PENERAPAN ALGORITMA *K-MEANS* UNTUK PENENTUAN STRATEGI PROMOSI (Studi Kasus: SMA PGRI 1 PURWAKARTA)

Karina Listiani Putri, Muhammad Agus Sunandar, Yusuf Muhyidin
Program Studi Teknik Informatika S1, STT Wastukancana Purwakarta
Jl Cikopak Sadang No 53 Purwakarta, Indonesia
karinalistiani61@wastukancana.ac.id

ABSTRAK

Data yang dikumpulkan setiap tahun tentang jumlah siswa baru belum sepenuhnya digunakan untuk mengevaluasi dan membuat strategi pemasaran yang efektif. Tidak dapat diabaikan betapa pentingnya membuat strategi promosi yang tepat dengan data dari masa lalu. Saat ini, pengolahan data menjadi hal yang sangat krusial karena dapat menghasilkan pengetahuan baru menjadi dasar dalam pengambilan keputusan institusi terkait strategi pemasaran penerimaan calon siswa baru. Data *mining* juga memegang peranan penting bagi perusahaan, baik perguruan tinggi maupun institusi, untuk mengungkap pola-pola tersembunyi dalam database yang dimiliki. Untuk menerapkan metode data *mining clustering*, penelitian ini akan menggunakan algoritma *K-Means*. Data yang digunakan dalam penelitian ini merupakan data penerimaan siswa pada tahun 2021-2022 dengan jumlah 30 siswa dan menggunakan 3 klaster. Hasil dari pengelompokan menggunakan algoritma *K-Means* dengan menggunakan *tool Rapidminer* versi 9.10 adalah mencari nilai K terbaik dari K2 hingga K30 berdasarkan performanya yang mendekati angka 0. Hasilnya menunjukkan bahwa K2 memiliki performa terbaik dengan nilai 0,648 pada iterasi pertama dari total iterasi 1 hingga 30. Performa algoritma *K-Means* dapat dilihat dari nilai *davies bouldin*, dimana semakin mendekati angka 0 nilai berarti algoritma semakin baik. Oleh karena itu, jumlah *cluster* terbaik dalam penelitian ini adalah 2, karena memiliki nilai *davies bouldin* yang mendekati 0.

Kata kunci: *Algoritma K-Means*, Strategi Promosi, *Rapid miner*

1. PENDAHULUAN

Salah satu tantangan yang dihadapi dalam era globalisasi di berbagai aspek kehidupan adalah adopsi teknologi informasi dan komunikasi. Teknologi informasi merupakan elemen yang tak terpisahkan dari dunia bisnis, terutama untuk menghadapi persaingan bisnis yang semakin ketat. Kehadiran teknologi informasi menjadi kebutuhan mendasar bagi perusahaan agar dapat bertahan dalam lingkungan bisnis yang penuh dengan kompetisi. Lebih lanjut, teknologi informasi telah mendorong perkembangan teknologi produk dan proses, serta mempengaruhi pembentukan masyarakat informasi [1].

SMA PGRI 1 Purwakarta berperan sebagai penyedia layanan pendidikan dan berusaha untuk memberikan layanan terbaik kepada siswa guna meningkatkan kepuasan terhadap pelayanan pendidikan. Sebagai institusi pendidikan tingkat menengah, SMA ini memiliki jumlah data akademik dan administrasi yang besar, namun hanya sebagian kecil dari data tersebut yang digunakan dalam proses penyusunan evaluasi diri. Data akademik siswa merupakan informasi yang dikumpulkan dari berbagai kegiatan.

Pada awal tahun ajaran baru di SMA PGRI 1 Purwakarta, selalu dilakukan proses penerimaan siswa baru yang menghasilkan banyak data. Namun, pendataan dan pemanfaatan data ini untuk keperluan evaluasi dan strategi pemasaran belum dilakukan secara optimal. Hingga saat ini, hanya dilakukan analisis sederhana dengan melihat data dari tahun sebelumnya untuk mengevaluasi dan memetakan profil siswa baru yang diterima. Proses ini dilakukan untuk sekadar mengetahui sebaran siswa baru yang berhasil diterima. Proses penerimaan siswa baru di Sekolah Menengah Atas akan terus berulang setiap tahunnya, menghasilkan data baru berupa profil siswa baru. Dengan pengolahan data yang tepat, berbagai informasi berharga dapat diperoleh dan digunakan untuk memberikan kontribusi dalam menentukan strategi promosi untuk penerimaan siswa baru pada tahun-tahun mendatang.

Data *mining* akan digunakan untuk mengolah data menggunakan metode Algoritma *K-Means*. Hasil dari pengolahan data ini diharapkan akan sangat membantu pihak sekolah. Dengan penelitian ini, diharapkan akan dapat diketahui sebaran siswa dari data penerimaan setiap tahun, sehingga akan diketahui wilayah mana yang menjadi prioritas untuk promosi ke depan berdasarkan jumlah siswa yang berasal dari wilayah tersebut. Atribut yang akan dimanfaatkan dalam pengolahan data ini mencakup kota asal, asal sekolah, jurusan, nilai matematika, nilai inggris, dan Nilai UN. Data kemudian akan dikelompokkan berdasarkan atribut tersebut menggunakan Algoritma *K-Means*.

2. TINJAUAN PUSTAKA

2.1. Data

Webster's New World's Dictionary menyatakan bahwa datum merujuk pada sesuatu yang diketahui atau diasumsikan. Dalam konteks ini, data

menggambarkan informasi tentang suatu keadaan atau permasalahan. *Oxford Dictionary* mendefinisikan data sebagai fakta-fakta. Oleh karena itu, data dapat didefinisikan sebagai informasi nyata yang diketahui atau diasumsikan yang digunakan untuk analisis, diskusi, uji statistik, atau presentasi ilmiah [2].

2.2. Data Mining

Proses data mining melibatkan pencarian pola atau informasi menarik dalam data yang telah dipilih menggunakan teknik atau metode tertentu. Ada berbagai macam teknik, metode, atau algoritma dalam data mining yang dapat digunakan. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan seluruh proses Knowledge Discovery in Database (KDD) [3].

2.3. Tujuan Data Minning

Berikut ini adalah beberapa tujuan data mining dilakukan [4]

1. Tujuan penjelasan (*Explanatory*): Data mining berperan sebagai alat untuk mengungkapkan atau menjelaskan kondisi penelitian atau fenomena tertentu.
2. Tujuan konfirmasi (*Confirmatory*): Data mining berfungsi untuk mengonfirmasi atau memverifikasi hipotesis atau pernyataan yang telah dibuat sebelumnya.
3. Tujuan eksplorasi (*Exploratory*): Data mining digunakan sebagai alat untuk melakukan eksplorasi guna menemukan pola-pola baru yang sebelumnya belum terdeteksi atau dipahami.

2.4. Pengelompokan Data Mining

Clustering dalam data mining Menurut Larose, data mining dikelompokkan berdasarkan tugas-tugas yang dapat dilakukan, yaitu [5]

1. Deskripsi dari data mining adalah untuk mencari metode yang dapat menggambarkan pola dan kecenderungan yang terdapat dalam data. Hasil dari proses deskripsi ini sering memberikan penjelasan tentang pola atau kecenderungan yang ada dalam data tersebut.
2. Estimasi hampir mirip dengan klasifikasi, namun variabel target pada estimasi cenderung bersifat numerik daripada kategorikal.
3. Prediksi hampir serupa dengan klasifikasi dan estimasi, tetapi fokusnya adalah pada nilai hasil yang akan terjadi di masa mendatang.
4. Klasifikasi terdapat variabel target berbentuk kategori. Sebagai contoh, penggolongan pendapatan dapat dibagi menjadi tiga kategori, yaitu pendapatan tinggi, sedang, dan rendah.
5. Pengklusteran adalah untuk membagi data secara keseluruhan menjadi kelompok yang homogen, di mana kesamaan antara data dalam satu kelompok mencapai tingkat tertinggi dan tingkat terkecil.
6. Data mining *associations* bekerja untuk menemukan fitur atau item yang sering muncul bersama dalam suatu waktu. Analisis keranjang belanja atau analisis asosiasi belanja adalah nama lain untuk pekerjaan ini di dunia bisnis.

2.5. Penerapan Data Mining

Berikut adalah beberapa sektor industri di mana data mining digunakan:

1. Bisnis
   Sektor bisnis menggunakan data mining untuk pemasaran, analisis pasar, dan analisis kebutuhan pelanggan
2. Pendidikan
   Sektor pendidikan menggunakan data mining untuk memahami karakteristik setiap siswa dan menemukan cara terbaik untuk menerapkan pelajaran.
3. Asuransi
   Data mining digunakan untuk menemukan penipuan dan risiko pengajuan klaim asuransi serta untuk memahami minat dan kebutuhan pelanggan.
4. Perbankan
   Sektor perbankan menggunakan data mining untuk memprediksi seberapa besar kemungkinan klien tidak dapat membayar pinjaman mereka, dengan tujuan mengurangi risiko kerugian.

2.6. Promosi dan Stategi

Menurut [6], *promotion, the fourth marketing mix tools, stand for various activities, the company undertakes to communicate its products merits and topersuade target customers to buy them*. Definisi tersebut bermakna bahwa promosi mencakup semua alat yang ada dalam bauran promosi, dengan peran utamanya adalah melakukan komunikasi yang bersifat membujuk.

2.7. Algoritma K-Means

Menurut [7] Algoritma *K-means* merupakan salah satu bidang penelitian yang berkaitan dengan data mining. Algoritma ini membagi data secara keseluruhan menjadi kelompok atau cluster yang memiliki kemiripan yang sama, meskipun pendekatan pengelompokannya bergantung pada kemiripan data yang tidak diawasi.

K-Means adalah algoritma clustering yang berulang-ulang. Algoritma ini dimulai dengan memilih secara acak nilai K, di mana K menunjukkan jumlah kluster yang ingin dibentuk. [8].

Langkah-langkah algoritma *K-Means* adalah sebagai berikut:

Persiapkan data set.

Tentukan jumlah *klister* (*K*).

Pilih titik *centroid*. Adapun untuk menentukan titik *centroid* ada beberapa pilihan diantaranya mengambil nilai tertinggi dan nilai terendah (*max*;*min*), berurutan dan secara acak.

Kelompokan data sehingga terbentuk *K* buah klister dengan titik *centroid* dari setiap *klister* dengan menggunakan metode *Euclidean distance*.

$$d(x,y)\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Dimana,

$d_{(x,y)}$ = jarak antara x ke data y

xi = data *testing* ke-i

yi = data *training* ke-i

perbarui nilai titik *centroid* dengan sebagai berikut :

$$\mu_k = \frac{1}{N_k}\sum_{i=1}^{N_k} x_i$$

Dimana,

μ_k = titik *centroid* dari klaster ke-*k*

N_k = banyaknya data pada *klaster* ke-*k*

x_i = data ke-I pada *klaster* ke-*k*

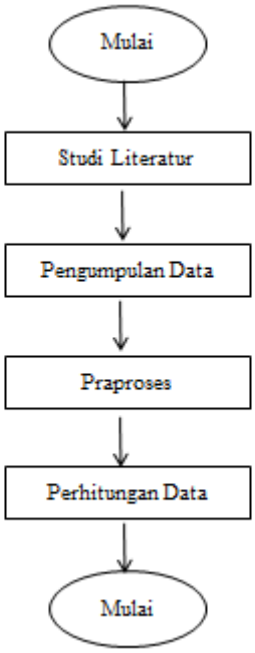
ulangi Langkah 3 sampai 5 sampai nilai *centroid* tidak lagi berubah [9].

3. METODE PENELITIAN

3.1. Kerangka Penelitian

Pada penelitian ini, digunakan pendekatan deskriptif kuantitatif. Artinya, penelitian deskriptif ini dilakukan dengan mencari informasi terkait gejala yang ada, merencanakan pendekatannya secara jelas dengan tujuan yang akan dicapai, dan mengumpulkan berbagai macam data sebagai dasar untuk membuat laporan. Selain itu, penelitian ini menggunakan pendekatan kuantitatif karena melibatkan penggunaan angka dalam semua tahapan, mulai dari pengumpulan data, penafsiran terhadap data, hingga penyajian hasilnya. [10]

Tahapan detail tentang K-Means dapat dilihat pada gambar dan pembahasan di bawah:



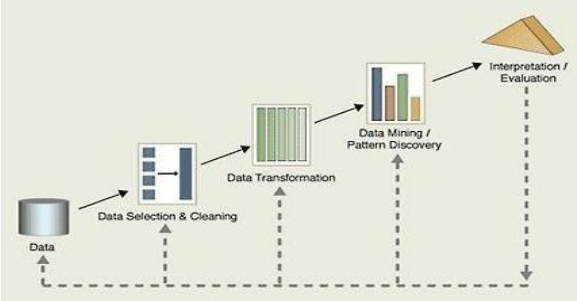
Gambar 1. Kerangka Penelitian

Centroid adalah rata-rata aritmatika bentuk objek dari setiap titik dalam objek. Penerapan klasterisasi *K-Means* dapat dilakukan dengan prosedur langkah demi langkah berikut.

1. Menyiapkan data training berbentuk vector
2. Siapkan nilai K *cluster*
3. Set nilai awal *centroids*.
4. Gunakan rumus jarak Euclidean untuk menghitungjarak antara data dan centroid
5. Partisi data berdasarkan nilai minimum
6. Kemudian ulangi saat partisi data masih bergerak(tidak perlu memindahkan objek ke partisi lain).
   Jika Anda masih belum, lanjutkan ke poin 3
7. Jika dataset saat ini sama dengan dataset sebelumnya, hentikan iterasi.
8. Data dibagi sesuai dengan nilai centroid akhir.

3.2. Tahapan Penelitian

Proses analisis data dalam penerapan data mining dilakukan melalui langkah-langkah knowledge discovery in databases (KDD) yang terdiri dari Data, *Data Cleaning, Data transformation, Data mining, Pattern evolution, knowledge:*



Gambar 2. Tahapan KDD
Sumber: Ahmad Haidar Mirza

Dibawah ini adalah pembahasan tentang tahap *knowledge discovery in databases (KDD)*

1. Data Selection
   Data PPDB pada SMA PGRI 1 Purwakarta diambil pada tahun pelajaran 2020/2021 Data yang digunakan dalam penelitian ini sebanyak 100 record
2. Data Cleaning
   Data yang diperoleh, baik dari database PPDB SMA PGRI 1 Purwakarta maupun dari survey, umumnya mengandung isian-isian yang tidak lengkap seperti data yang hilang, tidak valid, atau adanya kesalahan ketik.
3. Data integration
   Integrasi data dilakukan pada atribut yang mengidentifikasi entitas unik SMA PGRI 1 Purwakarta dengan 100 atribut Integrasi data perlu dilakukan dengan cermat, karena kesalahan dalam integrasi data dapat menyebabkan hasil yang bias dan

mempengaruhi keputusan yang diambil berikutnya.

4. Data transformation
   Merupakan proses transformasi dari data terpilih agar sesuai dengan persyaratan data mining.
5. Data *Mining*
   Data mining merupakan proses penemuan dan analisis sejumlah besar data dengan tujuan menemukan pola atau informasi yang menarik dari data yang disimpan. Proses ini dilakukan dengan menggunakan metode atau teknik tertentu. Proses pencarian pengetahuan dalam database KDD secara keseluruhan dan tujuan dari tahap ini sangat bergantung pada pemilihan metode, teknik, atau algoritma yang tepat. Tahap ini merupakan bagian penting dari proses KDD, yang dilakukan setelah data telah dibersihkan.
6. Pattern *evaluation*
   Pada langkah ini, hasil dari teknik data mining berupa pola dan model dievaluasi untuk menentukan apakah benar-benar dapat dicapai. Jika hasil yang diterima tidak konsisten, Beberapa opsi dapat diimplementasikan, seperti memberikan umpan balik untuk meningkatkan proses data mining lain yang lebih sesuai, atau menerima hasil yang tidak sesuai dengan harapan sebagai bahan pembelajaran.

4. HASIL DAN PEMBAHASAN

4.1. Pemahaman Bisnis

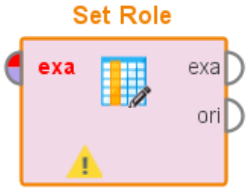
SMA PGRI 1 Purwakarta sekolah yang didirikan oleh yayasan. Pengelolaan sekolah tersebut sepenuhnya ada yayasan. Untuk mendapatkan murid, Setiap sekolah pasti akan melakukan penerimaan siswa baru setiap tahunnya sebagai cara untuk mendapatkan murid. Penerimaan siswa baru menjadi hal yang sangat penting bagi sekolah karena ini menjadi titik awal untuk kelancaran tugas dan proses belajar mengajar di sekolah tersebut.

4.2. Seleksi Data

Pada langkah ini, peneliti menggunakan data profil siswa untuk memilih atribut yang relevan untuk penelitian. Kota Asal, Sekolah Asal, dan Rata-Rata Nilai Rapor adalah lima atribut yang dipilih untuk penelitian, seperti yang terlihat di Tabel 4.1, yang menunjukkan data siswa yang digunakan untuk penelitian ini.

Gambar 3. Operator Read Excel

Koordinat & prediksi posisi yang akan pada masukan kedalam kategori. Agar ketika pengkategorian data dan terhitung & merubah hasil.



Gambar 4. Set Role

4.3. *Pre-processing*

Pre-Processing dibutuhkan untuk menghilangkan data yang tidak ada atau hilang dan mengidentifikasi data yang tidak lengkap. Untuk memudahkan pengelompokan, nilai matematika dan nilai Inggris dihitung rata-ratanya pada data di atas. Tabel berikut menunjukkan atribut yang diperoleh setelah pre-processing:

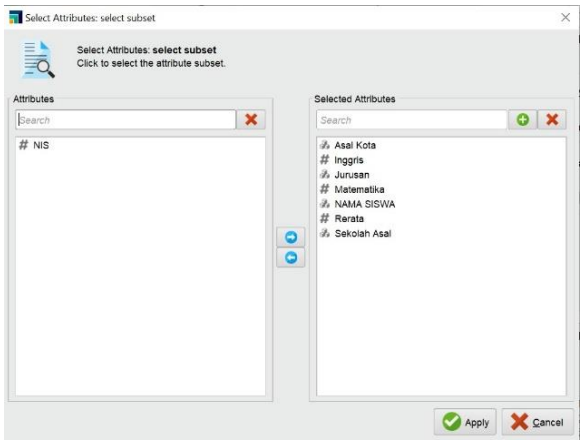
Tabel 1. Data dari excel

No	NIS	NAMA SISWA	Asal Kota	Jurusan	Sekolah Asal	MTK	INGG	UN
1	62487824	Ahmad Fakhrudin	Purwakarta	IPS	SMPN 1 Purwakarta	69	70	69
2	65704219	Ajeng wardah	Purwakarta	IPA	SMPN 1 Purwakarta	69	69	69
3	54561072	lfan Rio Desvandi	Purwakarta	IPS	SMPN 4 Purwakarta	68	69	69
4	65903246	Alifhya Putrie Wagia	Purwakarta	IPA	SMPN 2 Purwakarta	68	68	68
5	69417579	Arkan Lutfi Sonjaya	Purwakarta	IPS	SMPN 1 Purwakarta	68	68	68

Preprocessing data dilakukan dengan menggunakan *software rapidminer*. Setelah dilakukan *preprocessing* data.

Gambar 5. sesudah *preprocessing*

Selanjutnya menggunakan operator Select Atributes. Operator ini menyediakan berbagai jenis filter untuk mempermudah pemilihan atribut dapat dilakukan dengan berbagai cara, seperti pemilihan langsung atribut, seleksi dengan ekspresi reguler, atau memilih atribut tanpa nilai yang hilang. Hanya Atribut yang dipilih yang dikirim ke *port output*. Sisanya dihapus dari *ExampleSet*.



Gamar 6. Hasil Select Attributes

4.4. *Transformation*

Pada tahap transformasi, data diubah atau digabungkan untuk memungkinkan analisis menggunakan metode k-means. Variabel sumber pendidikan dikelompokkan menjadi 9 kelompok, jurusan menjadi 2 kelompok, asal kota menjadi 3 kelompok, nilai matematika menjadi 3 kelompok, nilai bahasa Inggris menjadi 3 kelompok, dan nilai rerata menjadi 3 kelompok.

Tabel 2. Transformation

Asal Kota	Jurusan	Sekolah Asal	MTK	Inggris	Rerata
1	2	1	1	1	1
1	1	1	1	1	1
1	2	4	1	1	1
1	1	2	1	1	1
1	2	1	1	1	1

4.5. Menentukan Titik Pusat *Cluster*

Dalam penelitian ini, telah dipilih 3 kelompok (cluster) yaitu C1, C2, dan C3. Berikut adalah data awal yang akan digunakan:

Tabel 3. Data Centroind

C1	62273413	Indri Sulastri	1	2	2	2	3	3
C2	62273414	Irvan Desamberi	1	2	1	2	3	3
C3	62273415	Japar Abubakar	1	2	6	2	3	3

4.6. Menghitung jarak data setiap cluster

Menghitung jarak semua data ke pusat *cetroid* menggunakan rumus *Euclidean distance*, dibawah ini:

$$\boldsymbol{d(p,q)} = \sqrt{(\boldsymbol{p1}-\boldsymbol{q1})^2+(\boldsymbol{p2}-\boldsymbol{q2})^2+(\boldsymbol{p3}-\boldsymbol{q3})^2}$$

Dengan,
p = nilai data, dan
q = nilai *centroid*.

4.7. Hasil Perhitungan Jarak Iterasi

Tabel 4. Hasil Perhitungan Jarak Iterasi

C1	C2	C3	Jarak Terdekat
2,872	3,041	3,354	2,872
2,872	4,031	3,041	2,872
3,354	4,924	5,679	3,354
3,000	3,000	6,633	3,000
3,000	6,633	3,000	3,000

4.8. Hasil Jarak Cluster

Setelah melakukan perhitungan jarak, langkah berikutnya adalah alokasi data ke dalam masing-masing kluster. Pengalokasian data ini didasarkan pada hasil jarak antara setiap data dengan setiap kluster. Proses pengelompokan data ini dilakukan secara iteratif. Anda dapat melihat hasilnya pada gambar berikut:

Tabel 5. Hasil Pengelompokan Data di Iterasi

C1	C2	C3	Jarak Terdekat	Cluster
2,872	3,041	3,354	2,872	C1
2,872	4,031	3,041	2,872	C1
3,354	4,924	5,679	3,354	C1
3,000	3,000	6,633	3,000	C1
3,000	6,633	3,000	3,000	C1

4.9. Menentukan Titik Pusat Cluster

Setelah penyelesaian proses pengelompokan, langkah berikutnya adalah menentukan titik pusat cluster baru dengan menghitung jumlah dari anggota cluster yang terbentuk, kemudian membaginya dengan jumlah anggota dalam cluster tersebut. Berikut adalah titik pusat cluster baru yang ditemukan dalam tabel di bawah ini:

Tabel 6. Titik Pusat Cluster Baru

C1	4,526	5,965	6,122
C2	4,761	6,247	4,761
C3	4,695	0,000	0,00

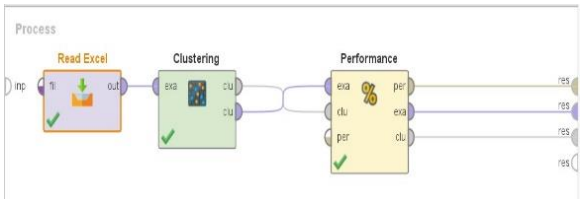
Langkah selanjutnya adalah melakukan penghitungan ulang jarak antara setiap objek dengan centroid baru. Setelah iterasi kedua selesai dan posisi cluster iterasi kedua telah diperoleh, langkah selanjutnya adalah menguji apakah nilai centroid baru dikurangi dengan nilai centroid lama sama dengan nol. Jika kondisi ini terpenuhi, maka iterasi dihentikan. Untuk mendapatkan nilai nol, penelitian memerlukan tiga iterasi tambahan.

Hasil akhir dari penelitian ini adalah pengelompokan berdasarkan kedekatan jarak antara titik pusat dengan semua atribut dari data yang dianalisis. proses menghasilkan clester model dengan 3 cluster dimana cluster 0 terdiri dari 18 item, cluster 1 terdiri dari 12 item, cluster 2 terdiri dari 20 item, cluster 3 terdiri 12 item, dan cluster 4 terdiri 1 item, sehingga total data sebanyak 63 siswa.

4.10. Proses Perhitungan

Dalam proses pembuatan cluster k-means, langkah pertama yang harus dilakukan adalah menentukan jumlah cluster yang optimal (k) berdasarkan jumlah anggota optimal dalam setiap cluster. Hasil dari pengelompokan menggunakan k-means akan menjadi acuan dalam menerima data siswa dari sekolah asal.

Berikut adalah desain proses untuk membuat cluster k-means:



Gambar 7. Desain K-Means

Cluster Model

```
Cluster 0: 18 items
Cluster 1: 12 items
Cluster 2: 20 items
Cluster 3: 12 items
Cluster 4: 1 items
Total number of items: 63
```

Gambar 8. Cluster model

Gambar di atas menampilkan operator utama dari algoritma clustering K-Means. Clustering adalah proses utama dari algoritma K-Means yang bertujuan untuk memodelkan dan menghasilkan hasil dari pengelompokan data. Algoritma K-Means bertugas untuk menentukan satu set k cluster dan mengalokasikan setiap contoh data ke dalam satu cluster yang sesuai dengan karakteristiknya.

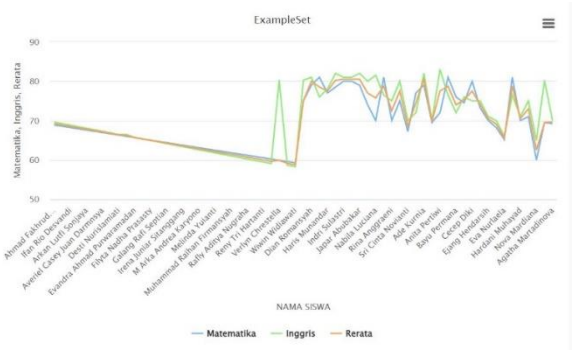
PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -11.990
Avg. within centroid distance_cluster_0: -9.881
Avg. within centroid distance_cluster_1: -23.733
Avg. within centroid distance_cluster_2: -8.828
Avg. within centroid distance_cluster_3: -9.677
Avg. within centroid distance_cluster_4: -0.000
Davies Bouldin: -0.634
```

Gambar 9. Hasil Performance Vector

Hasil dari pengelompokan menggunakan algoritma K-Means dengan menggunakan tool Rapid Miner versi 9.10 untuk mencari nilai K dari K2 hingga K30 menunjukkan bahwa K dengan performa terbaik yang mendekati angka 0 adalah K2. Nilai performa Davies-Bouldin untuk K2 adalah 0.634 pada iterasi pertama dari total iterasi 1 hingga 30. Performa algoritma K-Means dapat dilihat dari nilai Davies-Bouldin, di mana jika nilai mendekati angka 0, maka algoritma dianggap semakin baik. Oleh karena itu, dalam kasus ini, jumlah cluster terbaik adalah 2, yang memiliki nilai Davies-Bouldin yang mendekati 0 dan dianggap sebagai performa terbaik untuk pengelompokan data.

4.11. Evaluation



Gambar 10. Hasil Evaluasi

Setelah melakukan perhitungan clustering karakteristik debitur dengan menggunakan algoritma K-Means, peneliti melakukan evaluasi hasil clustering menggunakan metode Davies-Bouldin Index. Untuk mendapatkan nilai evaluasi tersebut, langkah pertama adalah mencari nilai Sum of Squance Within Cluster (SSW) menggunakan persamaan 3.1 dari data hasil perhitungan jarak data terhadap titik pusat cluster.

5. KESIMPULAN DAN SARAN

Setelah melalui tahapan penerapan algoritma k-means clustering, terdapat beberapa kesimpulan. Penentuan titik pusat pada tahap awal algoritma k-means memiliki pengaruh signifikan terhadap hasil klaster. Pengujian dengan menggunakan 63 dataset dan centroid yang berbeda menghasilkan klaster yang berbeda pula. Setelah melakukan pengelompokan data menggunakan metode k-means, terbentuk 3 klaster dengan rincian: cluster 1 dengan 12 item, cluster 2 dengan 20 item, dan cluster 3 dengan 12 item. Strategi promosi selanjutnya akan mengikuti klaster yang terbentuk berdasarkan minat yang paling tinggi.

Beberapa saran untuk penelitian selanjutnya, sebagai berikut: Disarankan untuk melakukan pengelompokan data siswa setiap tahun ajaran baru, Penelitian ini dapat dijadikan referensi bagi penelitian promosi sekolah di masa depan. Pada penelitian berikutnya, disarankan untuk mencoba menggabungkan metode clustering lainnya guna meningkatkan hasil penelitian yang lebih optimal.

DAFTAR PUSTAKA

[1] H. Indrayani, "PENERAPAN TEKNOLOGI INFORMASI DALAM PENINGKATAN EFEKTIVITAS, EFISIENSI DAN PRODUKTIVITAS PERUSAHAAN Oleh : Henni Indrayani Abstraksi," *Jurnal El-Riyasah*, vol. 3, no. 1, pp. 48–56, 2017.

[2] R. Setiawan, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta)," *Jurnal Lentera Ict*, vol. 3, no. 1, pp. 76–92, 2016.

[3] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4," *Jurnal Edik Informatika*, vol. 2, no. 2, pp. 213–219, 2019.

[4] L. Anang Setiyo and I. F. Bayu Andoro, "PENERAPAN ALGORITMA K-MEANS UNTUK MENENTUKAN STRATEGI PROMOSI (Studi Kasus: Universitas Katolik Widya Mandala Kampus Kota Madiun)," *Seminar dan Konferensi Nasional IDEC*, pp. 2579–6429, 2021.

[5] D. Triyansyah and D. Fitrianah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.

[6] D. Triyansyah and D. Fitrianah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.

[7] A. Rohmah, F. Sembiring, and ..., "Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: Smk Yaspim …," *... Sistem Informasi dan ...*, pp. 290–298, 2021.

[8] R. R. Putra and C. Wadisman, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 1, no. 1, pp. 72–77, Mar. 2018, doi: 10.31539/intecoms.v1i1.141.

[9] A. Rohman and ; Muhammad Rochcham, "Pengelompokan Mahasiswa Berdasarkan Indeks Prestasi Dengan Menggunakan Metode Clustering K-Means (Student Grouping Based on Achievement Index Using the K-Means Clustering Method)," 2020.

[10] T. Hartati, O. Nurdiawan, and E. Wiyandi, "Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari," *Jurnal Sains Teknologi Transportasi Maritim*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.51578/j.sitektransmar.v3i1.30.