapidminer AI Studio 2024. Visualisasi operator ini dapat dilihat pada Gambar 3.

Gambar 3. Operator *Read Excel*

Pada operator Read Excel parameters yang digunakan dapat dilihat pada Tabel 3. sebagai berikut.

Tabel 3. Parameters *Read Excel*

No.	Parameters	Isi
1	Sheet Selection	Sheet Number
2	Sheet Number	1
3	Import Cell	A1
4	Encoding	SYSTEM
5	Header Row	1
6	Date Format	-
7	Time Zone	SYSTEM
8	Locale	English

Dari hasil pembacaan operator Read Excel didapat informasi yang ditunjukkan pada Tabel 4.

Tabel 4. Hasil Operator Read Excel

No	Uraian	Keterangan
1	Record	276
2	Spesial Attribut	0
3	Reguler Attribute	13
4	Attribute :	
	Nama Siswa Baru	
	Alamat Lengkap	Polynominal, Missing 0
	Asal Sekolah	Polynominal, Missing 0
	Jenis Kelamin	Polynominal, Missing 0
	Kabupaten/Kota	Polynominal, Missing 0
	Kecamatan	Polynominal, Missing 0
	Kelurahan	Polynominal, Missing 0
	Pilihan 1	Polynominal, Missing 0
	Pilihan 2	Polynominal, Missing 33
	Provinsi	Polynominal, Missing 0
	RT	Integer, Missing 0
	RW	Integer, Missing 0

Langkah selanjutnya, operator *Set Role* digunakan untuk menetapkan peran pada kolom dataset, seperti mengubah atribut Nama Siswa Baru menjadi ID. Tampilan operator ini dapat dilihat pada Gambar 4.

Gambar 4. Operator Set Role

Dalam penggunaan operator Set Role parameters yang digunakan dapat dilihat pada Tabel 5 berikut.

Tabel 5. Parameters dan Atribut Set Role

No	Parameters	Isi	
1	Attribute Name	No.	
2	Target Role	id	
3	Set Adsitional Role	attribut name	target role
		Nama Siswa Baru	id

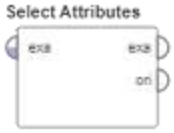
Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Adapun hasil dari penggunaan operator Set Role dapat dilihat pada Tabel 6 berikut.

Tabel 6. Hasil Set Role

No	Uraian	Keterangan
1	Record	276
2	Spesial Attribut	1
3	Reguler Attribute	12
4	Attribute :	
	Nama Siswa Baru(id)	Polynominal, Missing 0
	Alamat Lengkap	Polynominal, Missing 0
	Asal Sekolah	Polynominal, Missing 0
	Jenis Kelamin	Polynominal, Missing 0
	Kabupaten/Kota	Polynominal, Missing 0
	Kecamatan	Polynominal, Missing 0
	Kelurahan	Polynominal, Missing 0
	Pilihan 1	Polynominal, Missing 0
	Pilihan 2	Polynominal, Missing 33
	Provinsi	Polynominal, Missing 0
	RT	Integer, Missing 0
	RW	Integer, Missing 0

Tahapan selanjutnya dalam data selection adalah penggunaan operator Select Attributes, yang bertujuan untuk memilih atribut relevan, mengurangi dimensi dataset, mempercepat komputasi, dan mencegah overfitting. Operator ini ditunjukkan pada Gambar 5 berikut.



Gambar 5. Operator Select Attributes

Tabel 7 menunjukkan parameter operator Select Attributes yang digunakan dalam aplikasi Rapidminer AI Studio 2024, yang ditunjukkan pada Gambar 5 berikut.

Tabel 7. Parameters Select Attributes

No	Parameters	Isi
1	Type	include attribute
2	Attribute filter types	a subset
3	Select attribute	Alamat Lengkap
		Asal Sekolah
		Jenis Kelamin
		Kabupaten/Kota
		Kecamatan
		Kelurahan
		Pilihan 1
		Pilihan 2
		Provinsi
		RT
		RW

Dari hasil pembacaan operator Select Attributes didapat informasi yang dapat dilihat pada Tabel 8 berikut.

Tabel 8. Hasil Penggunaan Select Attributes

No	Uraian	Keterangan
1	Record	276
2	Spesial Attribut	1
3	Reguler Attribute	12
4	Attribute :	
	Nama Siswa Baru(id)	Polynominal, Missing 0

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Alamat Lengkap	Polynominal, Missing 0
Asal Sekolah	Polynominal, Missing 0
Jenis Kelamin	Polynominal, Missing 0
Kabupaten/Kota	Polynominal, Missing 0
Kecamatan	Polynominal, Missing 0
Kelurahan	Polynominal, Missing 0
Pilihan 1	Polynominal, Missing 0
Pilihan 2	Polynominal, Missing 33
Provinsi	Polynominal, Missing 0
RT	Integer, Missing 0
RW	Integer, Missing 0

Adapun model pada langkah data Selection dapat dilihat pada Gambar 6 sebagai berikut.

Gambar 6. Model Process Data Selection pada Rapidminer AI Studio

b. *Preprocessing*

Langkah selanjutnya adalah *pre-processing* data dengan nilai yang hilang. Atribut Pilihan 2 memiliki 33 data yang hilang, seperti terlihat pada Tabel 8. *Pre-processing* dilakukan untuk mengatasi nilai *missing* pada atribut dataset, menggunakan operator *Replace Missing Values*, yang dapat dilihat pada Gambar 7.

Gambar 7. Operator Replace Missing Values

Pada Tabel 9 Menunjukan parameters yang digunakan untuk mengubah data yang hilang menjadi data yang data digunakan.

Tabel 9. Parameters *Replace Missing Values*

No	Parameters	Isi
1	Attribute filter types	a subset
2	Default	Average
3	Attribute:	
	Nama Siswa Baru(id)	Polynominal, Missing 0
	Alamat Lengkap	Polynominal, Missing 0
	Asal Sekolah	Polynominal, Missing 0
	Jenis Kelamin	Polynominal, Missing 0
	Kabupaten/Kota	Polynominal, Missing 0
	Kecamatan	Polynominal, Missing 0
	Kelurahan	Polynominal, Missing 0
	Pilihan 1	Polynominal, Missing 0
	Pilihan 2	Polynominal, Missing 33
	Provinsi	Polynominal, Missing 0
	RT	Integer, Missing 0
	RW	Integer, Missing 0

Adapun hasil dari penggunaan operator Replace Missing Values dapat dilihat pada Tabel 10 berikut.

Tabel 10. Hasil Parameters Replace Missing Values

No	Uraian	Keterangan
1	Record	276
2	Spesial Attribut	1
3	Reguler Attribute	12
4	Attribute :	

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Nama Siswa Baru(id)	Polynominal, Missing 0
Alamat Lengkap	Polynominal, Missing 0
Asal Sekolah	Polynominal, Missing 0
Jenis Kelamin	Polynominal, Missing 0
Kabupaten/Kota	Polynominal, Missing 0
Kecamatan	Polynominal, Missing 0
Kelurahan	Polynominal, Missing 0
Pilihan 1	Polynominal, Missing 0
Pilihan 2	Polynominal, Missing 0
Provinsi	Polynominal, Missing 0
RT	Integer, Missing 0
RW	Integer, Missing 0

Model proses pada Rapidminer AI Studio pada Langkah Pra-pemprosesan dapat dilihat pada Gambar 8 berikut.

Gambar 8. Model Preprocessing

c. Data *Transformation*

Tahap *transformasi* bertujuan mengubah struktur data menjadi format yang sesuai untuk analisis. Operator *Nominal to Numerical* digunakan untuk mengubah atribut non-numerik menjadi numerik, seperti terlihat pada Gambar 9.

Gambar 9. Operator *Nominal to Numerical*

Parameter pada operator Nominal to Numerical yang digunakan tampak pada Tabel 11, berikut.

Tabel 11. Parameters *Nominal to Numerical*

No	Parameters	Isi
1	Attribute filter types	a subset
2	Coding type	unique integer
3	Attribute :	
	Alamat Lengkap	
	Asal Sekolah	
	Jenis Kelamin	
	Kabupaten/Kota	
	Kecamatan	
	Kelurahan	
	Pilihan 1	
	Pilihan 2	
	Provinsi	

Hasil dari pembacaan operator *Nominal to Numerical* didapat informasi yang dapat dilihat pada Tabel 12 berikut.

Tabel 12. Penggunaan *Nominal to Numerical*

No	Uraian	Keterangan
1	Record	276
2	Spesial Attribut	1
3	Reguler Attribute	12
4	Attribute :	
	Alamat Lengkap	Integer, Missing 0
	Asal Sekolah	Integer, Missing 0
	Jenis Kelamin	Integer, Missing 0

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

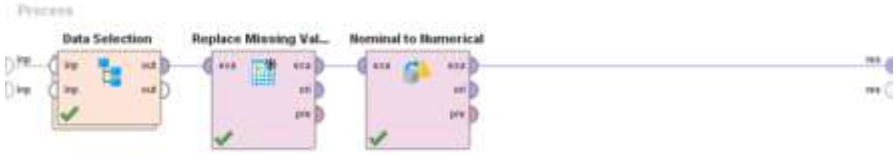
Kabupaten/Kota	Integer, Missing 0
Kecamatan	Integer, Missing 0
Kelurahan	Integer, Missing 0
Pilihan 1	Integer, Missing 0
Pilihan 2	Integer, Missing 0
Provinsi	Integer, Missing 0

Adapun hasil dari penggunaan operator dapat dilihat pada sampel Tabel 13 sebagai berikut.

Tabel 13. Hasil *Nominal to Numerical*

Uraian	Keterangan	
	Sesudah	Sebelum
Jenis Kelamin	0	PEREMPUAN
	1	LAKI-LAKI
Pilihan 1	0	PEMASARAN
	1	DESAIN KOMUNIKASI VISUAL
	2	OTOMOTIF
	3	PERHOTELAN
	4	TEKNIK LOGISTIK
Provinsi	0	JAWA BARAT

Model proses pada Rapidminer AI Studio pada Langkah *Transformation* dapat dilihat pada Gambar 10 berikut.



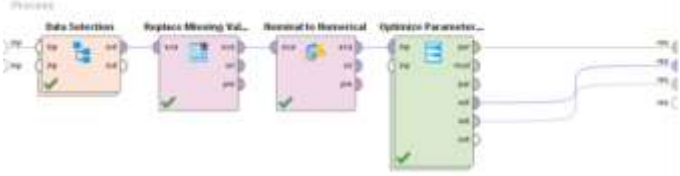
Gambar 10. Model *Transformation*

d. Data Mining

Proses data mining menggunakan algoritma K-Means untuk mengelompokkan data jurusan siswa baru, dengan RapidMiner sebagai software yang digunakan. Operator *Optimize Parameters Grid* diterapkan untuk menemukan parameter terbaik, seperti nilai K dan juga perhitungan jarak, guna menghasilkan pengelompokan yang akurat, pada Gambar 11 menunjukkan visualisasi operator tersebut.

Gambar 11. Operator *Optimize Parameters Grid*

Model proses pada Rapidminer AI Studio pada Langkah Data Mining menggunakan Operator Parameters Grid dapat dilihat pada Gambar 12 berikut.



Gambar 12. Model Proses Optimize Parameters Grid

Adapun penggunaan operator *Optimize Parameters Grid* terdapat beberapa parameters yang digunakan, dapat dilihat pada Tabel 14 sebagai berikut.

Tabel 14. Parameters *Optimize Parameters Grid*

Operator	Parameters	Selected Parameters	Value List
Clustering (K-Means)	K	Clustering.K Min = 2 Max = 10 Step = 10	

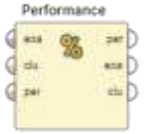
Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

	Numerical Measure	Scale = Linear Clustering.Numerical Measure	EuclideanDistance ManhattanDistance
Performance(Clustering Distance Performance)	Main Criterion	Performance.Main Criterion	Davies Bouldin

Operator *Optimize Parameters Grid* melibatkan dua subproses utama, Clustering (K-Means) dan Performance (Clustering Distance Performance), yang bekerja untuk mengoptimalkan clustering. Gambar 13 menampilkan visualisasi operator K-Means.

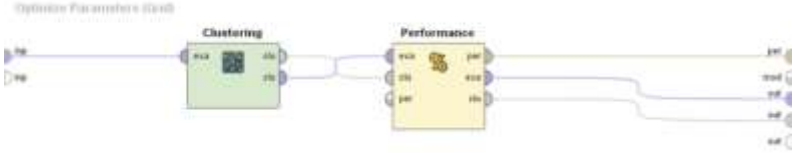
Gambar 13. Operator K-means

Adapun pada Gambar 14 menunjukkan visualisasi operator *Performance*, sebagai berikut.



Gambar 14. *Operator Performance*

Model Subproses *Optimize Parameters Grid* pada Rapidminer AI Studio pada Langkah Data Mining menggunakan dapat dilihat pada Gambar 15 sebagai berikut.



Gambar 15. Model Subproces

Berdasarkan Gambar 15 menggunakan algoritma K-Means untuk mengelompokkan data jurusan siswa baru. Metode ini melibatkan pengujian dengan variasi jumlah cluster (K=2 hingga K=10) dan penggunaan dua metrik jarak (Euclidean dan Manhattan) untuk mengukur kemiripan data. Hasil pengujian, yang disajikan dalam Tabel 15, merupakan pola atau kelompok yang signifikan dalam data jurusan siswa baru.

Tabel 15. Hasil *Optimize Parameters Grid*

No.	Cluster K	Clustering.numerical_measure	Davies Bouldin	No.	Cluster K	Clustering.numerical_measure	Davies Bouldin
1	2	EuclideanDistance	0.603	10	2	ManhattanDistance	0.612
2	3	EuclideanDistance	0.752	11	3	ManhattanDistance	0.749
3	4	EuclideanDistance	0.859	12	4	ManhattanDistance	0.899
4	5	EuclideanDistance	0.875	13	5	ManhattanDistance	0.847
5	6	EuclideanDistance	0.883	14	6	ManhattanDistance	0.905
6	7	EuclideanDistance	0.937	15	7	ManhattanDistance	0.949
7	8	EuclideanDistance	0.979	16	8	ManhattanDistance	0.983
8	9	EuclideanDistance	1.045	17	9	ManhattanDistance	1.124
9	10	EuclideanDistance	0.988	18	10	ManhattanDistance	1.089

Pada Tabel 15 menunjukkan hasil evaluasi Davies Bouldin Index (DBI) untuk berbagai jumlah klaster (K) menggunakan metrik jarak EuclideanDistance dan ManhattanDistance. Nilai DBI cenderung meningkat seiring bertambahnya jumlah klaster, dengan hasil terbaik pada K = 2 menggunakan *Euclidean Distance* dengan nilai DBI terendah 0.603, menunjukkan kinerja clustering yang optimal. Untuk menentukan perhitungan jarak Euclidean dan Mathattan menggunakan persamaan 1 dan 2 sebagai berikut.

$$d(i,j) = \sqrt{(x_{1i} - x_{1j}) + (x_{2i} - x_{2j}) + \cdots + (x_{ki} - x_{kj})^2} \quad (1)$$

di mana d(i, j) adalah jarak data antara pusat (i) dan (j), xni adalah data atribut ke-i pada data ke-k, dan xnj adalah data atribut ke-j pada data ke-k [10].

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 8 ; Nomor 1 ; Januari 2025 ; Page 25-35
E-ISSN : 2615-5133 **; P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

$$d(i,j) = |x_{i,1} - x_{j,1}| + |x_{i,2} - x_{j,2}| + \cdots + |x_{i,n} - x_{j,n}| \quad (2)$$

Di mana d(i, j) adalah jarak data antara pusat (i) dan (j), xni adalah data atribut ke-i pada data atribut ke-n, dan xnj adalah data atribut ke-j pada data atribut ke-n [11].

Pada Tabel 16 menampilkan Performance Vector yang mendukung hasil ini.

Tabel 16. *Performance Vector*

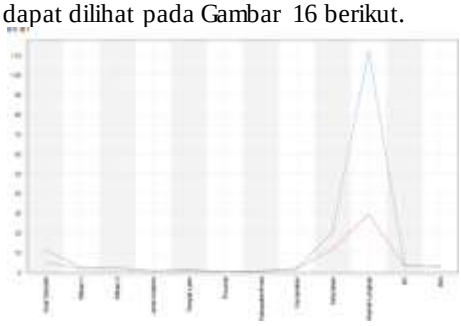
PerformanceVector
Avg. within centroid distance: -735.066
Avg. within centroid distance_cluster_0: -930.904
Avg. within centroid distance_cluster_1: -555.548
Davies Bouldin: -0.603

Dari hasil Tabel 15 bahpa pengelompokan terbaik berada pada K=2, dan Model cluster yang dihasilkan dapat dilihat pada Tabel 17 berikut.

Tabel 17. Cluster Model

No	Cluster	Keterangan
1	Cluster 0	132 items
2	Cluster 1	144 items
	total	276 items

Selain itu pada Proses data mining menghasilkan visualisasi titik centroid masing-masing cluster untuk mengidentifikasi karakteristiknya dapat dilihat pada Gambar 16 berikut.



Gambar 17. Hasil Grafik Atribut Data

e. *Pattern Evaluation*
Menunjukkan algoritma K-Means Clustering dengan k=2 mencapai hasil optimal yang dibuktikan dengan nilai Davies-Bouldin Index (DBI) sebesar 0.603 menggunakan metrik *Euclidean Distance*, dimana nilai ini mencerminkan pemisahan cluster yang baik dan konsistensi internal yang tinggi, didukung oleh parameter optimal dari Optimize Parameters Grid, sehingga membuktikan efektivitas algoritma dalam pengelompokan data[12].

3.2 Pembahasan

Hasil eksperimen dengan RapidMiner AI Studio 2024 menggunakan K-Means Clustering untuk pengelompokan data jurusan siswa baru ditunjukkan pada Tabel 15. Dua jenis jarak (Euclidean dan Manhattan) digunakan untuk menentukan nilai K terbaik, dengan jumlah cluster (K) antara 2 dan 10. Davies-Bouldin Index (DBI) digunakan untuk mengukur kualitas cluster, dan nilai DBI terendah (0.603) ditemukan pada *Euclidean Distance* dengan K=2, menunjukkan pemisahan cluster yang lebih baik. Menurut Zhou dkk, nilai DBI yang lebih rendah menunjukkan pemisahan cluster yang lebih jelas dan homogen, sehingga K=2 dengan *Euclidean Distance* adalah opsi terbaik berdasarkan DBI terendah[13].

4. KESIMPULAN

Berdasarkan Davies-Bouldin Index (DBI), nilai K terendah adalah K=2 dengan DBI 0.603, yang menunjukkan pemisahan cluster terbaik menggunakan Euclidean Distance. Oleh karena itu, jumlah cluster ideal untuk dataset ini adalah K=2.

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat terselesaikan dengan baik. Terima kasih yang mendalam penulis sampaikan kepada kedua orang tua dan keluarga atas doa dan dukungan yang tiada henti. Penulis juga mengucapkan terima kasih kepada Bapak/Ibu Dosen Pembimbing

yang telah memberikan arahan dan bimbingan selama proses penelitian. Tidak lupa ucapan terima kasih kepada rekan-rekan yang telah memberikan semangat dan bantuan hingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

[1] R. Sidik, N. Suarna, and A. Rinaldi Dikananda, "Analisa Data Set Peminatan Siswa Menggunakan Algoritma K-Means Dengan Optimize Parameter Di Sekolah Menengah Kejuruan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1197–1203, 2023, doi: 10.36040/jati.v7i2.6335.

[2] R. B. Ardi, F. E. Nastiti, and S. Sumarlinda, "ALGORITMA K-MEANS CLUSTERING UNTUK SEGMENTASI PELANGGAN (STUDI KASUS : FASHION VIRAL SOLO)," *INFOTECH J.*, vol. 9, no. 1, pp. 124–131, 2023, doi: https://doi.org/10.31949/infotech.v9i1.5214.

[3] T. Maulana, R. Astuti, and Y. Arie Wijaya, "Implementasi Algoritma K-Means Dengan Optimize Parameter Grid Pada Data Kecelakaan Lalu Lintas Di Kota Cirebon," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 310–317, 2024, doi: 10.36040/jati.v8i1.8430.

[4] J. Hutagalung, Y. H. Syahputra, and Z. P. Tanjung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022, doi: 10.35957/jatisi.v9i1.1516.

[5] E. Nurjannah, M. Nasution, and R. Muti, "Data Mining Clustering Analysis of Child Growth and Development Using the K-Means Method," vol. 8, no. 3, pp. 1909–1919, 2024, doi: https://doi.org/10.33395/sinkron.v8i3.13817 e-ISSN.

[6] A. Asmana, Y. Arie Wijaya, and M. Martanto, "Clustering Data Calon Siswa Baru Menggunakan Metode K-Means Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 552–559, 2022, doi: 10.36040/jati.v6i2.5236.

[7] M. Zubair, M. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, "An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling," *Ann. Data Sci.*, vol. 11, no. 5, pp. 1525–1544, 2022, doi: 10.1007/s40745-022-00428-2.

[8] S. S. Patel, N. Kumar, J. Aswathy, S. K. Vaddadi, S. A. Akbar, and P. C. Panchariya, "K-Means Algorithm: An Unsupervised Clustering Approach Using Various Similarity/Dissimilarity Measures BT - Intelligent Sustainable Systems," J. S. Raj, R. Palanisamy, I. Perikos, and Y. Shi, Eds., Singapore: Springer Singapore, 2022, pp. 805–813.

[9] S. N. Safitri, H. Setiadi, and E. Suryani, "Educational Data Mining Using Cluster Analysis Methods and Decision Trees based on Log Mining," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 3 SE-Information Systems Engineering Articles, Jul. 2022, doi: 10.29207/resti.v6i3.3935.

[10] A. Chaerudin, D. T. Murdiansyah, and M. Imrona, "Implementation of K-Means ++ Algorithm for Store Customers Segmentation Using Neo4J," vol. 6, no. April, pp. 53–60, 2021, doi: 10.34818/indojc.2021.6.1.547.

[11] D. Jollyta, D. Priyanto, A. Hajjah, and Y. N. Marlim, "Comparison of Distance Measurements Based on k-Numbers and Its Influence to Clustering," vol. 23, no. 1, pp. 93–102, 2023, doi: 10.30812/matrik.v23i1.3078.

[12] F. Nie, Z. Li, R. Wang, and X. Li, "An Effective and Efficient Algorithm for K-Means Clustering With New Formulation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3433–3443, 2023, doi: 10.1109/TKDE.2022.3155450.

[13] S. Zhou, F. Liu, and W. Song, "Estimating the Optimal Number of Clusters Via Internal Validity Index," *Neural Process. Lett.*, vol. 53, no. 2, pp. 1013–1034, 2021, doi: 10.1007/s11063-021-10427-8.