

PENGELOMPOKAN ASAL SEKOLAH SISWA MENGGUNAKAN ALGORITMA K-MEANS

Nanita, Martanto, Umi Hayati

Teknik Informatika, STMIK IKMI, Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135

Nitaoup123@gmail.com

ABSTRAK

Sekolah Menengah Kejuruan adalah jenjang pendidikan menengah yang menawarkan program keahlian atau jurusan tertentu yang berfokus pada penguasaan keterampilan praktis. Peningkatan efektivitas strategi pendidikan membutuhkan pemahaman mendalam terhadap profil siswa, khususnya asal sekolah mereka. Dengan banyaknya data calon penerimaan siswa baru yang mendaftar, menimbulkan permasalahan yaitu pihak SMK belum memanfaatkan data asal sekolah tersebut dengan baik. Adapun metode yang dipakai yaitu metode clustering untuk mengelompokkan siswa berdasarkan asal sekolah menjadi sebuah pendekatan untuk memahami dan meningkatkan manajemen siswa baru di Sekolah Menengah Kejuruan Wahidin. Penelitian ini bertujuan untuk melakukan clustering terhadap asal siswa Sekolah Menengah Kejuruan Wahidin di Kota Cirebon menggunakan metode K-Means. Hasil penelitian menunjukkan bahwa pengelompokan siswa berdasarkan asal sekolah mampu membentuk kelompok-kelompok yang memperhatikan perbedaan karakteristik antar kelompok. Dalam penelitian ini peneliti melakukan uji coba dari $K=2$ sampai dengan $K=10$ dan menghasilkan nilai Davies Bouldin Index terbaik yaitu sebesar 0,245 pada $K=10$. Hasil cluster asal sekolah siswa menggunakan metode k-means dapat disimpulkan cluster tersebut dapat dianalisis untuk mendapatkan informasi yang lebih akurat sebagai rujukan dalam menyusun strategi promosi kepada calon siswa baru. Dalam konteks penempatan di kelas, penelitian ini memberikan kontribusi penting dalam pengoptimalan distribusi siswa di berbagai program studi, memungkinkan personalisasi pembelajaran yang lebih efektif. Implikasi praktis dari penelitian ini mencakup penyempurnaan manajemen sekolah, penerapan strategi pembelajaran yang lebih sesuai, dan peningkatan interaksi sosial antar siswa. Dengan adanya informasi ini, lembaga pendidikan dapat meningkatkan efisiensi pengelolaan siswa baru, meningkatkan kualitas pendidikan, dan menciptakan lingkungan belajar yang lebih inklusif. Studi ini juga membuka peluang untuk penelitian lanjutan dalam pengembangan model clustering yang lebih kompleks dan penyesuaian pendekatan pendidikan untuk mendukung keberagaman siswa.

Kata kunci : K-Means, Clustering, Nilai DBI, Strategi Pemasaran, Siswa SMK Wahidin

1. PENDAHULUAN

Pembelajaran merupakan proses interaksi pendidik dan peserta didik. “Pembelajaran adalah proses komunikasi dua arah, mengajar dilakukan oleh pihak guru sebagai pendidik, sedangkan belajar dilakukan oleh peserta didik sebagai murid” [1].

Berdasarkan pengertian ini menambahkan: dalam pembelajaran terdapat kegiatan memilih, menetapkan, dan mengembangkan metode untuk mencapai hasil pembelajaran yang diinginkan. Pemilihan, penetapan, dan pengembangan metode ini didasarkan pada kondisi pembelajaran yang ada. Selain itu, pembelajaran merupakan suatu proses mencapai tujuan belajar, proses mengalami pengetahuan baik dari segi kognitif, afektif, dan psikomotorik Menurut Gagne belajar didefinisikan sebagai suatu proses organisasi perubahan perilaku sebagai akibat dari pengalaman. Pendidikan tingkat menengah kejuruan, khususnya di Sekolah Menengah Kejuruan (SMK) Wahidin di Kota Cirebon, memiliki tantangan unik dalam merancang strategi pendidikan yang efektif. Sehingga dibutuhkan informasi mengenai asal sekolah siswa mana yang banyak mendaftar sehingga dapat dijadikan sebagai rujukan dalam menentukan strategi promosi.

Sekolah Menengah Kejuruan Wahidin di Kota Cirebon sebagai lembaga pendidikan yang fokus pada pengembangan keterampilan teknis, dihadapkan pada kebutuhan untuk memperbaiki proses penempatan siswa baru. Asal sekolah siswa dapat mencerminkan berbagai faktor, termasuk kurikulum sebelumnya, pendekatan pembelajaran, dan kualitas pendidikan di tempat asal siswa. Oleh karena itu, penggunaan metode clustering untuk mengelompokkan siswa berdasarkan asal sekolah menjadi sebuah pendekatan yang menarik untuk memahami dan meningkatkan manajemen siswa baru di Sekolah Menengah Kejuruan Wahidin. Dengan banyaknya data calon penerimaan siswa baru SMK Wahidin belum memanfaatkannya dengan baik. Sehingga SMK Wahidin masih sulit dalam melakukan strategi pemasaran dengan tepat. Permasalahan pertama dalam klastering asal sekolah siswa menggunakan metode k-mean adalah menentukan jumlah klaster yang optimal. Penentuan jumlah klaster yang tidak tepat dapat menghasilkan klaster yang tidak mewakili dengan baik struktur data sebenarnya. Metode k-mean sangat sensitif terhadap posisi awal centroid klaster. Pemilihan posisi awal yang buruk dapat menyebabkan konvergensi ke solusi lokal yang suboptimal, sehingga menimbulkan masalah stabilitas

pada hasil klustering. K-mean rentan terhadap pengaruh outlier karena bekerja berdasarkan rata-rata. Outlier dapat mempengaruhi perhitungan rata-rata di dalam klaster, sehingga menghasilkan klaster yang tidak merepresentasikan dengan baik sebagian besar data. K-mean cenderung menghasilkan klaster yang berbentuk bulat dan sejajar, sehingga tidak efektif untuk mengatasi data dengan struktur klaster yang asimetris atau berbentuk aneh. Hal ini dapat menjadi permasalahan jika asal sekolah siswa memiliki struktur yang kompleks. Jika data asal sekolah siswa tidak terdistribusi merata di seluruh wilayah atau memiliki kepadatan yang berbeda-beda, metode k-mean mungkin tidak efektif. Algoritma ini menganggap kepadatan yang seragam di dalam klaster, sehingga bisa tidak optimal untuk kasus-kasus di mana kepadatan berbeda di setiap klaster.

Kajian terdahulu pernah dilakukan oleh [2] yang berjudul Analisis dan Penerapan Algoritma K-Means Clustering Sebagai Strategi Promosi Penerimaan Mahasiswa Baru Pada Universitas Wiraraja. Penelitian ini bertujuan untuk menganalisis data dalam klasterifikasi promosi mahasiswa baru menggunakan data mining. Penelitian dilakukan menggunakan dataset yang berasal dari data penerimaan mahasiswa baru, pada penelitian ini metode yang digunakan algoritma K-Means Clustering. Jumlah Cluster ditentukan mengacu pada Davies Bouldin Index (DBI) dari beberapa percobaan. Didapatkan nilai terkecil dari DBI yaitu 0.878 pada nilai $k=2$ maka Cluster optimal yang digunakan berjumlah 2. Cluster_0/cluster pertama berjumlah 1460 yang merupakan Cluster Tinggi dan cluster_1/cluster kedua berjumlah 551 yang merupakan Cluster rendah. Perbandingan persentase kedua cluster adalah 73% : 23%. Promosi yang bisa dilakukan dengan memfokuskan dan meningkatkan promosi di kabupaten sumenep dengan memprioritaskan beberapa kecamatan yaitu Kota Sumenep, Kec Kepulauan, Kalianget dan Saronggi. Kajian terkait juga pernah dilakukan oleh [3].

Hasil Riset Yang Dilakukan Oleh [4] Pandemi Covid-19 telah mengubah berbagai aspek kehidupan, termasuk pendidikan, dengan dampak negatif pada minat belajar dan pemahaman konsep matematika siswa. Penelitian ini, berfokus pada siswa kelas sepuluh SMK, menganalisis penggunaan media E-comic dalam materi aritmatika sosial selama pandemi. Dalam penelitian kuantitatif deskriptif ini, 15 siswa dipilih secara acak sebagai sampel. Angket minat belajar siswa disebarakan melalui Google Form, terdiri dari 20 pernyataan. Hasilnya menunjukkan efektivitas media E-comic dalam meningkatkan minat belajar siswa di masa pandemi Covid-19, dengan tingkat persetujuan mencapai 80,67%. Media E-comic dianggap sangat efektif dalam pembelajaran matematika selama pandemi. Hasil Riset Yang Dilakukan Oleh [5]

Literatur ini mencermati dampak negatif dari motivasi belajar yang rendah, seperti

ketidaksemangan dalam belajar dan kesulitan berkonsentrasi. Penyebab utamanya adalah keterbatasan media pembelajaran dan kondisi pembelajaran yang kurang optimal. Hasilnya menunjukkan bahwa 10 artikel relevan dengan topik penelitian ini mengkonfirmasi keefektifan video animasi sebagai media pembelajaran yang meningkatkan motivasi dan hasil belajar peserta didik. Dari tanggapan siswa, guru, dan uji validitas, disimpulkan bahwa media ini layak digunakan untuk meningkatkan semangat dan prestasi belajar.

Penelitian ini bertujuan untuk menggali potensi metode K-Means dalam meningkatkan manajemen siswa baru di Sekolah Menengah Kejuruan Wahidin Kota Cirebon berdasarkan asal sekolah. Melalui pendekatan ini, diharapkan dapat ditemukan pola-pola unik yang dapat membantu lembaga pendidikan dalam mengoptimalkan penempatan siswa dan merancang program pembelajaran yang lebih terfokus. Untuk mengeksplorasi apakah terdapat kelompok siswa yang memiliki kesamaan asal sekolah, dengan harapan dapat memberikan wawasan tentang faktor-faktor tertentu yang memengaruhi pola perilaku atau kinerja akademis. Melalui analisis clustering, penelitian ini bertujuan untuk meningkatkan pemahaman tentang dinamika dan hubungan antara siswa dari berbagai latar belakang sekolah.

Hasil dari penelitian ini diharapkan dapat memberikan informasi yang bermanfaat bagi pihak sekolah dalam mengembangkan strategi pendidikan yang lebih terarah dan mendukung keberhasilan siswa. Dengan menggabungkan teknologi clustering dan konteks pendidikan menengah kejuruan, penelitian ini diharapkan dapat memberikan kontribusi positif terhadap pengelolaan siswa baru di Sekolah Menengah Kejuruan Wahidin dan memperkaya literatur mengenai penggunaan metode K-Means dalam konteks pendidikan tingkat menengah sehingga bisa mempermudah untuk mengetahui banyaknya siswa yang masuk agar bisa meningkatkan tingkat promosi ke daerah yang memiliki calon siswa terbanyak.

Menganalisis Penelitian bertujuan untuk menggunakan metode k-mean dalam klustering untuk menganalisis dan mengidentifikasi pola distribusi siswa berdasarkan asal sekolah, sehingga dapat memahami secara holistik struktur spasial dari populasi siswa. Penelitian ini bertujuan untuk mengoptimalkan penempatan fasilitas dan sumber daya pendidikan dengan memahami pembagian siswa ke dalam klaster yang homogen berdasarkan sekolah asal, memberikan wawasan yang berharga bagi perencanaan dan pengelolaan sumber daya. Penelitian ini bertujuan untuk memahami karakteristik khusus dari kelompok siswa yang berasal dari sekolah-sekolah tertentu dengan menggunakan klustering k-mean, sehingga dapat memberikan wawasan lebih mendalam terkait kebutuhan pendidikan mereka. Dengan mengelompokkan siswa berdasarkan asal

sekolah menggunakan metode k-mean, penelitian ini bertujuan untuk meningkatkan efisiensi program pendidikan regional dengan memberikan informasi yang lebih presisi tentang distribusi spasial siswa, membantu dalam perencanaan dan evaluasi program pendidikan. Mengembangkan model yang memiliki kemampuan umum untuk mengatasi variasi dalam data siswa dari berbagai konteks sekolah, sehingga dapat diterapkan secara luas dan relevan di berbagai lingkungan pendidikan.

Hasil dari penelitian ini menghasilkan 3 kluster yaitu cluster_0, cluster_1, dan cluster_2 yang akan membantu dalam merancang strategi dalam mempromosikan sekolah. Dengan mengetahui informasi kluster dapat memberikan pemahaman lebih mendalam tentang struktur demografis siswa dan mendukung pengembangan strategi promosi yang lebih terarah. bertujuan untuk memahami karakteristik khusus dari kelompok siswa yang berasal dari sekolah-sekolah tertentu dengan menggunakan klustering k-means. Sehingga dapat memberikan wawasan lebih mendalam terkait kebutuhan pendidikan mereka. mengelompokkan siswa berdasarkan asal sekolah menggunakan metode k-means.

Hasil dari penelitian ini bertujuan untuk meningkatkan efisiensi program pendidikan regional dengan memberikan informasi yang lebih presisi tentang distribusi spasial siswa. Mengembangkan model yang memiliki kemampuan umum untuk mengatasi variasi dalam data siswa dari berbagai konteks sekolah, sehingga dapat diterapkan secara luas dan relevan di berbagai lingkungan pendidikan. Melalui pendekatan ini, diharapkan dapat ditemukan pola-pola unik yang dapat membantu lembaga pendidikan dalam mengoptimalkan penempatan siswa dan merancang program pembelajaran yang lebih terfokus. Untuk mengeksplorasi apakah terdapat kelompok siswa yang memiliki kesamaan asal sekolah.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Sebagai bidang ilmu yang relatif baru, data mining saat ini mendapatkan perhatian di kalangan ilmuwan dan profesional. Penambangan data melibatkan pengambilan informasi berguna dari database besar. Ini menjadi informasi baru dan perlu digali untuk membantu pengambilan keputusan. Data mining adalah proses menganalisis data dari berbagai sumber dan menggabungkannya untuk menciptakan informasi, pengetahuan, atau pola yang penting untuk meningkatkan keuntungan, mengurangi biaya, atau keduanya [6].

2.2. Clustering

Clustering adalah salah satu dari banyak teknik penambangan data yang bertujuan untuk mengelompokkan beberapa objek berdasarkan properti serupa. [7].

Clustering adalah proses mengelompokkan objek objek serupa ke dalam kelompok-kelompok berbeda, dan lebih khusus lagi, membagi kumpulan data menjadi beberapa subset sedemikian rupa sehingga data dalam setiap subset memiliki arti yang berguna. Sebuah cluster terdiri dari kumpulan objek-objek yang serupa satu sama lain dan berbeda dengan objek-objek yang terdapat dalam cluster lainnya.

Algoritma pengelompokan terdiri dari dua bagian: hierarki dan terpartisi. Algoritme hierarki mencari cluster secara berurutan dimana cluster telah ditentukan sebelumnya, sedangkan algoritma parsial mencari semua grup pada titik waktu tertentu [8].

2.3. Algoritma K-Means

Algoritma K-Means merupakan algoritma yang mengelompokkan data menggunakan sekumpulan parameter dan membagi data tersebut ke dalam cluster-cluster berdasarkan kemiripan antara data pada suatu cluster dengan data pada cluster lainnya. Adapun metode yang dipakai yaitu metode clustering untuk mengelompokkan siswa berdasarkan asal sekolah menjadi sebuah pendekatan untuk memahami dan meningkatkan manajemen siswa baru di Sekolah Menengah Kejuruan Wahidin.

Penelitian ini bertujuan untuk melakukan clustering terhadap asal siswa Sekolah Menengah Kejuruan Wahidin di Kota Cirebon menggunakan metode K-Means. Hasil penelitian menunjukkan bahwa pengelompokan siswa berdasarkan asal sekolah mampu membentuk kelompok-kelompok yang memperhatikan perbedaan karakteristik antar kelompok. Dalam penelitian ini peneliti melakukan uji coba dari $K=2$ sampai dengan $K=10$ dan menghasilkan nilai Davies Bouldin Index terbaik yaitu sebesar -0,245 pada $K=10$. Hasil cluster asal sekolah siswa menggunakan metode k-means dapat disimpulkan cluster tersebut dapat dianalisis untuk mendapatkan informasi yang lebih akurat sebagai rujukan dalam menyusun strategi promosi kepada calon siswa baru. Dalam konteks penempatan di kelas, penelitian ini memberikan kontribusi penting dalam pengoptimalan distribusi siswa di berbagai program studi, memungkinkan personalisasi pembelajaran yang lebih efektif.[9]

K-means merupakan teknik analisis kelompok yang membagi N objek yang diamati menjadi K kelompok (cluster). Setiap objek yang diamati termasuk dalam kelompok yang nilai mean-nya paling dekat. Ini mirip dengan algoritma maksimalisasi ekspektasi campuran Gaussian dan mencoba keduanya. Temukan pusat grup pada data dengan jumlah iterasi penyempurnaan yang sama yang dilakukan oleh kedua algoritma.[10]

2.4. Strategi Promosi

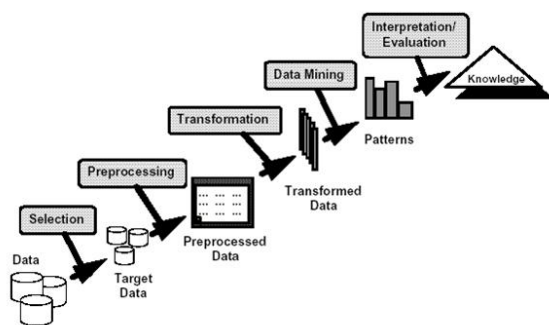
Strategi promosi adalah rencana yang dirancang untuk memperkenalkan, memasarkan, dan mempromosikan produk atau layanan agar dapat mencapai tujuan tertentu, seperti meningkatkan

penjualan, kesadaran merek, atau loyalitas pelanggan. Tujuan utama dari strategi promosi adalah untuk mempengaruhi perilaku target pasar dan membuat mereka tertarik atau terlibat dengan produk atau layanan yang ditawarkan.[7]

3. METODE PENELITIAN

3.1. Metode Penelitian

Pada penelitian ini menggunakan metode penelitian K-Means Clustering dengan pendekatan kuantitatif yang dimana pada hasilnya akan menampilkan angka statistic dalam bentuk visualisasi grafik. Pada metode penelitian ini menggunakan tahapan KDD (Knowledge Discovery Database). KDD merupakan sebuah Langkah praktis dalam melakukan tahapan data mining untuk mencari, mengidentifikasi pola pada data, dimana pola yang ditemukan bersifat sah dan bermanfaat dan mudah dimengerti. Dalam tahapan KDD terdapat beberapa tahapan diantaranya pembersihan data, integrasi data, pemilihan data, seleksi data, transformasi data. Berikut adalah gambar mengenai tahapan – tahapan dalam KDD (Knowledge Discovery Database):



Gambar 1. Tahapan KDD

Pada penelitian ini menggunakan metode penelitian K-Means Clustering dengan pendekatan kuantitatif yang dimana pada hasilnya akan menampilkan angka statistic dalam bentuk visualisasi grafik adapun penjelasan dari tahapan KDD di atas :

a. Data

Data yang diperoleh secara langsung dari SMKN Wahidin dengan format. Akan memproses data yang sudah diperoleh untuk diolah kedalam Rapidminer.

b. Data Selection

Pada tahap ini dilakukan proses pemilihan data dari sekumpulan data operasional yang perlu dilakukan sebelum melakukan tahap KDD (Knowledge Discovery in Databases). Data yang telah terpilih dan akan digunakan dalam proses data mining kemudian disimpan dalam suatu berkas yang terpisah dari basis data operasional. Yang selanjutnya data yang terpilih akan diimport kedalam Rapidminer untuk dilakukan pemrosesan data.

c. Pre-processing

Pada tahap Pre-processing, dilakukan pemeriksaan terhadap nilai atau atribut yang tidak akan digunakan atau tidak relevan agar data yang digunakan menjadi valid. Membersihkan data atau atribut yang tidak memiliki korelasi menggunakan operator Correlation Matrix, selanjutnya setelah kita mengetahui atribut apa saja yang tidak digunakan akan kita hilangkan menggunakan operator Select Atribut.

d. Transformasi Data

Pada tahap ini digunakan untuk melakukan inisialisasi terhadap data yang memiliki nilai normal menjadi nilai numerik. Pada proses transformasi data mengubah nilai normal menjadi Numerik menggunakan operator Nominal to Numerical.

e. Data Mining

Pada tahap ini, dilakukan proses pemilihan algoritma atau teknik data mining yang akan digunakan untuk mencari informasi yang relevan dalam data yang telah dipilih. Algoritma yang digunakan adalah Algoritma K-MEANS Clustering, pada tahap ini untuk menentukan nilai klaster dan mencari nilai DBI menggunakan operator K-MEANS Clustering dan operator Cluster Distance Performance .

f. Interpretasi dan Evaluasi

Tahap evaluasi digunakan untuk mengevaluasi hasil dari tahap data mining dan memastikan bahwa hasil tersebut dapat menjawab tujuan yang telah ditetapkan. Evaluasi melibatkan profilisasi pada setiap cluster yang terbentuk dan analisis lebih lanjut untuk menghubungkannya dengan atribut yang relevan.

3.2. Sumber Data

Sumber data ini mencakup data siswa smk wahidin. Pengambilan data dimulai dari 20 Oktober 2023 sampai dengan 10 Desember 2023. Data yang digunakan dalam penelitian ini adalah data siswa kelas X yang diperoleh langsung dari smk wahidin kota cirebon, Data-data tersebut berisi Nis, Nama lengkap, Jenis kelamin, Asal SMP/MTS, Nama desa/Kelurahan, Kecamatan, Kota/Kabupaten, Dan Provinsi. Jumlah data yang digunakan dalam penelitian ini, berjumlah 779 data siswa kelas X tahun ajaran 2022-2023. Data siswa ini untuk mengetahui asal siswa terbanyak yang masuk ke smk wahidin.

3.3. Populasi dan Sempel

Populasi adalah seluruh kelompok individu atau objek yang menjadi subjek penelitian. Sampel adalah sebagian elemen yang dipilih secara acak atau dengan kriteria tertentu dari populasi yang lebih besar, yang digunakan sebagai perwakilan atau representasi dari keseluruhan populasi dalam konteks penelitian. Dalam penelitian ini populasi yang di dapat banyaknya asal siswa yang masuk ke SMK Wahidin

dari daerah setiap tahunnya Populasi yang digunakan adalah seluruh siswa SMK Wahidin yaitu sebanyak 779. Teknik Pengambilan sampel adalah sampling jenuh, jadi semua populasi dijadikan sampel yaitu sebanyak 779.

3.4. Teknik pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah dengan survey observasi yang dilakukan langsung ke SMK Wahidin Kampus 2 Kota Cirebon yang menghasilkan data siswa kelas X periode 2022-2023 sebanyak 779 data siswa.

3.5. Teknik Analisis Data

Setelah melakukan pengumpulan data, data tersebut akan dianalisis serta diidentifikasi menggunakan analisis statistik deskriptif yang bertujuan untuk menggambarkan data statistik dengan numerik ataupun dalam bentuk visualisasi grafis. Pada Teknik analisis data membutuhkan beberapa tahapan algoritma k-means clustering yaitu KDD diantaranya :

- Pengumpulan Data, yaitu untuk melakukan Pengumpulan data, yaitu untuk Melakukan pengumpulan data dari berbagai sumber data yang telah didapat.
- Pembersihan data, yaitu melakukan pembersihan pada data yang tidak konsisten atau memiliki nilai dan atribut kosong.
- Integrasi data, yaitu menggabungkan data dari setiap database ke dalam database baru.
- Pemilihan data, yaitu menyeleksi data digunakan untuk menganalisis atribut pada data yang sesuai dan relevan untuk dilakukan metode data mining sehingga data akan diseleksi terlebih dahulu.
- Transformasi data yaitu, mengubah dan menggabungkan data kedalam sebuah format yang tepat untuk dilakukan proses data mining
- Data Mining yaitu, Menerapkan data mining pada data yang akan diolah dan dianalisis menggunakan system pengelompokan yaitu clustering dan algoritma K-means, serta menggunakan tools Rapidminer
- Evaluation, yaitu mengevaluasi pola pada hasil yang telah ditemukan pada penerapan data mining K-mean Clustering tersebut untuk melihat keakuratan dan ketepatan hasil

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Data yang dipilih pada penelitian ini adalah data siswa kelas X pada tahun 2022-2023. Yang merupakan data real dari SMK Wahidin Kota Cirebon, dan data ini memiliki 779 record dan Dari 9 atribut yang ada pada jenis data, Hasil selection dari dataset yang digunakan hanya 3 attribute memiliki nilai yang sama pada masing-masing atributnya seperti tampilan table berikut:

Tabel 1. Hasil Penelitian

No.	Atribut	Type Data	Keterangan
1.	No	Integer	No
2.	Nis	Numeric	Nomer Induk Siswa
3.	Nama Lengkap	Numeric	Nama Siswa
4.	Jeni kelamin	Numeric	Gender
5.	Asal sekolah	Polynomial	Asal Sekolah
6.	Desa/kelurahan	Numeric	Kelurahan
7.	Kecamatan	Numeric	Kecamatan
8.	Kota/KabupKaten	Numerik	Kabupaten
9.	Provinsi	Numerik	Provinsi

No, Kecamatan, Nama Desa, Nama Lengkap, Nis, Provinsi. walaupun pada atribut description masih ada lagi beberapa yang dapat dijadikan atribut baru tetapi dapat diabaikan karena banyak memiliki nilai missing. Sehingga dari 9 atribut menjadi seperti tampak pada tabel berikut:

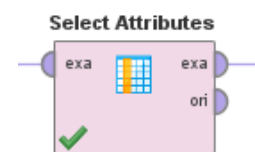
Tabel 2. Atribut Data Siswa

No.	Atribut	Type Data	Keterangan
1.	Asal Sekolah	Polynomial	Asal Sekolah
4.	Jenis Kelamin	Numeric	Gender
5.	Kota/Kabupaten	Numeric	Kabupaten

File data siswa setelah diekstrak dari description, disimpan dengan nama yang sama tetapi berekstension xls. Untuk membaca dataset dalam bentuk file excel, menggunakan operator Read Excel seperti tampak pada gambar 2.



Gambar 2. Atribut Read Exsel



Gambar 3. Operator Read Excel

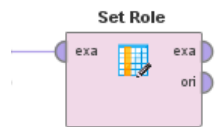
Parameter pada operator Read Excel, menggunakan operator default. Untuk seleksi selanjutnya pada dataset, digunakan operator *Select Attribute* seperti tampak pada gambar 3.

Atribut yang tidak dipilih yaitu :

- Nis, karena atribut ini tidak dibutuhkan
- Nama, karena atribut ini tidak dibutuhkan
- Kelurahan, karena atribut ini tidak dibutuhkan
- Kecamatan, karena atribut ini tidak dibutuhkan
- Provinsi, karena atribut ini tidak dibutuhkan

Proses cleansing atau pembersihan data yang missing atau memiliki nilai yang tidak konsisten pada langkah preprocessing. Sebelum melakukan proses ini, dilakukan analisa terlebih dahulu apakah atribut

pada dataset yang dipilih memiliki nilai missing atau tidak serta memiliki data yang konsisten atau tidak. dataset memiliki data yang konsisten terhadap nilainya. Ternyata atribut Office memiliki nilai NA (no answer) atau dengan kata lain tidak memiliki nilai atau missing, selanjutnya menambahkan operator Set Role.



Gambar 4. Set Role

4.2. Pre-Processing

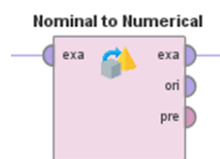
Pada tahap cleaning ini, data dibersihkan dari duplikat, kesalahan pada data, dan perubahan data. Sebelum pada tahap ini dilakukan terlebih dahulu analisa pada atribut dataset yang akan digunakan apakah memiliki missing value atau tidak serta apakah data yang akan digunakan konsisten atau tidak seperti tampak pada gambar dibawah ini.

Name	Type	Missing	Statistics	Filter (5/5 attributes)	Search for Attributes
Id	Integer	0	Min: 1 Max: 779 Average: 390		
cluster	Nominal	0	Least: cluster_2 (6) Most: cluster_0 (456)		
Asal SMP/ITS	Polynomial	0	Least: UPT SMP (1) Most: SMP N 13 (54) Average: 0.368		
Jenis Kelamin	Numeric	0	Min: 0 Max: 2		
Kota/Kabupaten	Numeric	0	Min: 0 Max: 10 Average: 2.270		

Gambar 5. Pre Processing

4.3. Transformasi Data

Data transformasi pada tahap ini digunakan untuk melakukan inisialisasi terhadap data yang memiliki nilai normal menjadi nilai numerik karena algoritma k-means harus bertipe numerik. Pada tahap ini Asal Sekolah dan Kota/Kabupaten di transformasikan menjadi nilai numerik menggunakan operator Nominal to Numerical seperti pada gambar dibawah ini.



Gambar 6. Nominal to Numerical

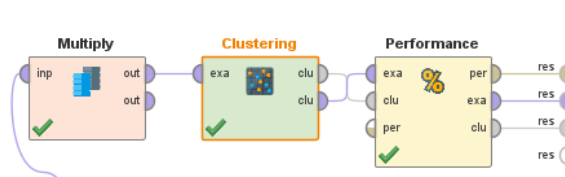
Operator Nominal to Numerical membantu proses dataset untuk dapat diproses pada data mining dengan cara merubah data polynominal menjadi numerical, tahap ini bertujuan untuk memudahkan proses clustering.

4.4. Data Mining

Pada tahap ini dilakukan penerapan algoritma dan metode data mining, Adapun algoritma yang digunakan adalah algoritma k-means dan menggunakan metode clustering. Di tahap ini

langkah-langkah algoritma kmeans diterapkan untuk melihat hasil akhir yang telah diproses oleh data mining sesuai dengan atribut pada data yang telah di transformasi.

Untuk tahap awal menentukan jumlah k pada cluster, hal ini bertujuan untuk melihat jumlah yang efektif dalam metode clustering. Untuk uji cluster yang akurat, pada rapid miner akan membandingkan menjadi 9 kelompok. Dibandingkan dengan cluster jumlah k=2, k=3, k=5, k=6, k=7, k=8, k=9 dan k=10. Untuk dapat menambah operator cluster lebih dari satu pada rapidminer, bisa menggunakan operator multiply yang bertujuan untuk menduplikat operator pada rapidminer.



Gambar 7. Rapid Miner

4.5. Evaluation

Pada tahap evaluasi dilakukan setelah melakukan tahap data mining. Tahapan ini untuk menentukan hasil keakuratan yang didapatkan oleh proses data mining, untuk menentukan seberapa akurat dan ketepatan dari hasil yang diperoleh dengan menyesuaikan dengan nilai performa pada data mining. Semakin kecil hasil perfor

Pada tahap ini hasil dari masing-masing cluster akan ma pada data mining maka semakin baik dan akurat. dilihat nilai performanya baik yang menggunakan k=2, k=3, dan k=4. Untuk melihat nilai performa pada cluster, dalam rapidminer bisa menggunakan operator cluster distance performance.



Gambar 8. operator cluster distance performance

Dan untuk nilai k=10 didapatkan hasil, dengan nilai dbi (davies bouldin index) 0.245. Berdasarkan hasil pada masing-masing nilai DBI (Davies Bouldin Index) hasil yang paling rendah diantara masing-masing cluster adalah dengan jumlah k=4 dengan nilai DBI sebesar 0.131, maka cluster dengan jumlah k=4 lebih baik dibandingkan dengan cluster jumlah k=2, k=3, k=5, k=6, k=7, k=8, k=9 dan k=10.

5. KESIMPULAN DAN SARAN

Penentuan jumlah cluster optimal dapat ditentukan menggunakan tiga kali pengujian nilai k berbeda dengan menggunakan jarak Euclidean

distance dan uji nilai DBI. Evaluasi yang digunakan menggunakan DBI (Davies Bouldin Index) dengan mendapatkan pengujian nilai k, untuk nilai k=4 menjadi nilai k yang paling rendah berdasarkan hasil performa pada DBI dengan nilai akurasi sebesar 0.131 dan dinyatakan sebagai jumlah cluster yang optimal. Dalam pengelompokan yang dilakukan oleh penulis dapat memiliki nilai akurasi yang lebih baik dan akurat. Dengan Analisa dataset siswa kelas X dengan menggunakan *K-Means* ini semoga selanjutnya dapat membandingkan dengan algoritma yang lain.

DAFTAR PUSTAKA

- [1] H. Pratama *et al.*, “Analisis Kesulitan Belajar Siswa Di Smk Pgri 1 Mejayan Pada Materi Kelistrikan,” *Pros. Semin. Nas. Pendidik. Sains*, no. November, pp. 267–275, 2015.
- [2] I. Iddrus and F. Helmi, “Analisis dan Penerapan Algoritma K-Means Clustering Sebagai Strategi Promosi Penerimaan Mahasiswa Baru Pada Universitas Wiraraja,” *JUSTIFY J. Sist. Inf. Ibrahimy*, 2023, [Online]. Available: <https://journal.ibrahimy.ac.id/index.php/JUSTIFY/article/view/3205>
- [3] A. L. Maukar, F. Marisa, and ..., “Analisis Data Penerimaan Mahasiswa Baru Berbasis K-Means,” *JIKO (Jurnal Inform. ...)*, 2022, [Online]. Available: <https://ejournal.akakom.ac.id/index.php/jiko/article/view/558>
- [4] Y. A. BatuBara, Z. Zetriuslita, A. Dahlia, and ..., “Analisis minat belajar siswa menggunakan media pembelajaran e-comic aritmatika sosial masa pandemi Covid-19,” *J. Deriv. J. ...*, 2021.
- [5] S. Maulani, N. Nuraisyah, D. Zarina, I. Velinda, and ..., “Analisis Penggunaan Video sebagai Media Pembelajaran Terpadu terhadap Motivasi Belajar Siswa,” *J. Pendidik. dan ...*, 2022.
- [6] A. K. Wahyudi, N. Azizah, and ..., “DATA MINING KLASIFIKASI KEPRIBADIAN SISWA SMP NEGERI 5 JEPARA MENGGUNAKAN METODE DECISION TREE ALGORITMA C4. 5,” *J. Inf. ...*, 2022, [Online]. Available: <https://journal.unisnu.ac.id/JISTER/article/view/392>
- [7] A. Bellanov, “K-Means Clustering Analysis Untuk Menentukan Strategi Promosi Kampus,” *J. Tek. Ind. J. Has. Penelit. ...*, 2023, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/jti/article/view/22492>
- [8] R. Pormes and D. H. F. Manongga, “Penggunaan Algoritma Clustering K-means Untuk Melihat Daerah-Daerah Penyuplai Mahasiswa Di Universitas Kristen Satya Wacana, Salatiga.,” *J. Tek. Inform. dan ...*, 2019, [Online]. Available: <https://journal.maranatha.edu/index.php/jutisi/article/view/1968>
- [9] M. A. Hairudin, H. Zainuddin, and ..., “Rekomendasi Strategi Sosialisasi Program Studi Melalui Jalur Undangan Menggunakan Algoritma ID3 dan K-Means,” *JITCE (Journal ...)*, 2022, [Online]. Available: <http://jitce.fti.unand.ac.id/index.php/JITCE/article/view/80>
- [10] S. Syahputra, S. Ramadani, and ..., “Menentukan Strategi Promosi Menggunakan Algoritma Clustering K-Means,” *JOISIE (Journal ...)*, 2020, [Online]. Available: <https://www.ejournal.pelitaindonesia.ac.id/ojs32/index.php/JOISIE/article/view/510>