

KLASTERISASI DATA KEGEMARAN MEMBACA MENGGUNAKAN ALGORITMA K-MEANS DI SMA AL-ISLAM CIREBON

Luthfia Fahmi Az Zahroh, Nining Rahaningsih, Raditya Damar Dana

Teknik Informatika, STMIK IKMI Cirebon
Jl.Perjuangan No 02B Majasem Cirebon Jawa Barat
Fluthfia42@gmail.com

ABSTRAK

Kesejahteraan siswi yang tinggi terjadi apabila tingkat kegemaran membaca juga tinggi. Dalam upaya meningkatkan kegemaran membaca diperlukan suatu tolak ukur penelitian yang menjadi dasar dari kebijakan yang akan dilakukan maupun suatu pengambilan keputusan. Tujuan dari penelitian ini untuk menganalisis pola kegemaran membaca buku pada suatu populasi menggunakan metode pengelompokan data dengan algoritma *K-Means*. Data kegemaran membaca buku ini dianalisis menggunakan algoritma *K-Means* untuk membentuk kelompok-kelompok dengan karakteristik kegemaran membaca yang serupa. Dari hasil kajian nilai indeks kegemaran membaca dapat diketahui hasil fenomena kegemaran membaca pada siswi SMA Al Islam Cirebon. Hasil analisis ini diharapkan dapat membantu memberikan wawasan mendalam mengenai ragam preferensi dalam hal membaca buku di antara individu-individu dalam populasi tertentu. Hasil penelitian ini menunjukkan bahwa terdapat data yang dikelompokkan menjadi beberapa *cluster* berdasarkan kegemaran membaca buku. Selain itu, penelitian ini menunjukkan bahwa algoritma *K-Means* dapat digunakan sebagai metode analisis *cluster* yang efektif pada data kegemaran membaca buku. Didapatkan hasil pada analisis *cluster* pada data kegemaran membaca buku sebanyak empat *cluster*, dengan jumlah *cluster* optimal yang didapatkan dari penelitian yaitu $k=2$ dengan nilai *davies bouldin* 0.970.

Kata Kunci: Analisis, Kegemaran, K-Means, Pola.

1. PENDAHULUAN

Dalam era perkembangan teknologi pesat di bidang informatika, transformasi digital telah menciptakan dampak yang signifikan pada berbagai lapisan kehidupan, termasuk dalam aspek kegemaran membaca buku. Membaca sangat penting bagi kehidupan manusia. Kegiatan membaca buku merupakan kegiatan kognitif yang mencakup proses penyerapan pengetahuan, pemahaman, kemampuan analisis, kemampuan sintesis, dan kemampuan evaluasi[1]. Dengan terbiasa membaca maka seseorang akan memiliki cakrawala pengetahuan yang luas, kreativitas terbuka, imajinasi tinggi, pemikiran yang maju dan berkembang serta menjadi cikal bakal pemberdayaan manusia yang cerdas dan berintelektual. Membaca adalah wujud dari sifat pembelajar. Sangat pantas jika buku disebut sebagai jendela dunia dan membaca merupakan investasi masa depan, tapi mengapa masyarakat kita kurang gemar membaca?[1]. Pada saat ini khususnya bagi para siswi baik dari jenjang pendidikan dasar sampai perguruan tinggi kegiatan membaca cenderung sangat jarang sekali dilakukan sehingga hal ini menyebabkan susahny siswi itu sendiri untuk proses pengetahuannya[2].

Penelitian ini bertujuan untuk menganalisis kegemaran membaca buku pada data yang tersedia menggunakan metode *K-Means* dan mengatasi permasalahan yang muncul dalam penggunaannya. Selain itu, Penelitian ini memiliki signifikansi yang besar dalam mengisi kesenjangan pengetahuan terkait dengan penggunaan algoritma *K-Means* dalam konteks kegemaran membaca buku. Selain itu, hasil

dari penelitian ini diharapkan mampu memberikan banyak kontribusi pada pengembangan metode analisis data yang lebih efektif dan efisien pada data kegemaran membaca buku. Secara efisien, hasil dari penelitian ini digunakan untuk membantu dalam memahami pola kegemaran membaca buku, yang dapat memiliki implikasi pada pengelolaan perpustakaan, penerbitan buku, dan bahkan pengembangan kebijakan pendidikan. Dengan demikian, penelitian ini memiliki potensi manfaat yang luas dalam berbagai aspek kehidupan.

Untuk mengeksplorasi lebih dalam tentang preferensi membaca buku dan sifat kelompok yang muncul, analisis *cluster* pada penelitian ini akan menggunakan metode *K-Means Clustering* pada dataset preferensi membaca buku. Analisis *cluster* merupakan proses pengelompokan data berdasarkan tingkat kemiripan atau kedekatan antar data[3]. Dengan menerapkan teknik *clustering*, dapat dikenali pola atau karakteristik yang dimiliki oleh setiap kelompok data. *K-Means Clustering* merupakan algoritma *Machine Learning* yang populer untuk analisis data tanpa supervisi. Algoritma ini membagi pengamatan menjadi kelompok sedemikian rupa sehingga tiap pengamatan tergabung dalam *cluster* dengan jarak rata-rata terdekat. Salah satu algoritma data mining, algoritma *K-Means*, dapat melakukan pengelompokan (*clustering*) data[4].

2. TINJAUAN PUSTAKA

Dalam penelitian ini metode literature review yang digunakan adalah dengan cara mereview artikel-artikel terdahulu terkait dengan analisis *cluster* pada

data kegemaran membaca buku menggunakan algoritma *K-Means*. Adapun metode untuk melakukan literature review melalui empat tahapan, pertama formulasi permasalahan, kedua pencarian literature, ketiga evaluasi data, dan terakhir analisis dan interpretasi[5].

Penelitian terdahulu dari [6] yang membahas mengenai “Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan *K-Means Clustering*”. Penelitian yang dilakukan bertujuan untuk melakukan klasterisasi penyebaran penyakit pada pasien untuk menentukan pola berdasarkan usia pasien dengan menggunakan metode analisis *K-Means Clustering*. Dalam penelitian tersebut, digunakan teknik *data mining* dan algoritma *K-Means Clustering* untuk menempatkan pasien dalam kelompok usia berdasarkan kelompok usia yang mungkin memiliki penyebaran penyakit. Berdasarkan dari hasil *cluster* pada dataset kunjungan pasien, kesimpulan dari jurnal ini bahwa pasien dewasa dan anak – anak kemudian disusul dengan pasien lanjut usia dan terakhir pasien remaja merupakan kelompok usia yang paling rentan mengidap penyakit pada Puskesmas Muara Enim adalah pada setiap tahunnya[2].

Penelitian selanjutnya dari [7]. Penelitian ini melibatkan analisis *data mining* pada data penjualan merek kacamata Optik Chantika menggunakan metode clustering *K-Means*. Tujuan penelitian ini adalah untuk menemukan pola penjualan baru berdasarkan usia pasien menggunakan algoritma *clustering K-Means*. Sebuah kunci-kunci yang diperoleh dari penelitian ini yaitu algoritma *K-Means* sebagai metode untuk mengelompokkan data penjualan merek kacamata menjadi tiga kelompok: penjualan tertinggi, penjualan sedang dan penjualan terendah. Dari penelitian menunjukkan bahwa data pasien anak-anak dan pasien dewasa merupakan kelompok rentang usia yang paling tinggi mengidap penyakit, diikuti oleh pasien lansia, dan pasien remaja setiap tahunnya. Dengan menerapkan metode *K-Means*, Optik Chantika dapat meningkatkan kepuasan pelanggan dan efisiensi manajemen data penjualan. Hasil penelitian ini menunjukkan bahwa lima puluh persen merek kacamata berada dalam kategori penjualan tertinggi, 27,27 persen berada dalam kategori sedang, dan 22,72 persen berada dalam kategori terendah[7].

Penelitian selanjutnya dari artikel [8]. Dalam penelitian yang berjudul "Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma *K-Means*", Data Jamaah Umrah pada Tour Tanurmutmainah dikelompokkan dengan teknik *K-Means Clustering*. Berikut adalah beberapa parafrase dari kalimat tersebut:

a. Penelitian ini bertujuan untuk menggunakan metode *K-Means Clustering* untuk mengelompokkan data Jamaah Umrah pada Tanurmutmainah Tour.

- b. *K-Means Clustering Method* membagi data menjadi beberapa kelompok dan menerima masukan berupa data tanpa label kelas..
- c. Algoritma *K-Means Clustering* merupakan algoritma non-hirarki yang berfungsi untuk mengelompokkan data dengan karakteristik yang sama.
- d. Proses klasterisasi *K-Means Clustering* melibatkan beberapa langkah, seperti mengatur nilai k, penentuan sentris, pembagian data, dan iterasi hingga nilai sentris tidak berubah.
- e. Hasil penelitian ini mungkin memberikan informasi tentang pola Jamaah Umrah pada Tanurmutmainah Tour, yang dapat membantu dalam pengelolaan dan pengembangan strategi bisnis.

Dalam konteks penelitian ini, dalam mengelompokkan data jamaah umrah pada Tanurmutmainah Tour digunakan metode *K-Means Clustering*, yang dapat membantu dalam pengelolaan dan pengembangan strategi bisnis. Hasil penelitian ini mungkin memberikan informasi tentang pola jamaah umrah pada Tanurmutmainah Tour, yang dapat membantu dalam pengelolaan dan pengembangan strategi bisnis[4].

Penelitian selanjutnya dari artikel [9] dengan judul Artikel “Peningkatan Tingkat Kegemaran Membaca (TGM) Siswi Melalui Pengukuran Indeks Gemar Membaca (IGM)” membahas tentang bagaimana mengukur budaya kegemaran membaca buku pada siswi dilakukan untuk memberikan tanggapan terhadap indikator kinerja utama pada perpustakaan, prinsip penyusunan program, dan strategi pengembangan perpustakaan, serta landasan untuk menyusun arah kebijakan perpustakaan. Rendahnya Indeks Literasi Siswi(ILM) menjadi permasalahan dan metode analisis Structural Equation Modeling (SEM) digunakan untuk mengidentifikasi komponen yang paling signifikan yang mempengaruhi kegemaran dan kebiasaan membaca[9].

Penelitian selanjutnya dari [4] dengan judul artikel “Analisis Klasterisasi Siswi Baru dalam Memilih Program Studi dengan Menggunakan Algoritma *K-Means*” pada tahun 2019 sampai 2020 siswi baru di Universitas Islam Kuantan Singingi terjadi penurunan yang cukup signifikan dalam jumlah penerimaan siswi baru. Penurunan ini mengakibatkan ketidakproporsionalan antara jumlah dosen dan siswi di beberapa program studi, yang dapat memberikan dampak negatif pada institusi. Pada permasalahan tersebut, penelitian ini membuat klasterisasi pada siswi baru untuk menentukan program studi yang sesuai dan diminati siswi, klasterisasi ini dilakukan untuk mengevaluasi sejauh mana peningkatan atau penurunan jumlah siswi di setiap program studi. Tujuan utamanya adalah memberikan informasi yang berfungsi sebagai pertimbangan pimpinan dalam menentukan keputusan terkait program studi dan

untuk menentukan strategi dalam meningkatkan jumlah siswi kedepannya. Kesimpulan dari penelitian ini terbentuk dua kelompok program studi yang paling diminati dan kelompok program studi yang kurang diminati. Hasil penelitian ini dapat digunakan sebagai evaluasi untuk meningkatkan jumlah siswi baru di setiap program studi[6].

Berdasarkan penelitian-penelitian terdahulu yang dipaparkan diatas, dapat ditarik kesimpulan bahwa metode *K-Means clustering* merupakan salah satu metode yang populer dan sering digunakan dalam banyak bidang, seperti kesehatan, bisnis, pendidikan, dan pemerintahan. Dengan cara ini, data dapat dikelompokkan menjadi beberapa kategori berdasarkan karakteristik yang sama. Berikut adalah beberapa kesimpulan umum dari penelitian-penelitian terdahulu tersebut:

- Data dapat dikelompokkan dengan metode clustering K-Means. dengan berbagai macam karakteristik, seperti data numerik, data kategorikal, dan data campuran.
- Hasil klasterisasi *K-Means clustering* dapat dipengaruhi oleh beberapa faktor, seperti jumlah cluster (k), jenis jarak (distance metric), dan jumlah iterasi.
- Metode *K-Means clustering* dapat digunakan untuk berbagai macam tujuan, seperti analisis data, pengambilan keputusan, dan pengembangan strategi bisnis.

Pada penelitian-penelitian terdahulu, metode *K-Means clustering* telah berhasil digunakan untuk berbagai macam tujuan. Hasil klasterisasi dapat digunakan untuk memberikan insight baru tentang data, membantu dalam pengambilan keputusan, dan mengembangkan strategi bisnis. Dari literatur yang telah disusun, ditarik kesimpulan bahwa adanya perbedaan yang signifikan dalam acuan-acuan yang digunakan dan tidak didapati mengenai klasterisasi dalam hal kegemaran membaca buku pada siswi kecuali pada jurnal "Peningkatan Tingkat Kegemaran Membaca (TGM) Siswi Melalui Pengukuran Indeks Gemar Membaca (IGM)" oleh [9] yang memiliki data, metode dan pendekatan yang berbeda. Kesimpulan ini menyoroti subjek penelitian, keragaman variabel data, tempat penelitian, pendekatan, dan fokus penelitian yang ada dalam literatur tersebut. Oleh karena itu, pemahaman mendalam terhadap perbedaan-perbedaan ini menjadi krusial untuk mengarahkan penelitian selanjutnya dan memastikan relevansi serta keberlanjutan pengembangan pengetahuan di bidang terkait.

2.1. Algoritma K-Means

Algoritma K-means menugaskan setiap titik ke cluster yang pusatnya juga disebut centroid terdekat. Pusat adalah rata-rata seluruh titik dalam cluster, yaitu koordinatnya merupakan rata-rata aritmatika untuk setiap dimensi secara terpisah atas semua titik dalam cluster[10].

2.2. RapidMiner Studio

RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi[11].

3. METODE PENELITIAN

Pengertian dari metode itu sendiri merupakan kerangka kerja dalam melaksanakan suatu tindakan atau kerangka berpikir yang berfungsi sebagai penyusun gagasan yang lebih terorganisir dan relevan dengan tujuan yang diinginkan[5]. Dorongan keingintahuan terhadap peristiwa di sekitar mendorong dilakukannya penelitian, yang bertujuan untuk menemukan jawaban terhadap permasalahan yang sedang terjadi[5]. Dalam bab ini metode penelitian yang dipilih untuk melakukan analisis data kegemaran membaca melalui metode algoritma clustering K-Means adalah penelitian kuantitatif. Pendekatan kuantitatif berarti pendekatan yang bersifat "menjumlahkan atau mengumpulkan". Dalam penelitian sosial, metode statistik merupakan representasi metode kuantitatif yang paling jelas, karena di dalam metode ini ada proses "kuantifikasi", yaitu proses memberi angka terhadap "kualitas" sesuatu hal[12]. Dimana penulis membutuhkan informasi tentang perilaku membaca siswi.

Subyek penelitian ini adalah siswi SMA Al Islam Boarding School Cirebon. Penelitian ini dimulai pada tanggal 14 Maret 2024 dengan lokasi penelitian di SMA Al Islam Boarding School beralamat di Jl.Kondangsari Kec.Beber Kab.Cirebon Jawa Barat. Populasi yang dimiliki sebesar 220 orang dan error margin 1%, maka jumlah sampel minimum yang diperlukan adalah:

$$n = 220 / (1 + (220 \times 0,01^2))$$

$$n = 220 / (1 + (220 \times 0,001))$$

$$n = 220 / (1 + 0,022)$$

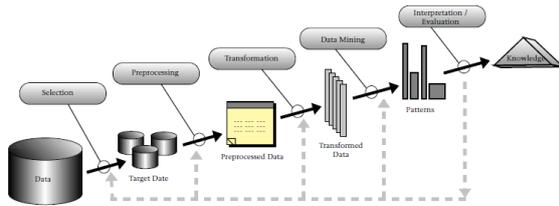
$$n = 220 / 1,022$$

$$n \approx 215,26$$

Jadi, jumlah sampel minimum yang diperlukan dari jumlah populasi (220 orang) adalah 215 orang.

4. HASIL DAN PEMBAHASAN

Proses Knowledge Discovery in Database (KDD) merupakan salah satu metode yang dapat diterapkan dalam melakukan data mining. mendefinisikan KDD sebagai proses menggunakan metode data mining untuk menemukan informasi berharga dan pola yang terdapat dalam data, melibatkan algoritma untuk mengidentifikasi pola tersebut. Dunham (2003) menggambarkan proses KDD dalam beberapa tahap, melibatkan seleksi data, pra-proses data, transformasi data, data mining, serta interpretasi dan evaluasi. Berikut adalah ilustrasi dan penjelasan mendetail mengenai proses KDD:



Gambar 1. KDD

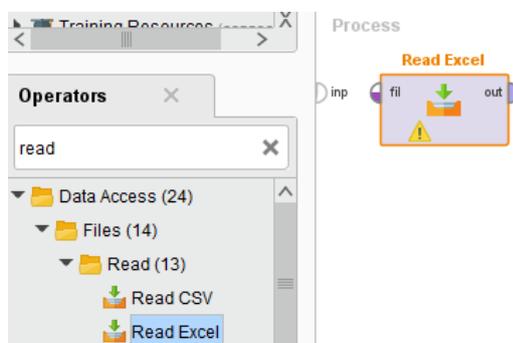
4.1. Pengumpulan Data

Dalam tahap ini, metode yang diterapkan dalam penelitian ini adalah metode clustering dengan menggunakan algoritma K-Means. Data diambil melalui survey kepada siswi SMA Al-Islam. Setelah data terkumpul langkah selanjutnya adalah melakukan pengolahan data dengan menggunakan metode clustering dengan tahapan awal seleksi data, preprocessing, transformasi dan data mining dengan menggunakan aplikasi RapidMiner. RapidMiner adalah perangkat lunak yang memanfaatkan prinsip dan algoritma pemrosesan data untuk mengolah data.

Table 1. Data Kegemaran Membaca

No	Atribut	Tipe data	Keterangan
1	Var01	nominal	Nama
2	Var02	Nominal	kelas
3	var03	Nominal	Jurusan
4	Var04	Integer	Jumlah buku yang dibaca dalam sebulan
5	Var05	Nominal	Genre buku
6	Var06	Nominal	Alasan membaca
7	Var07	integer	Skala kegemaran membaca
8	Var08	integer	Durasi membaca buku (menit)
9	Var09	Nominal	Kelompok

Data yang telah terkumpul akan dilanjutkan dengan proses pengolahan menggunakan aplikasi RapidMiner Studio. Sebelum masuk pada tahap seleksi data, data yang telah terkumpul diimpor terlebih dahulu ke dalam aplikasi RapidMiner Studio menggunakan operators read excel karena data yang digunakan berekstensi excel pada gambar berikut:



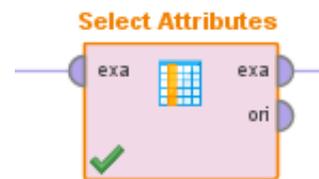
Gambar 2. operator read excel

Jika sudah muncul seperti tampilan diatas pilih import wizard data excel dan pilih cell range yang ingin diimpor lalu klik next untuk melanjutkan dan finish jika sudah selesai.

4.2. Data Selection

Sebelum tahap penggalian informasi dalam Knowledge Discovery in Database (KDD) dimulai, data yang dipilih dari sekumpulan data operasional harus dipilih dan disimpan dalam suatu berkas yang berbeda dari basis data operasional[13]. Pemilihan data dilakukan untuk menentukan dataset yang akan digunakan dalam proses data mining. Data yang dipilih kemudian disimpan terpisah dari basis data operasional.

Data kegemaran membaca buku melalui tahap penyaringan awal yang melibatkan ekstraksi beberapa atribut tertentu dari tabel guna keperluan analisis. Lima atribut yang diambil melibatkan informasi mengenai nama, jurusan, jumlah buku yang dibaca, durasi membaca buku, tingkat kegemaran membaca buku. Proses pengambilan data ini mengacu pada atribut yang terlihat pada gambar dibawah, dengan penyeleksian data menggunakan operators select attribute lalu pada parameter memilih filter type subset untuk dapat menyeleksi beberapa data dengan memilih exclude attributte.



Gambar 3. operator select attribute

Table 2. Hasil data selection

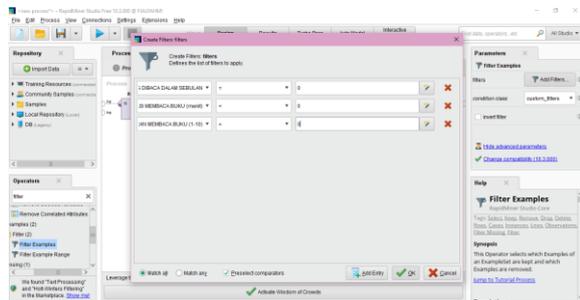
No	Atribut	Tipe data	Keterangan
1	Var01	nominal	Nama
2	Var02	Nominal	kelas
3	var03	Nominal	Jurusan
4	Var04	Integer	Jumlah buku yang dibaca dalam sebulan
7	Var07	integer	Skala kegemaran membaca
8	Var08	integer	Durasi membaca buku (menit)

Selanjutnya jika masih tersisa data yang tidak relevan dan akan mengganggu hasil akhir seperti data yang berbeda dari yang lainnya, maka dilakukan tahap filtering pada proses selanjutnya.

4.3. Preprocessing

Tahap pembersihan data bertujuan untuk meningkatkan kualitas data dengan menghapus duplikasi, memeriksa konsistensi, dan memperbaiki kesalahan. Proses enrichment juga dapat dilakukan untuk menambahkan data relevan. Sebelum melanjutkan dengan proses data mining, diperlukan

tahap pembersihan data pada fokus KDD. Tahap pembersihan ini melibatkan langkah-langkah seperti menghapus data yang terduplikasi, memeriksa inkonsistensi dalam data, dan memperbaiki kesalahan pada data, termasuk kesalahan cetak[14].

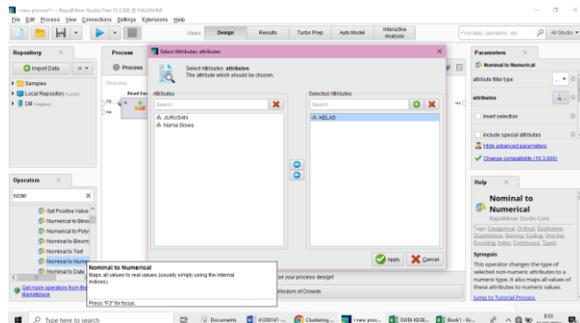


Gambar 4. proses filtering

Operators filter examples digunakan untuk memilih atribut atau baris tertentu dari sebuah dataset berdasarkan kriteria tertentu. Ini membantu menyaring data menjadi data yang relevan untuk analisis dan menghindari kerusakan *value* data. Pada tahap ini atribut tidak ada yang berkurang hanya saja jika terdapat nilai integer = 0 maka data akan terhapus secara otomatis.

4.4. Transformation

Pada tahapan ini, dilakukan proses transformasi data untuk memungkinkan pengolahan data menggunakan algoritma k-Means *clustering*. Data yang bisa diolah adalah data yang berupa *numerik* sehingga data yang berupa *non-numerik* harus diubah menjadi data yang berupa *numerik* agar bisa dilanjutkan ke dalam proses selanjutnya. Data yang bersifat *non-numeric* akan mengalami inisialisasi agar dapat diubah menjadi bentuk *numerik*.



Gambar 5. Transformasi data

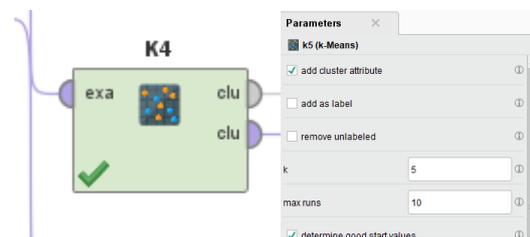
Table 3. Hasil transformasi data

No	Atribut	Tipe data	Keterangan
1	Var01	nominal	Nama
2	Var02	Nominal	kelas
3	var03	Integer	Jurusan
4	Var04	Integer	Jumlah buku yang dibaca dalam sebulan
7	Var07	integer	Skala kegemaran membaca
8	Var08	integer	Durasi membaca buku (menit)

Dengan demikian, pada karakteristik atribut kelas, ditemukan 2 data beserta frekuensinya dan inisial. Jurusan IPA setelah diinisialisasi menjadi 1 dan IPS menjadi 2.

4.5. Data Mining

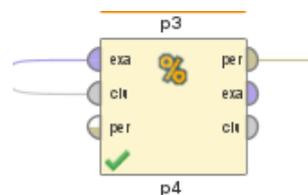
Data mining adalah tahap di mana teknik atau metode khusus digunakan untuk mengekstrak pola atau informasi menarik dari data terpilih. Pemilihan metode bergantung pada tujuan dan keseluruhan proses KDD. Data mining adalah analisis peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dari sebelumnya sehingga pemilik data dapat memahami dan menggunakannya(Larose,2005) dalam [11].



Gambar 6. operator K-Means Clustering

Hasil pengolahan data kegemaran membaca buku pada penelitian ini menggunakan bantuan *software* RapidMiner studio dengan algoritma *clustering* k-means. Pada parameters K-Means terdapat k sebagai kluster yang berbeda. K2 dengan jumlah k=2, k3 dengan jumlah k=3, k4 dengan jumlah k=4 dan k5 dengan jumlah k=5 dengan masing-masing *max runs* 10. Parameter ini mendefinisikan jumlah kali k-Means *algorithm* dijalankan dengan nilai inisialisasi yang berbeda.

Setelah melakukan klusterisasi, untuk melihat daftar nilai kriteria kinerja berdasarkan centroid kluster, dibutuhkan *operators performance*. Operator ini digunakan untuk mengevaluasi kinerja model clustering berbasis centroid seperti K-Means dan K-Medoids.



Gambar 7. operator performance

4.6. Evaluasi Menggunakan Davies Bouldin Index

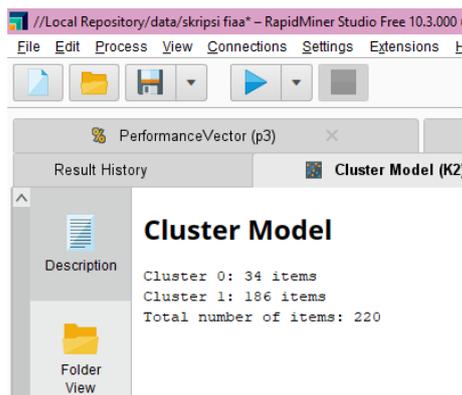
Dalam menentukan jumlah cluster optimal pada RapidMiner, dapat dilakukan dengan memanfaatkan Index Davies Bouldin (IDB). Metode ini merupakan salah satu metode evaluasi internal untuk mengukur kualitas cluster dengan memperhatikan nilai terendah. Penilaian ini didasarkan pada kesejajaran (kohesi) dan pemisahan (separasi) dalam cluster-cluster tersebut[15]. Semakin kecil nilai Davies-Bouldin Index akan memberikan hasil yang baik[16]. Berikut

ini disajikan hasil perhitungan nilai Davies- Bouldin Index untuk jumlah masing-masing klaster.

Tabel 2. Nilai *Davies Bouldin Index*

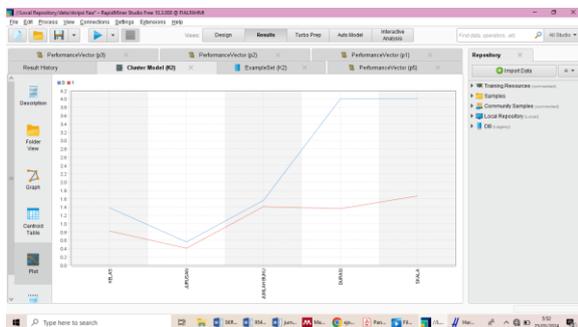
K	DBI
2	0.927
3	1.131
4	1.073
5	0.970

Pada tabel diatas diketahui nilai DBI terkecil berada pada k=2 dengan nilai DBI 0.970 dengan hasil cluster sebagai berikut:



Gambar 8. hasil cluster optimal

Diperoleh klasterisasi optimal pada cluster model (k2) dengan jumlah cluster 0 sebanyak 34 items dan cluster 1 sebanyak 186 items, yang dijabarkan sebagai berikut.



Gambar 9.karakteristik pada tiap cluster

Dari gambar diatas dapat ditarik kesimpulan hasil klasterisasi data kegemaran membaca buku pada siswi SMA Al-Islam Boarding School Cirebon seperti pada berikut:

- a. (Cluster_0) Pengelompokan berdasarkan Intensitas Membaca:
 - Kelompok dengan intensitas membaca tinggi: terdiri dari siswi yang membaca banyak buku dalam satu bulan dan memiliki durasi membaca yang lama. Kelompok ini adalah siswi yang sangat antusias terhadap membaca dan seringkali membaca dengan tujuan belajar atau peningkatan pengetahuan. Kelompok ini cenderung menggunakan buku sebagai sumber

informasi utama dan memiliki kebiasaan membaca yang intensif.

- Kelompok pembaca serius: memiliki intensitas membaca yang tinggi dan membaca dengan tujuan belajar atau peningkatan pengetahuan. Kelompok ini memanfaatkan buku sebagai sarana untuk mengembangkan diri dan memperoleh informasi baru.
- b. (Cluster_1)
 - Kelompok dengan intensitas membaca rendah: terdiri dari siswi yang membaca sedikit buku dalam satu bulan dan memiliki durasi membaca yang singkat. Cenderung membaca hanya sesekali atau hanya jika ada kebutuhan tertentu. Siswi dalam kelompok ini mungkin memiliki keterbatasan waktu atau minat terhadap membaca. Buku mungkin tidak menjadi prioritas utama dalam kehidupan mereka.
 - Kelompok pembaca hobi: terdiri dari siswi yang memiliki intensitas membaca sedang. Kelompok ini membaca buku sebagai bentuk hiburan dan kesenangan pribadi. Aktivitas membaca ini tidak hanya meningkatkan pengetahuan, tetapi juga memberikan kepuasan secara emosional dan intelektual.

5. KESIMPULAN DAN SARAN

Kesimpulan pada penelitian ini adalah merujuk kepada tujuan penelitian, dan menjawab dari rumusan masalah. Berdasarkan penelitian klastering data kegemaran membaca menggunakan algoritma K-Means Clustering, ditemukan bahwa nilai indeks Davies-Bouldin pada k=2 sebesar 0.970.

Dengan karakteristik tertentu yaitu kelompok cluster pertama siswi dengan intensitas membaca tinggi dan kelompok pembaca serius, kelompok cluster kedua merupakan kelompok siswi dengan intensitas membaca rendah dan menjadikan membaca buku hanya sekedar hobi atau kesenangan saja.

DAFTAR PUSTAKA

- [1] Shofaussamawati, “Menumbuhkan Minat Baca dengan Pengenalan Perpustakaan pada Anak Sejak Dini,” *J. Perpust. Libr.*, vol. 2, no. 1, hal. 46–59, 2018.
- [2] Ruslan & Wibayanti, “Pentingnya Meningkatkan Minat Baca Siswa,” *Pros. Semin. Nas. Pendidik. Progr. Pascasarj. Univ. Pgrri Palembang*, hal. 767–775, 2019, [Daring]. Tersedia pada: www.perpusnas.go.id
- [3] D. E. Prakosa, “Laporan Kuliah Kerja Nyata Institut Seni Indonesia Surakarta,” hal. 42, 2020.
- [4] H. Nopriandi dan F. Haswan, “Analisis Klasterisasi Mahasiswa Baru dalam Memilih Program Studi dengan Menggunakan Algoritma K-Means,” *J. Inf. Syst. Res.*, vol. 3, no. 4, hal. 666–671, 2022, doi: 10.47065/josh.v3i4.1986.
- [5] P. Zainal A. Hasibuan, *METODOLOGI PENELITIAN PADA BIDANG ILMU KOMPUTER DAN TEKNOLOGI INFORMASI*,

- vol. 20, no. 1. 2007. doi: 10.55601/jsm.v20i1.649.
- [6] M. B. Fajri dan S. D. Purnamasari, "Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Means Clustering," *J. Inf. Technol. Ampera*, vol. 3, no. 3, hal. 317–334, 2022, [Daring]. Tersedia pada: <https://journal-computing.org/index.php/journal-ita/index>
- [7] M. Rizki dan M. Mulyawan, "PENERAPAN METODE K-MEANS CLUSTERING PADA DATA PENJUALAN OPTIK CHANTIKA," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, hal. 1303–1307, 2023, doi: 10.36040/jati.v7i2.6562.
- [8] M. Djaka Permana, A. Lia Hananto, E. Novalia, B. Huda, dan T. Paryono, "Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means," *J. KomtekInfo*, vol. 10, hal. 15–20, 2023, doi: 10.35134/komtekinfo.v10i1.332.
- [9] E. Fat, "Peningkatan Tingkat Kegemaran Membaca (TGM) Masyarakat Melalui Pengukuran Indeks Gemar Membaca (IGM)," *Tik Ilmeu J. Ilmu Perpust. dan Inf.*, vol. 6, no. 2, hal. 341, 2022, doi: 10.29240/tik.v6i2.5498.
- [10] M. G. H. Omran, A. P. Engelbrecht, dan A. Salman, "An overview of clustering methods," *Intell. Data Anal.*, vol. 11, no. 6, hal. 583–605, 2007, doi: 10.3233/ida-2007-11602.
- [11] Aprilla Dennis, "Belajar Data Mining dengan RapidMiner," *Innov. Knowl. Manag. Bus. Glob. Theory Pract. Vols 1 2*, vol. 5, no. 4, hal. 1–5, 2013, [Daring]. Tersedia pada: <http://esjournals.org/journaloftechnology/archiv>
- e/vol1no6/vol1no6_6.pdf%5Cnhttp://www.aircse.org/journal/nsa/5413nsa02.pdf
- [12] S. K. Donatus, "Pendekatan Kuantitatif dan Kualitatif Dalam Penelitian Ilmu Sosial," *JournalStudia Philos. Theol.*, vol. 16, no. 2, hal. 197–210, 2016.
- [13] Yuli Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, hal. 213–219, 2019.
- [14] F. Nasari dan S. Darma, "Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama)," *STMIK AMIKOM Yogyakarta*, hal. 6–8, 2015.
- [15] Y. Sopyan, A. D. Lesmana, dan C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, hal. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.
- [16] A. Badruttamam, S. Sudarno, dan D. A. I. Maruddani, "PENERAPAN ANALISIS KLASSTER K-MODES DENGAN VALIDASI DAVIES BOULDIN INDEX DALAM MENENTUKAN KARAKTERISTIK KANAL YOUTUBE DI INDONESIA (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade)," *J. Gaussian*, vol. 9, no. 3, hal. 263–272, 2020, doi: 10.14710/j.gauss.v9i3.28907.