

IMPLEMENTASI ALGORITMA K-MEANS PADA PENGELOMPOKAN DATA PENERIMAAN PESERTA DIDIK BARU DI SMKN 1 BALONGAN

Suiroh¹, Rini Astuti², Fadhil Muhamad Basysyar³

¹Teknik Informatika, STMIK IKMI Cirebon

²Sistem Informasi, STMIK LIKMI Bandung

³Sistem Informasi, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat 45131, Indonesia.

Suirohikmi2020@gmail.com

ABSTRAK

Penerimaan Peserta Didik Baru (PPDB) di SMKN 1 Balongan memiliki peranan penting dalam mengelola lembaga pendidikan yang berlangsung setiap tahun menjelang dimulainya tahun ajaran baru. Masalah pada penelitian ini yaitu SMKN 1 Balongan masih sering menemui strategi yang kurang tepat dalam promosi penerimaan siswa baru dikarenakan belum memiliki informasi dari hasil pengelompokan data penerimaan peserta didik baru yang mendaftar. Penelitian ini bertujuan untuk mendapatkan informasi dari hasil pengelompokan data penerimaan peserta didik baru dengan melalui penerapan algoritma *K-Means Clustering*. Selain itu, penelitian ini bertujuan untuk mendapatkan parameter *Numerical Measures Type* apa yang menghasilkan kelompok terbaik pada data PPDB. Pada penelitian ini metode yang digunakan yaitu *Knowledge Discovery in Database (KDD)*. Hasil dari pengelompokan menunjukkan bahwa dari jumlah *cluster* K2 sampai K10 menghasilkan K mana yang terbaik performanya yaitu yang mendekati angka 0. Nilai K yang paling mendekati angka 0 adalah K2. diperoleh hasil terbaiknya yaitu pada cluster k-2 dengan nilai DBI sebesar 0,508 untuk parameter *Numerical Measures Type ChebychevDistance dan Correlaton Matrix*. Untuk mengevaluasi kinerja algoritma *K-means* dapat dilihat dari *performance*, dimana nilai *Davies Bouldin* yang mendekati 0 mengindikasikan kualitas algoritma yang semakin baik. Dengan demikian, maka jumlah *cluster* terbaik dalam percobaan ini adalah K2 atau yang mendekati 0.

Kata kunci : Analisis data, PPDB, SMKN 1 Balongan, K-Means Clustering, Data Mining, Knowledge Discovery in Database (KDD).

1. PENDAHULUAN

Kemajuan teknologi informasi sekarang sudah semakin berkembang pesat dan hampir mencakup di segala bidang kehidupan [1], termasuk dalam bidang politik, ekonomi, kebudayaan, seni, dan bahkan pendidikan [2]. Dimana pendidikan sendiri merupakan proses untuk membentuk karakteristik dan menciptakan mutu yang berkualitas [3]. Institusi pendidikan (sekolah) pada tingkat menengah kejuruan (SMK) menjadi pintu gerbang dalam memberikan bekal bagi peserta didik untuk memasuki dunia industri atau melanjutkan studi ke perguruan tinggi [4]. Seiring perkembangan teknologi, PPDB juga semakin terdorong untuk mengadopsi sistem yang efisien dan transparan, menjadikan penerapan teknologi informasi dalam proses ini sebagai langkah yang semakin relevan dan mendukung terwujudnya pendidikan yang inklusif serta memberikan akses yang lebih adil bagi seluruh masyarakat [5].

Penerimaan Peserta Didik Baru (PPDB) adalah salah satu kegiatan yang rutin diselenggarakan setiap tahun sebagai media untuk mendaftarkan calon peserta didik baru [6]. Di Dalam pendidikan, data sangat diperlukan untuk melengkapi setiap proses administrasi siswa mulai dari diterimanya sebagai siswa aktif sampai akhirnya menjadi alumni [7]. Proses penerimaan peserta didik baru di SMKN 1 Balongan masih sering menemui strategi yang kurang tepat dalam promosi penerimaan siswa baru dikarenakan belum memiliki informasi dari hasil

pengelompokan data penerimaan peserta didik baru yang mendaftar [8]. Untuk menarik lebih banyak calon peserta didik baru setiap tahunnya, perlu ditunjukkan upaya dan strategi yang efektif. Sejumlah besar data dihasilkan dalam bentuk profil siswa yang terdaftar selama proses penerimaan peserta didik baru [9]. Pengolahan data dalam ukuran besar yang didalamnya terdapat banyak record tidak bisa dilakukan dengan mudah dan dalam waktu yang singkat untuk mendapatkan informasi yang diinginkan. Tentunya pengolahan data ini dapat dilakukan dengan menggunakan tool untuk mengolah data dalam ukuran yang besar, yaitu dengan menerapkan metode *Data Mining*. [10]. *Data mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data [9].

Penelitian ini bertujuan melakukan pengelompokan (*Clustering*) untuk mendapatkan informasi dari hasil analisis pengelompokan. selain itu, penelitian ini juga bertujuan untuk mendapatkan Parameter *Numerical Measures Type* apa yang menghasilkan kelompok terbaik pada data penerimaan peserta didik baru dengan melalui penerapan algoritma *K-Means Clustering*. Jumlah *Cluster* ditentukan mengacu pada *Davies Bouldin Index (DBI)* [11]. Melalui tahapan analisis tersebut, diharapkan dapat memanfaatkan informasi yang diperoleh sebagai dasar strategi promosi untuk penerimaan peserta didik baru di tahun ajaran berikutnya.

Clustering merupakan metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain [12]. *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (unsupervised) [13]. Tujuan dari proses *clustering* ini adalah untuk meminimalkan fungsi tujuan yang ditetapkan, yang umumnya bertujuan untuk mengurangi variasi di dalam suatu cluster dan meningkatkan variasi antar cluster [14].

K-Means merupakan salah satu algoritma yang banyak digunakan dalam pengelompokan karena kesederhanaannya dan efisiensi secara partisi yang memisahkan data ke dalam kelompok yang berbeda-beda [12]. Proses *clustering* menggunakan algoritma *K-Means* lebih optimal dari segi waktu dan output yang dihasilkan lebih berkualitas meskipun menggunakan data dalam jumlah besar [15].

Penilaian Indeks *Davies-Bouldin* menampilkan sistem evaluasi klaster internal, di mana hasil klaster yang baik atau buruk ditentukan oleh jumlah dan kedekatan hasil klaster. Semakin mendekati 0 angka DBI nya semakin baik *clustering* yang dihasilkan [16]. Metode DBI dipilih karena dapat diimplementasikan dalam semua data baik ukuran besar dan kecil sehingga sangat cocok dan dapat diterima untuk perhitungan akurasi *clustering* [14].

2. TINJAUAN PUSTAKA

Pada penelitian yang dilakukan oleh Maukar, Anastasia L Marisa, Fitri Widodo, dan Anang Aris yang berjudul “Analisis Data Penerimaan Mahasiswa Baru Berbasis *K-Means*”. Hasil *clustering* ini memberikan wawasan tentang pola minat calon mahasiswa baru dan dapat digunakan untuk mengatur strategi promosi universitas kepada sekolah menengah atas. Metode ini efisien dalam mengelompokkan calon mahasiswa berdasarkan jurusan, asal sekolah, dan IPK, sehingga memberikan wawasan yang berguna untuk pengambilan keputusan strategis dan perencanaan promosi universitas [10].

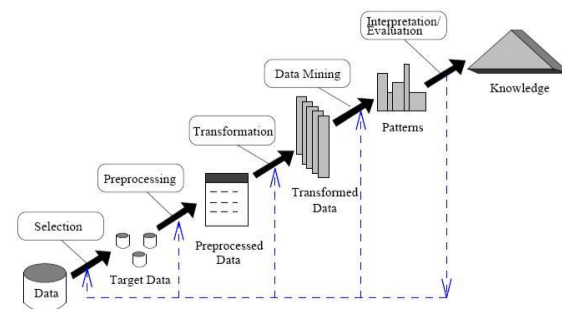
Penelitian dengan judul “*Clustering* Data Calon Siswa Baru Menggunakan Metode *K-Means* Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon” yang dilakukan oleh Asmana, Yudhistira Arie Wijaya, dan Martanto pada tahun 2022. Dalam hal ini yang menjadi permasalahan adalah SMK Wahidin Kota Cirebon masih sering menemui strategi yang tidak tepat dalam promosi penerimaan siswa dikarenakan tidak memiliki data induk dalam bentuk pengelompokan data asal sekolah siswa yang mendaftar di SMK Wahidin Kota Cirebon. Hasil penelitian ini memberikan wawasan yang bisa dijadikan sumber informasi dari hasil pengelompokan yang dapat dijadikan strategi promosi ditahun ajaran yang akan datang di SMK Wahidin Kota Cirebon [10].

Penelitian dengan judul penelitian “Penerapan Data Mining Pada Penerimaan Mahasiswa Baru Dengan Algoritma *K-Means Clustering*” yang dilakukan oleh Septian Isnanto dan Suryarini Widodo

pada tahun 2021. Hasil dari penelitian ini dapat disimpulkan bahwa Penerapan algoritma *K-means* dalam penelitian ini efisien karena mampu mengelompokkan data akademik dari berbagai program studi dengan hasil yang bermakna. Algoritma ini dapat menghasilkan cluster-cluster yang memiliki karakteristik yang berbeda, sehingga memberikan informasi yang berguna bagi panitia penerimaan mahasiswa baru untuk memprioritaskan jurusan dan jenis sekolah tertentu pada proses seleksi penerimaan mahasiswa baru [17].

3. METODE PENELITIAN

Teknik analisis data merupakan metode atau pendekatan yang digunakan untuk mengolah, menginterpretasi, dan menyajikan data yang telah dikumpulkan dalam rangka mendapatkan informasi yang bermakna. Teknik analisis data berperan penting dalam menghasilkan temuan, pola, atau informasi yang relevan dari dataset. Dalam penelitian mengenai Analisis *Clustering* Data Penerimaan Peserta Didik Baru Di Smkn 1 Balongan Dengan Menggunakan Algoritma *K-Means*, terdapat beberapa langkah teknis yang dapat dilakukan dalam analisis data. Adapun Proses *Knowledge Discovery in Database* (KDD) pada penelitian ini dapat dilihat pada Gambar 2 dibawah ini.



Gambar 1. *Knowledge Discovery in Database* (KDD)

Berikut adalah penjelasan dari setiap tahapannya:

3.1. Selection

Pada tahap ini, dilakukan seleksi data yang relevan berdasarkan data penerimaan peserta didik baru di SMKN 1 Balongan. Data yang berhasil dikumpulkan mencakup informasi seperti nama pendaftar, tahun ajaran, jenis kelamin, tempat lahir, tanggal lahir, bulan lahir, tahun lahir, kejuruan, jumlah rombongan, jalur pendaftar, dan asal sekolah. Data yang digunakan dalam penelitian ini sebanyak 837 record.

3.2. Preprocessing

Pada tahap preprocessing dalam penelitian ini, dilakukan proses pembersihan data dengan menghilangkan beberapa komponen yang tidak diperlukan dalam perhitungan algoritma.

3.3. Transformation

Pada tahap transformasi hal yang dilakukan yaitu merubah data menjadi bentuk yang sesuai untuk dilakukan proses data mining. Tujuan menghilangkan atribut data yang tidak relevan adalah untuk memudahkan proses pengelompokan.

3.4. Data Mining

Proses data mining ini melibatkan penerapan algoritma atau pencarian pengetahuan, dan dalam penelitian ini, digunakan algoritma *K-means clustering* pada *tools rapidminer* versi 10.2.

3.5. Interpretation/Evaluation

Tahapan ini merupakan tahap akhir dari proses data mining. Tahap ini dilakukan setelah data diolah dengan algoritma yang dipilih menggunakan *tools rapidminer* dan tujuannya adalah untuk melakukan evaluasi terhadap hasil proses data mining.

4. HASIL DAN PEMBAHASAN

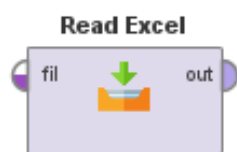
Dataset yang akan diolah dalam penelitian ini adalah data penerimaan peserta didik baru di SMKN 1 Balongan tahun ajaran 2023/2024 dengan jumlah 837 record. Data diperoleh melalui observasi, dan dokumentasi langsung di sekolah dengan tujuan mendapatkan informasi terkait variabel-variabel data penerimaan siswa baru. Variabel yang dicakup antara lain nama pendaftar, tahun ajaran, jenis kelamin, tempat lahir, tanggal lahir, bulan lahir, tahun lahir, jurusan, jumlah rombel, jalur pendaftar, dan asal sekolah. Selanjutnya dataset tersebut akan digunakan untuk proses analisis kluster dengan algoritma *K-Means Clustering* menggunakan aplikasi *RapidMiner* versi 10.2. Data PPDB yang akan diolah dapat dilihat pada Gambar 2 di bawah ini.

No	Nama Pendaftar	Jenis Kelamin	Tahun Ajaran	Tempat Lahir	Tanggal Lahir	Bulan Lahir	Tahun Lahir	Kejuruan	Jumlah Rombel	Jalur Pendaftaran	Asal Sekolah
1	Ahmad Henggo Prasetyo	Laki-laki	2023/2024	Subang	18	Januari	2007	TKJ	2	Pindah Ortu	Smp Negeri 1 Cusmen
2	Achwan Asnur Hakim	Laki-laki	2023/2024	Indramayu	20	Oktober	2007	TKJ	2	Persiapan Kelas Industri	Smp Negeri 1 Sityeg
3	Adella Balgis Zuliani	Perempuan	2023/2024	Indramayu	26	Februari	2008	TKJ	2	KETM	Smp muhammadiyah eksakurip
4	Adha Nur Sabrina	Perempuan	2023/2024	Indramayu	5	Mei	2006	TKJ	2	KETM	Mtsn 2 Indramayu
5	Agila Pujiastuti	Perempuan	2023/2024	Indramayu	23	Januari	2007	TKJ	2	KETM	Smp negeri 1 balongan
6	Agusta Noviana Putri	Perempuan	2023/2024	Indramayu	25	Mei	2007	TKJ	2	Persiapan Kelas Industri	Smp Negeri 1 Sityeg
7	Ahmad Munawar Kholid	Laki-laki	2023/2024	Indramayu	25	Agustus	2006	TKJ	2	Prioritas Jarak	Smp negeri 3 balongan
8	Aina Cahyaning Ratri	Perempuan	2023/2024	Indramayu	14	Agustus	2007	TKJ	2	Persiapan Kelas Industri	Smp Negeri 1 Sityeg
9	Aisha Vania Zaida Mustafa	Perempuan	2023/2024	Indramayu	3	Februari	2008	TKJ	2	Prioritas Jarak	Smp negeri 1 balongan
10	Aisy Kamila	Perempuan	2023/2024	Indramayu	17	Januari	2008	TKJ	2	Persiapan Kelas Industri	Smp negeri 1 balongan
...	...	...	...	...	...	...	...	...	...	...	...
836	Yuditha Andriani Hidayat	Laki-laki	2023/2024	Indramayu	28	Juli	2007	TKM	2	Prioritas Jarak	Smp muhammadiyah eksakurip
837	Zalfa Sababilla	Perempuan	2023/2024	Indramayu	3	Januari	2008	TKM	2	Persiapan Kelas Industri	Smp Negeri 1 Jundinyat

Gambar 2. Dataset PPDB

4.1. Selection

Langkah pertama adalah mencari lokasi file Itu sebelumnya dibuat dalam format xlsx dengan operator *Read Excel*. Fungsi dari Operator *Read Excel* ini adalah untuk membaca data yang tersimpan dalam format xlsx.



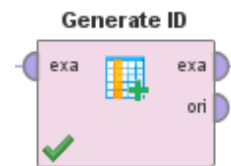
Gambar 3. Operator Read Excel

Parameter pada operator *Read Excel* menggubakan parameter default. Dari hasil pembacaan operator *Read Excel* didapat informasi sebagai berikut.

Tabel 1. Statistik Data PPDB

No	Uraian	Keterangan
1	Record	873
2	Special Attributes	0
3	Reguler Attributes	11
4	Attributes:	
	Nama Pendaftar	Polynomial, Missing 0
	Jenis Kelamin	Polynomial, Missing 0
	Tahun Ajaran	Polynomial, Missing 0
	Tempat Lahir	Polynomial, Missing 0
	Tanggal Lahir	Integer, Missing 0
	Bulan Lahir	Polynomial, Missing 0
	Tahun Lahir	Integer, Missing 0
	Kejuruan	Polynomial, Missing 0
	Jumlah Rombel	Integer, Missing 0
	Jalur Pendaftaran	Polynomial, Missing 0
	Asal Sekolah	Polynomial, Missing 0

Kemudian menggunakan operator *Generate ID*, Pemilihan *Generate ID* ini karena dataset dalam penelitian ini tidak memiliki atribut yang dapat dijadikan sebagai ID.



Gambar 4. Operator Generate ID

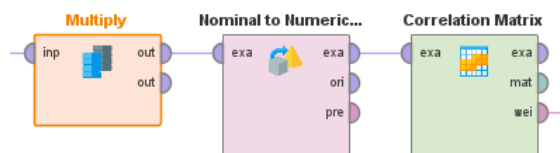
Untuk parameter yang digunakan pada operator ini adalah parameter *default*. Dari hasil pembacaan operator *Generate ID* didapat informasi yang dapat dilihat pada Tabel 2 dibawah ini.

Tabel 2. Hasil Operator *Generate ID*

No	Nama Kolom	Keterangan
1	Record	873
2	Special Attributes	1
3	Reguler Attributes	11
4	Attributes:	
	ID	Integer, Missing 0
	Nama Pendaftar	Polynomial, Missing 0
	Jenis Kelamin	Polynomial, Missing 0
	Tahun Ajaran	Polynomial, Missing 0
	Tempat Lahir	Polynomial, Missing 0
	Tanggal Lahir	Integer, Missing 0
	Bulan Lahir	Polynomial, Missing 0
	Tahun Lahir	Integer, Missing 0
	Kejuruan	Polynomial, Missing 0
	Jumlah Rombel	Integer, Missing 0
	Jalur Pendaftaran	Polynomial, Missing 0
	Asal Sekolah	Polynomial, Missing 0

Sebelum melakukan seleksi data dan menentukan data yang akan digunakan, serta menghapus data yang tidak diperlukan, langkah awal yang akan dilakukan

adalah proses penentuan atribut. Proses ini melibatkan operator multiply, nominal to numerical, dan correlation matrix. Tujuan dari langkah ini adalah untuk mengidentifikasi atribut-atribut yang tidak memiliki korelasi atau memiliki nilai korelasi yang sama dengan 0. Jika hasil dari correlation matrix terdapat atribut yang memiliki nilai korelasi 0, maka atribut tersebut akan dihapus pada tahap seleksi atribut berikutnya. Gambar 5 merupakan operator yang digunakan untuk membantu proses correlation matrix.



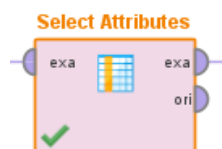
Gambar 5. Proses penentuan atribut dengan correlation Matrix

Adapun hasil dari proses penentuan atribut dengan menggunakan correlation matrix dapat dilihat pada Tabel 3 dibawah ini.

Tabel 3. Hasil Operator Correlation Matrix

No	Attribute	Weight
1	Tahun Ajaran	0
2	Jumlah Rombel	0
3	Nama Pendaftar	0.876
4	Kejuruan	0.879
5	Asal Sekolah	0.991
6	Jalur Pendaftaran	0.993
7	Tempat Lahir	0.998
8	Jenis Kelamin	0.999
9	Tahun Lahir	1.000
10	Bulan Lahir	1.000
11	Tanggal Lahir	1

Berdasarkan Tabel 3 diatas atribut tahun ajaran dan jumlah rombongan memiliki nilai 0 atau tidak memiliki korelasi yang artinya kedua atribut tersebut akan dihapus pada tahap seleksi atribut berikutnya. Select Attributes, operator ini digunakan untuk memisahkan data apa saja yang akan dipergunakan dalam pemrosesan data. Atribut dalam dataset Penerimaan Peserta Didik Baru di filter dari 12 menjadi 10 atribut yang digunakan.



Gambar 6. Operator Select Attributes

Parameter pada operator Select Attributes yang digunakan tampak pada Tabel 5 dibawah ini.

Tabel 4. Parameter dan atribut yang dipilih pada operator Select Attributes

No	Parameter	Isi
1	Type	Include attributes
2	Attribute filter type	a subset
3	Select attribute	ID, nama pendaftar, jenis kelamin, tempat lahir, tanggal lahir, bulan lahir, tahun lahir, kejuruan, jalur pendaftar, dan asal sekolah

Dari hasil pembacaan operator Select Attributes didapat informasi sebagai berikut.

Tabel 5. Statistik Dataset

No	Nama Kolom	Keterangan
1	Record	873
2	Special Attributes	1
3	Reguler Attributes	9
4	Attributes:	
	ID	Integer, Missing 0
	Nama Pendaftar	Polynomial, Missing 0
	Jenis Kelamin	Polynomial, Missing 0
	Tempat Lahir	Polynomial, Missing 0
	Tanggal Lahir	Integer, Missing 0
	Bulan Lahir	Polynomial, Missing 0
	Tahun Lahir	Integer, Missing 0
	Kejuruan	Polynomial, Missing 0
	Jalur Pendaftaran	Polynomial, Missing 0
	Asal Sekolah	Polynomial, Missing 0

4.2. Preprocessing

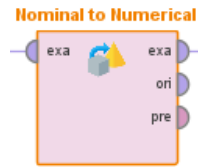
Pada tahapan preprocessing ini bertujuan untuk melakukan persiapan data sebelum masuk ke proses analisis lebih lanjut. Proses ini melibatkan serangkaian langkah untuk membersihkan, memformat, dan menyesuaikan data agar sesuai dengan kebutuhan analisis atau pemodelan yang akan dilakukan. Tujuan utama dari preprocessing data adalah meningkatkan kualitas data, menghilangkan noise atau gangguan, serta memastikan data siap untuk diolah dengan tepat. Dikarenakan pada dataset penerimaan peserta didik baru tidak ditemukan missing value atau data yang tidak memiliki nilai maka preprocessing tidak dilakukan. Adapun hasil statistik dari data PPDB dapat dilihat pada Gambar 7 dibawah ini.

No	Statistik	Detail
1	Nama Pendaftar	873 records, 1 missing value
2	Jenis Kelamin	873 records, 1 missing value
3	Tempat Lahir	873 records, 1 missing value
4	Tanggal Lahir	873 records, 1 missing value
5	Bulan Lahir	873 records, 1 missing value
6	Tahun Lahir	873 records, 1 missing value
7	Kejuruan	873 records, 1 missing value
8	Jalur Pendaftaran	873 records, 1 missing value
9	Asal Sekolah	873 records, 1 missing value

Gambar 7. Hasil Statistik

4.3. Transformation

Nominal to Numerical, digunakan untuk mengubah tipe atribut yang bersifat non-numerik menjadi tipe numerik. Dalam dataset Penerimaan peserta didik baru atribut asal sekolah, bulan lahir, jalur pendaftaran, jenis kelamin, kejuruan, nama pendaftaran, tahun ajaran, tempat lahir.



Gambar 8. Operator *Nominal to Numerical*

Tabel 6. Parameter pada Operator *Nominal to Numerical*

No	Parameter	Isi
1	Attribute Filter Type	Subset
2	Attributes	Asal sekolah, bulan lahir, jalur pendaftaran, jenis kelamin, kejuruan, nama pendaftaran, tahun ajaran, tempat lahir.
3	Coding Type	Unique Integers

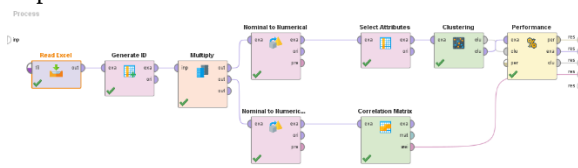
Dari hasil pembacaan operator *Nominal to Numerical* didapat informasi sebagai berikut.

Tabel 7. Hasil Pembacaan Operator *Nominal to Numerical*

No	Nama Kolom	Keterangan
1	Record	873
2	Special Attributes	1
3	Reguler Attributes	9
4	Attributes:	
	ID	Integer, Missing 0
	Nama Pendaftar	Numeric, Missing 0
	Jenis Kelamin	Numeric, Missing 0
	Tempat Lahir	Numeric, Missing 0
	Tanggal Lahir	Integer, Missing 0
	Bulan Lahir	Numeric, Missing 0
	Tahun Lahir	Integer, Missing 0
	Kejuruan	Numeric, Missing 0
	Jalur Pendaftaran	Numeric, Missing 0
	Asal Sekolah	Numeric, Missing 0

4.4. Data Mining

Berikut adalah model proses *Clustering* dengan menggunakan Algoritma K-Means pada data penerimaan peserta didik baru menggunakan tool RapidMiner versi 10.2. :



Gambar 9. Model Clustering Menggunakan Algoritma K-Means

Gambar 10 merupakan operator utama Clustering K-Means.



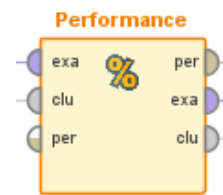
Gambar 10. *K-Means Clustering*

Parameter pada operator *K-Means Clustering* yang digunakan tampak pada tabel dibawah ini. Parameter pada operator *K-Means Clustering* akan dilakukan eksperimen dengan mengubah nilai K dari 2 sampai 10, *measure type* diubah menjadi *Numerical Measure* dengan jenis *Chebychev distance* dan *Correlation Similarity*.

Tabel 8. Parameter pada Operator *K-Means Clustering*

No	Parameter	Isi
1	K	2-10
2	Numerical Measure	<i>ChebychevDistance</i> dan <i>Correlation Similarity</i>

Proses selanjutnya yaitu menambahkan Operator *Performance* dengan *Main Criterion* Davies-Bouldin yang bertujuan untuk mengetahui nilai DBI nya.



Gambar 11. Operator *Performance*

Operator *Performance* digunakan untuk mengevaluasi kinerja model atau proses dalam analisis data. Dalam konteks K-Means clustering, operator ini memberikan informasi tentang seberapa baik model K-Means membentuk kelompok atau kluster yang sesuai dengan pola data sebenarnya. Parameter pada operator *Performance* yang digunakan tampak pada Tabel 10 dibawah ini.

Tabel 9. Parameter pada Operator *Performance*

No	Parameter	Isi
1	<i>Main Criterion</i>	<i>Devies Bouldin</i>

4.5. Hasil Eksperimen Kluster (*Numerical Measures Type - ChebyChev Distance*)

Hasil pengelompokan didapatkan melalui sembilan uji coba yang dilakukan dengan menggunakan nilai K sebanyak 2, 3, 4, 5, 6, 7, 8, dan 9 untuk *Numerical Measures Type = Chebychev Distance*. Tahap berikutnya adalah menghitung nilai DBI untuk setiap pengelompokan percobaan tersebut. Hasil dari nilai DBI sebagaimana terlihat dalam tabel 3 di bawah ini.

Tabel 10. Kluster yang dihasilkan
(Numerical Measures Type - Euclidean Distance)

Numerical Measures	K	DBI
ChebyChev Distance	2	0.508
	3	0.509
	4	0.518
	5	0.526
	6	0.535
	7	0.544
	8	0.555
	9	0.567
	10	0.582

4.6. Hasil Eksperimen Kluster (Numerical Measures Type – Correlation Similarity)

Hasil pengelompokan didapatkan melalui lima uji coba yang dilakukan dengan menggunakan nilai K sebanyak 2, 3, 4, 5, 6, 7, 8, dan 9 untuk *Numerical Measures Type = Correlation Similarity*. Tahap berikutnya adalah menghitung nilai DBI untuk setiap pengelompokan percobaan tersebut. Hasil dari nilai DBI sebagaimana terlihat dalam tabel 11 di bawah ini.

Tabel 11. Kluster yang dihasilkan
(Numerical Measures Type – Correlation Similarity)

Numerical Measures	K	DBI
Correlation Similarity	2	0.508
	3	0.509
	4	0.518
	5	0.527
	6	0.536
	7	0.546
	8	0.556
	9	0.568
	10	0.582

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 14533.172
Avg. within centroid distance_cluster_0: 14391.917
Avg. within centroid distance_cluster_1: 14672.749
Davies Bouldin: 0.508
```

Gambar 12. Hasil *Performance Vector*

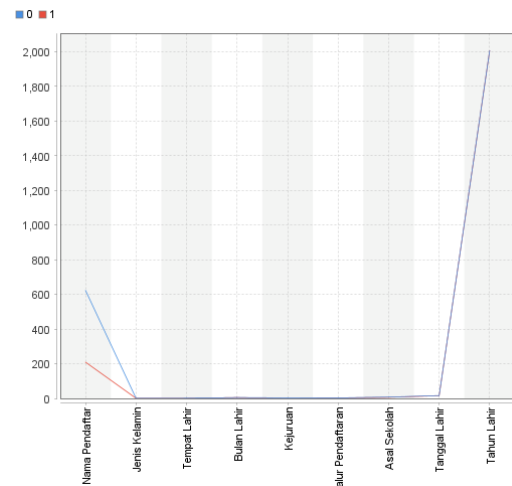
4.7. Interpretation/Evaluation

Dari hasil penelitian dengan Cluster Distance Performance berdasarkan Davies-Bouldin Index (DBI) menggunakan Numerical Measure Chebychev Distance dan Correlation Similarity, ditemukan bahwa nilai optimal K adalah 2. Cluster dengan K=2 dipilih karena memiliki nilai DBI yang paling rendah atau mendekati nol. Pengukuran nilai max runs dari 1 sampai 10 dengan numerical measure Chebyshev Distance dan Correlation Similarity menunjukkan konsistensi nilai DBI, yaitu sebesar 0.508. Terdapat 416 item dalam cluster 0 dan 421 item dalam cluster.

Tabel 12. Hasil Clustering K=2 yang merupakan nilai DBI terbaik

No	Cluster	Keterangan
1	Cluster 0	416 Items
2	Cluster 1	421 Items
	Total Number of Items	837 Items

Berikut adalah hasil grafik dari setiap atribut yang divisualisasikan dalam plot menggunakan RapidMiner 10.2.



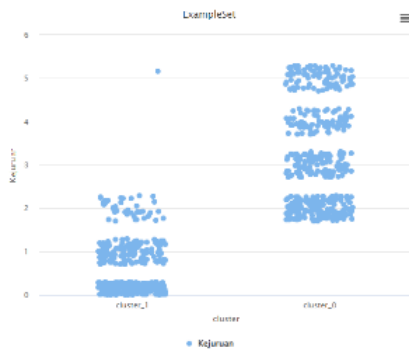
Gambar 13. Hasil grafik pada atribut data PPDB

Dari Gambar 13 di atas, dapat disimpulkan bahwa hasil pengelompokan data penerimaan peserta didik baru menggunakan Algoritma K-Means didasarkan pada atribut jenis kelamin, tempat lahir, bulan lahir, kejuruan, jalur pendaftaran, asal sekolah, tanggal lahir, dan tahun lahir pada kedua kluster memberikan hasil yang serupa, serta memiliki Evaluasi dan nilai Support tertinggi dengan akurasi sebesar 2,000. Sementara itu, atribut Nama Pendaftar menghasilkan nilai akurasi yang berbeda, di mana kluster 0 memiliki akurasi sebesar 600 dan kluster 1 memiliki akurasi sebesar 200.

Berikut adalah hasil analisis karakteristik dari setiap cluster pada atribut kejuruan dan asal sekolah yang dapat digunakan sebagai informasi sebagai dasar strategi promosi untuk penerimaan peserta didik baru di tahun ajaran berikutnya, yang di visualisasikan pada plot type Scatter / Bubble dapat dilihat pada gambar dibawah ini.

4.5.1. Cluster terhadap Jalur Pendaftaran

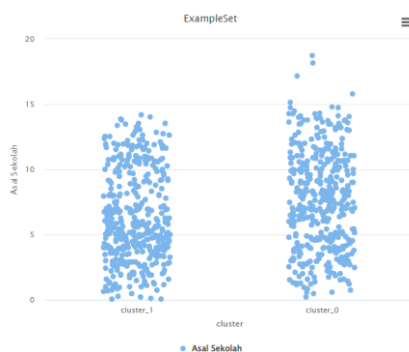
Untuk Atribut kejuruan pada cluster_1 didominasi oleh jenis kejuruan TKJ, DKV, TKRO, dan TKM. Sedangkan pada cluster_0 didominasi oleh jenis kejuruan TKRO, TITL, DPIB, TKM dan tidak mencakup jenis kejuruan TKJ dan DKV.



Gambar 14. Titik Persebaran Atribut Kejuruan

4.5.2. Cluster Terhadap Atribut Asal Sekolah

Untuk atribut asal sekolah pada cluster_1 didominasi oleh jenis asal sekolah seperti Smp Negeri 1 Ciasem, Smp Negeri 1 Sliyeg, Smp Muhammadiyah Sukaurip, Mtsn 2 Indramayu, Smp Negeri 1 Balongan, Smp Negeri 3 Balongan, Smp Negeri 1 Karangampel, Smp Negeri 2 Balongan, Smp Negeri 1 Juntinyuat, Smp Negeri 1 Indramayu, Smp Pgri Juntinyuat, Smp Negeri 2 Juntinyuat, Smp Plus Nurul Islam, dan Smp Negeri 3 Sindang. Sedangkan pada cluster_0 mencakup hampir semua jenis asal sekolah diantaranya seperti Smp Negeri 1 Sliyeg, Smp Muhammadiyah Sukaurip, Mtsn 2 Indramayu, Smp Negeri 1 Balongan, Smp Negeri 3 Balongan, Smp Negeri 1 Karangampel, Smp Negeri 2 Balongan, Smp Negeri 1 Juntinyuat, Smp Negeri 1 Indramayu, Smp Pgri Juntinyuat, Smp Negeri 2 Juntinyuat, Smp Plus Nurul Islam, Smp Negeri 3 Sindang, Smp Negeri 10 Kota Bandung, Mts Muhammadiyah Baruamba, Smp Negeri 1 Ciawigebang, Smp Negeri 2 Cimalaka, Smp Negeri 1 Margasari, terkecuali jenis asal sekolah dari Smp Negeri 1 Ciasem.



Gambar 15. Titik Persebaran Atribut Asal Sekolah

5. KESIMPULAN DAN SARAN

Dari hasil eksperimen dengan menggunakan measure type numerical measure chebychev distance dan correlation similarity menghasilkan kelompok terbaik $k=2$ dengan nilai DBI 0.508, dengan jumlah cluster 0 sebanyak 421 items sedangkan jumlah Cluster 1 sebanyak 416. Untuk Cluster_0 pada atribut kejuruan di dominasi oleh jenis kejuruan TKRO, TITL, DPIB, TKM. Dan untuk atribut asal sekolah pada Cluster_0 di dominasi oleh jenis asal sekolah

diantaranya seperti Smp Negeri 1 Sliyeg, Smp Muhammadiyah Sukaurip, Mtsn 2 Indramayu, Smp Negeri 1 Balongan, Smp Negeri 3 Balongan, Smp Negeri 1 Karangampel, Smp Negeri 2 Balongan, Smp Negeri 1 Juntinyuat, Smp Negeri 1 Indramayu, Smp Pgri Juntinyuat, Smp Negeri 2 Juntinyuat, Smp Plus Nurul Islam, Smp Negeri 3 Sindang, Smp Negeri 10 Kota Bandung, Mts Muhammadiyah Baruamba, Smp Negeri 1 Ciawigebang, Smp Negeri 2 Cimalaka, Smp Negeri 1 Margasari dan cluster 1 berjumlah 300 item didominasi atribut... Hasil penelitian ini bisa menjadi sebuah informasi yang dapat digunakan oleh SMKN 1 Balongan untuk melakukan strategi promosi yang lebih efektif dan mampu meningkatkan minat siswa untuk mendaftar di SMKN 1 Balongan. Untuk Cluster 1 pada atribut kejuruan di dominasi oleh jenis kejuruan TKJ, DKV, TKRO, dan TKM. Dan untuk asal sekolah pada Cluster_1 di dominasi oleh jenis asal sekolah seperti Smp Negeri 1 Ciasem, Smp Negeri 1 Sliyeg, Smp Muhammadiyah Sukaurip, Mtsn 2 Indramayu, Smp Negeri 1 Balongan, Smp Negeri 3 Balongan, Smp Negeri 1 Karangampel, Smp Negeri 2 Balongan, Smp Negeri 1 Juntinyuat, Smp Negeri 1 Indramayu, Smp Pgri Juntinyuat, Smp Negeri 2 Juntinyuat, Smp Plus Nurul Islam, dan Smp Negeri 3 Sindang.

Dari hasil penelitian ini, beberapa saran strategis dapat diajukan untuk meningkatkan pemahaman dan pengembangan sebagai berikut : Mengacu pada hasil pengolahan data PPDB menggunakan pendekatan clustering algoritma k-means, ditemukan pola-pola data yang dapat dijadikan informasi untuk menentukan strategi promosi PPDB ke depannya. Secara spesifik, promosi dapat difokuskan pada setiap jurusan dan sekolah asal yang paling diminati di setiap klaster yang terbentuk. Dengan demikian, promosi PPDB dapat dilakukan lebih tepat sasaran. Pihak sekolah juga dapat mempromosikan jurusan-jurusan dengan peminat yang relatif sedikit kepada sekolah asal dengan jumlah pendaftar yang tinggi. Sebagai contoh, promosi program kejuruan TITL, DPIB dan sejenisnya yang kurang diminati dapat diprioritaskan ke SMP-SMP favorit atau SMP dengan jumlah pendaftar tinggi di wilayah tersebut. Sementara itu, upaya promosi perlu ditingkatkan ke sekolah asal yang memiliki sedikit pendaftar ke SMKN 1 Balongan, guna meningkatkan minat yang lebih merata dari beragam latar belakang sekolah. Untuk penelitian selanjutnya dapat mengeksplorasi lebih lanjut dalam pemilihan metode pengukuran numerik dan nilai K yang optimal, serta menguji efektivitasnya pada berbagai jenis dataset. Penelitian juga dapat menelusuri potensi penggunaan algoritma clustering dalam konteks aplikasi praktis atau industri tertentu. Fokus pada interpretabilitas hasil clustering dan pemahaman lebih mendalam terhadap konteks domain spesifik diharapkan dapat memperkaya kontribusi penelitian. Terakhir, penelitian berikutnya bisa mempertimbangkan pengembangan model prediktif atau integrasi dengan teknik pembelajaran mesin lainnya untuk memberikan

solusi yang lebih menyeluruh dan relevan terhadap tantangan praktis di bidang pengelompokan data.

DAFTAR PUSTAKA

- [1] M. R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," *J. Softw. Eng. Ampera*, vol. 1, no. 1, pp. 1–14, 2020, doi: 10.51519/journalsea.v1i1.10.
- [2] A. Maritsa, U. Hanifah Salsabila, M. Wafiq, P. Rahma Anindya, and M. Azhar Ma'shum, "Pengaruh Teknologi Dalam Dunia Pendidikan," *Al-Mutharahah J. Penelit. dan Kaji. Sos. Keagamaan*, vol. 18, no. 2, pp. 91–100, 2021, doi: 10.46781/al-mutharahah.v18i2.303.
- [3] N. A. Rahmalinda and A. Jananto, "Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru," *J. Tekno Kompak*, vol. 16, no. 2, pp. 163–175, 2022.
- [4] M. Solikhin and Naimah, "Praxis Blue Ocean Strategic: (Student Selection)Ppdb Process in Educational Institutions During the Covid-19 Pandemic," *J. Manaj. Pendidik. Islam*, vol. 6, no. 2, pp. 286–297, 2019, [Online]. Available: <http://e-journal.ikhac.ac.id/index.php/nidhomulhaq>
- [5] N. S. PERDANA, "Implementasi Ppdb Zonasi Dalam Upaya Pemerataan Akses Dan Mutu Pendidikan," *J. Pendidik. Glas.*, vol. 3, no. 1, p. 78, 2019, doi: 10.32529/glasser.v3i1.186.
- [6] A. Widyaningtyas, N. Nugraha, and S. Diana, "Antropocene : Jurnal Penelitian Ilmu Humaniora Persepsi Masyarakat Terhadap Sistem Penerimaan Peserta Didik Baru," vol. 1, no. 1, pp. 29–37, 2021.
- [7] D. Udariansyah and D. Rahmat Ibrahim, "Klasifikasi Data Penerimaan Mahasiswa Baru Pada Universitas Bina Darma Menggunakan Algoritma K-Means Clustering," *J. Pendidik. dan Konseling*, vol. 4, no. 4, pp. 2692–2701, 2022, [Online]. Available: <https://journal.universitaspahlawan.ac.id/index.php/jpdk/article/view/5879>
- [8] 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.
- [9] A. Agneresa, A. L. Hananto, S. S. Hilabi, A. Hananto, and T. Tukino, "Strategi Promosi Penerapan Data Mining Mahasiswa Baru Dengan Metode K-Means Clustering," *Dirgamaya J. Manaj. dan Sist. Inf.*, vol. 2, no. 2, pp. 25–34, 2022, doi: 10.35969/dirgamaya.v2i2.275.
- [10] A. L. Maukar, F. Marisa, and A. A. Widodo, "Analisis Data Penerimaan Mahasiswa Baru Berbasis K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, p. 142, 2022, doi: 10.26798/jiko.v6i2.558.
- [11] I. F. Helmi, "JUSTIFY : Jurnal Sistem Informasi Ibrahimy Analisis dan Penerapan Algoritma K-Means Clustering Sebagai Strategi Promosi Penerimaan Mahasiswa Baru Pada Universitas Wiraraja," vol. 2, no. 1, pp. 1–11, 2023, doi: 10.35316/justify.v2i1.3205.
- [12] W. Rosida and Y. A. Wijaya, "Klasterisasi Penyakit HIV/AIDS di Jawa Barat Menggunakan Algoritma K-Means Clustering," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 306–315, 2023, doi: 10.56211/blendsains.v1i4.235.
- [13] A. Setiadi and E. D. Sikumbang, "K-Means Clustering Dalam Penerimaan Karyawan Baru," *INFORMATICS Educ. Prof. J. Informatics*, vol. 4, no. 2, p. 103, 2020, doi: 10.51211/itbi.v4i2.1304.
- [14] A. Permadi and Y. A. Wijaya, "Pengelompokan Terbaik Menggunakan Algoritma K-Means Pada Dataset Bus Biskita Bogor," *Intern. (Information Syst. J.)*, vol. 6, no. 1, pp. 88–100, 2023, [Online]. Available: <http://jurnal.masoemuniveristy.ac.id/index.php/interanal>
- [15] Raoda, R. Retnawati, Penelitian, H. Ismail, O. R. Imawan, and U. Cendrawasih, "Analisis Klaster Ujian Nasional : Pengelompokan sekolah untuk menjaga keberlangsungan mutu sekolah menengah pada tingkat SD , SMP , SMA , SMK , dan sederajat . UN bertujuan untuk mengukur dan menilai mengacu pada Standar Kompetensi Lulusan . Selain itu ,," vol. 8, no. 1, pp. 13–23, 2022.
- [16] J. Malau, M. A. Mulki, N. K. Fitri, N. Yusiana, and N. K. Hanum, "p-ISSN: <https://doi.org/10.5281/zenodo.7571470> Accredited by Directorate General of Strengthening for Research and Development Available online at <https://jurnal.peneliti.net/index.php/JIWP>," vol. 9, no. 2, pp. 259–269, 2023.
- [17] S. Isnanto and S. Widodo, "Penerapan Data Mining Pada Penerimaan Mahasiswa Baru Dengan Algoritma K-Means Clustering," *J. Tek. Inf. dan Komput.*, vol. 4, no. 2, p. 158, 2021, doi: 10.37600/tekinkom.v4i2.367.