

BAB II

LANDASAN TEORI

2.1. Minat Baca

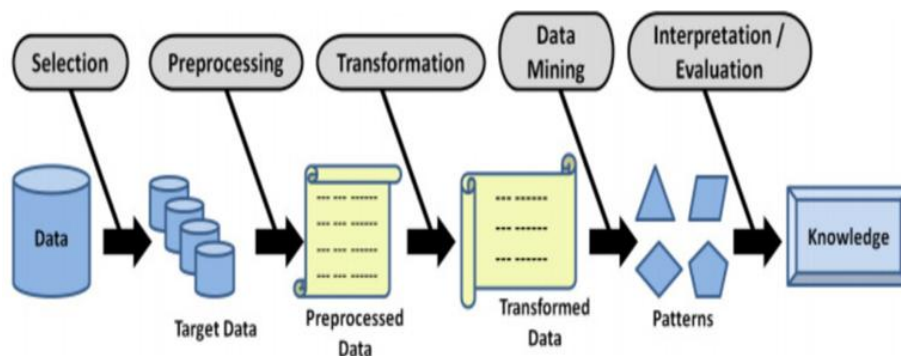
Minat baca merupakan kesadaran individu untuk membaca yang berasal dari dorongan diri masing-masing yang didukung dengan lingkungan [1]. Anak yang membaca dengan minat akan lebih memahami bacaan yang sedang dibaca, karena anak akan membaca dengan sepenuh hati. Agar siswa dapat mengetahui makna bacaan, dibutuhkan minat yang baik dalam membaca. Minat baca adalah kecenderungan jiwa seseorang secara mendalam yang ditandai dengan perasaan senang serta berkeinginan kuat untuk membaca tanpa adanya paksaan [2]. Berdasarkan pengertian diatas, minat baca dapat diartikan sebagai ketertarikan untuk membaca terhadap suatu hal dan menaruh perhatian pada suatu pembelajaran tertentu dan disertai harapan untuk mengetahui, mempelajari, dan membuktikannya melalui partisipasi aktif juga keinginan besar untuk membaca. Kemauan juga menjadi faktor penting dalam mengembangkan kesadaran siswa akan pentingnya keinginan membaca serta tercapainya tujuan dan hasil yang diinginkan oleh pembaca.

Minat baca tidak hadir langsung ketika kita dilahirkan di dunia. Minat baca ada dan menjadi bagian didalam diri karena adanya usaha didalam diri dan faktor lingkungan yang juga mendukung. Minat baca merupakan faktor yang penting untuk meningkatkan keaksaraan atau penambahan kosa kata siswa, selain itu minat baca membantu siswa untuk sukses disekolah. Hal ini menyadarkan kita betapa pentingnya minat baca [3].

Perpustakaan sekolah menjadi sarana untuk menumbuhkan minat baca siswa sebagai sumber informasi dan menyediakan bahan bacaan serta fasilitas yang memadai. Perpustakaan sekolah digunakan sebagai sarana dan prasarana yang diharapkan dapat mendukung pembelajaran sekolah yang mendorong terwujudnya pendidikan yang berkualitas untuk mencerdaskan kehidupan bangsa. Perpustakaan sekolah dapat berperan dalam memajukan siswa melalui ilmu pengetahuan dan informasi, sehingga perpustakaan dapat dijadikan sebagai sumber informasi bagi guru dan siswa melalui kegiatan membaca [4].

2.2. *Knowledge Discovery in Database*

Data mining dan *Knowledge Discovery in Database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. *Knowledge Discovery Data* (KDD) merupakan sejumlah data yang besar yang diekstrak melalui tahapan penggalian dan analisis untuk mendapatkan informasi dan pengetahuan yang berguna (Marisa et al, 2021) [5]. Data mining adalah salah satu langkah dalam proses KDD secara keseluruhan. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis detail dapat dijelaskan pada Gambar 2.1.



Gambar 2. 1. Proses *Knowledge Discovery in Database* (KDD) [6]

Keterangan Gambar:

1. *Selection*

Pemilihan (*selection*) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining disimpan dalam suatu berkas, terpisah dari basis data operasional. *Selection* digunakan untuk menentukan variabel yang akan diambil agar tidak ada kesamaan dan terjadi perulangan yang tidak diperlukan dalam pengolahan *data mining*.

2. *Preprocessing*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Pada *preprocessing* terdapat dua tahapan, yaitu sebagai berikut:

a. *Data Cleaning*

Menghilangkan data yang tidak diperlukan seperti menangani *missing value*, *noise* data serta menangani data-data yang tidak konsisten dan relevan.

b. *Data Integration*

Dilakukan terhadap atribut yang mengidentifikasi entitas yang unik.

3. *Transformation*

Merubah data sesuai format *ekstention* yang sesuai dalam pengolahan data mining karena beberapa metode pada data *mining* memerlukan format khusus sebelum dapat diproses pada data mining. *Coding* adalah

transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

Proses utama pada metode yang diterapkan untuk mendapatkan pengetahuan baru dari data yang diproses.

5. *Evaluation/ Interpretation*

Mengidentifikasi pola-pola yang menarik kedalam *knowledgebase* yang diidentifikasi. Pada tahap ini, menghasilkan pola-pola khas maupun model prediksi yang dievaluasi untuk menilai kajian yang ada sudah memenuhi target yang diinginkan.

6. *Knowledge*

Pola-pola yang dihasilkan akan dipresentasikan kepada pengguna. Pada tahapan ini pengetahuan baru yang dihasilkan bisa dipahami semua orang yang akan dijadikan acuan pengambilan keputusan.

KDD memiliki 4 fungsi dasar yang biasa kita temui yaitu [7]:

1. Fungsi Klasifikasi, fungsi ini merupakan fungsi untuk menemukan model/fungsi untuk menggambarkan suatu *class* dari data. Dengan klasifikasi mampu mempelajari data sehingga mampu meramalkan kecenderungan data.
2. Fungsi Prediksi, fungsi ini digunakan untuk menemukan pola data menggunakan variable dalam memprediksi variabel lain yang belum diketahui jenisnya.

3. Fungsi Deskripsi, fungsi ini digunakan untuk menemukan karakteristik data.

Fungsi Asosiasi, fungsi ini digunakan untuk menemukan hubungan yang ada pada nilai atribut dari banyaknya data.

2.3. Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Database* (KDD) [8]. Data mining dapat membantu dalam memprediksi hasil, seperti dalam penentuan penerima program bantuan pemerintah [9]. Selain itu, teknik data mining, seperti algoritma Apriori, dapat digunakan untuk menganalisis data transaksi guna mendukung strategi penjualan [10]. Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Proses data mining terdiri dari pengumpulan data, ekstraksi data, analisa data, dan statistik data. Data Mining juga umum dikenal sebagai *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lainnya [11].

Proses data mining melibatkan beberapa langkah utama yang saling terkait, dimulai dari pemrosesan data hingga analisis hasil. Langkah-langkah tersebut adalah sebagai berikut :

1. Langkah pertama adalah pengumpulan data, di mana data yang relevan dikumpulkan dari berbagai sumber.

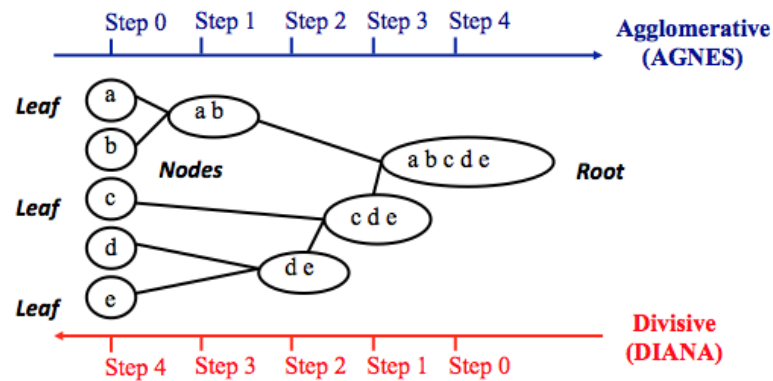
2. Selanjutnya, tahap pemrosesan data (*preprocessing*) dilakukan untuk membersihkan dan menyiapkan data, termasuk penghapusan noise dan pengisian nilai yang hilang
3. Setelah data siap, teknik data mining seperti klasifikasi, asosiasi, dan klustering diterapkan untuk mengekstrak pola dan informasi yang berguna
4. Setelah penerapan teknik mining, hasil yang diperoleh perlu dievaluasi dan dianalisis untuk memastikan keakuratan dan relevansinya
5. Langkah terakhir adalah penyajian hasil, di mana informasi yang diperoleh disajikan dalam format yang mudah dipahami untuk pengambilan keputusan lebih lanjut
6. Proses ini mencerminkan kompleksitas dan interkoneksi antara berbagai langkah dalam data mining, yang merupakan bagian integral dari penemuan pengetahuan dalam basis data.

2.4. Clustering

Metode *Clustering* adalah metode pengelompokan yang bersifat *unsupervised* atau tanpa arahan (yaitu metode dilakukan tanpa adanya *training* atau latihan) serta tidak memerlukan target atau sasaran luaran [12]. Clustering adalah salah satu teknik data mining yang memungkinkan untuk mengidentifikasi kelompok yang dihasilkan dari pengelompokan elemen yang lebih kecil berdasarkan kemiripan satu sama lain [13]. Metode *clustering* secara umum dapat dibagi menjadi dua yaitu *hierarchical clustering* dan *partitional clustering*. Selain itu, terdapat pula metode Density-Based (berbasis kepadatan) dan Grid-Based (berbasis grid) yang digunakan dalam implementasi clustering. Berikut penjelasannya:

1. Hierarchical Clustering

Hierarchical clustering adalah pengelompokan data melalui diagram yang berbentuk hirarki, dimana setiap iterasinya menggabungkan dua grup yang terdekat atau membagi seluruh dataset ke dalam *cluster*.

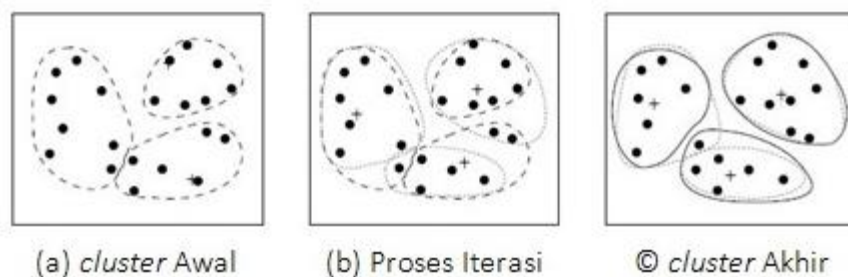


Gambar 2. 2. *Algoritma Hierarchical Clustering* [14]

Contoh metode *hierarchy clustering* adalah: Single Linkage, Complete Linkage, Average Linkage, dan Average Group Linkage.

2. Partitional Clustering

Partitional clustering yaitu data dikelompokkan ke dalam sejumlah cluster tanpa adanya struktur hirarki antara satu dengan yang lainnya. Pada metode partitional clustering, setiap cluster memiliki titik pusat cluster (centroid) dan metode ini memiliki tujuan meminimumkan jarak (dissimilarity) dari seluruh data ke pusat cluster. Contoh metode partitional clustering: K-Means, Fuzzy K-means dan Mixture Modelling.



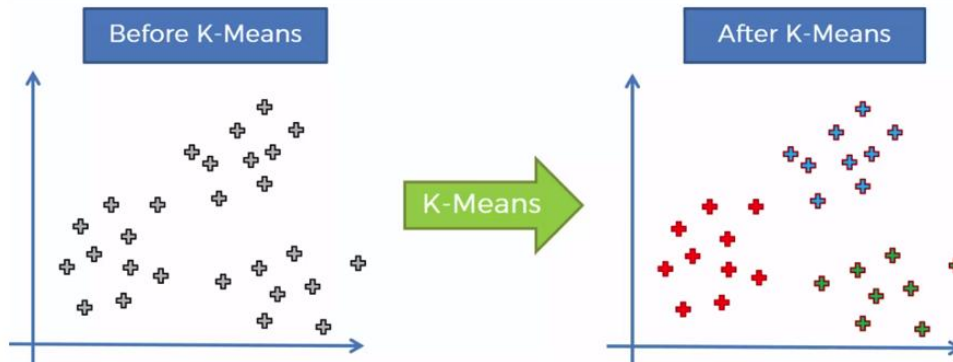
Gambar 2. 3. *Partitional Clustering* [15]

2.5. Algoritma K-Means Clustering

Algoritma K-Means adalah salah satu teknik klasterisasi paling populer yang digunakan dalam analisis data [16]. K-Means adalah metode klasterisasi yang umum digunakan dalam analisis data untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan kedekatan data dengan centroid [17]. Metode ini efektif dalam mengelompokkan data berdasarkan parameter tertentu. Clustering merujuk pada pengelompokan dokumen, observasi atau kasus pada kelas yang objeknya mirip [18]. Klaster adalah kumpulan dokumen yang mirip satu sama lain dan berbeda dengan dokumen pada klaster lain. Clustering berbeda dengan Clasification, pada Clustering tidak ada target variabel untuk dikelompokkan. Algoritma Clustering mencoba untuk membagi kumpulan data menjadi klaster yang anggotanya relatif sama, dimana kemiripan dokumen di klaster yang sama tinggi, dan kemiripan dokumen di klaster lain kecil. Clustering dianggap sebagai metode pembelajaran tanpa pengawasan yang paling penting, di mana masalah seperti itu adalah tentang menemukan pola dalam kumpulan data yang tidak berlabel. Cluster Clustering membagi kumpulan data menjadi beberapa kelompok dimana kesamaan pada suatu kelompok tertentu lebih besar daripada pada kelompok lainnya [19].

Penggunaan algoritma pengelompokan tergantung pada jenis data yang tersedia untuk tujuan dan aplikasi tertentu. Jika analisis klaster digunakan sebagai alat deskriptif atau eksplorasi, beberapa algoritma dapat dicoba pada data yang sama untuk mendapatkan apa yang diungkapkan oleh data tersebut. Secara umum metode Clustering dapat diklasifikasikan menjadi beberapa kategori, salah satunya adalah kategori metode partisi. Metode partisi ini didasarkan pada awalnya menentukan jumlah grup dan kemudian menetapkan kembali objek secara iteratif untuk menemukan grup yang terletak di suatu titik. Salah satu algoritma yang

populer dalam penerapan metode partisi ini adalah algoritma K-Means dan algoritma K-Medoids [19].



Gambar 2. 4. *Algoritma K-Means Clustering [20]*

K-Means Clustering dapat diterapkan dalam konteks medis, seperti dalam pengelompokan pasien hipertensi berdasarkan karakteristik tertentu, di mana evaluasi jumlah cluster yang optimal dilakukan menggunakan Davies Bouldin Index (DBI) Qirom [21]. Hal ini menunjukkan bahwa K-Means tidak hanya efektif dalam pengelompokan, tetapi juga dalam memberikan wawasan yang lebih dalam tentang data yang dianalisis. K-Means adalah metode pembelajaran tanpa pengawasan yang digunakan untuk mengelompokkan data yang belum dilabel ke dalam cluster yang berbeda [22]. K-Means Clustering dalam kerangka KDD merupakan alat yang sangat berguna untuk menganalisis dan mengelompokkan data dalam berbagai bidang. Dengan pendekatan yang sistematis, peneliti dapat mengidentifikasi pola dan hubungan dalam data yang kompleks, yang pada gilirannya dapat mendukung pengambilan keputusan yang lebih baik.

Metode K-Means memiliki keunggulan yang signifikan dalam pengelompokan data, termasuk kemudahan implementasi dan kecepatan dalam proses klusterisasi. K-Means efektif dalam mengelompokkan data yang memiliki struktur yang jelas dan dapat digunakan dalam berbagai aplikasi, seperti pemasaran

dan analisis data sosial [23]. Namun, metode ini juga memiliki keterbatasan, terutama terkait dengan pemilihan titik awal centroid yang dapat mempengaruhi hasil klusterisasi dan berpotensi menyebabkan konvergensi pada solusi lokal. Selain itu, K-Means kurang efektif dalam menangani data dengan distribusi yang tidak teratur atau adanya outlier, yang dapat menyebabkan hasil kluster yang tidak representative. Oleh karena itu, meskipun K-Means merupakan alat yang berguna dalam analisis data, penting untuk mempertimbangkan konteks dan karakteristik data sebelum penerapannya. Kelebihan K-means adalah kemampuannya untuk menangani dataset besar dan kompleks dengan efisiensi yang tinggi. Namun, metode ini juga memiliki kelemahan, seperti sensitivitas terhadap pemilihan titik awal centroid dan ketidakmampuan untuk menangani cluster dengan bentuk yang tidak bulat [24]. Oleh karena itu, pemahaman yang mendalam tentang karakteristik data sangat penting untuk penerapan K-means yang efektif.

Tahapan data mining yang mendukung implementasi metode K-Means meliputi beberapa langkah penting, yaitu pengumpulan data, pra-pemrosesan, pemilihan fitur, pengelompokan, dan evaluasi hasil. Pengumpulan data dilakukan untuk memperoleh informasi yang relevan dari sumber yang beragam, seperti data penjualan atau data Kesehatan. Selanjutnya, pra-pemrosesan data penting untuk membersihkan dan menyiapkan data agar bebas dari noise dan konsisten, yang merupakan langkah krusial sebelum pengelompokan. Setelah data siap, pemilihan fitur dilakukan untuk menentukan atribut yang paling relevan untuk analisis. Metode K-Means kemudian diterapkan untuk mengelompokkan data berdasarkan kesamaan karakteristik, seperti dalam pengelompokan pelanggan atau produk. Terakhir, evaluasi hasil clustering penting untuk menilai efektivitas pengelompokan

yang dilakukan, seringkali menggunakan metrik seperti silhouette score atau elbow method untuk menentukan jumlah cluster yang optimal. Dengan demikian, tahapan ini sangat mendukung implementasi metode K-Means dalam berbagai konteks analisis data.

2.6. Alat Bantu Analisis

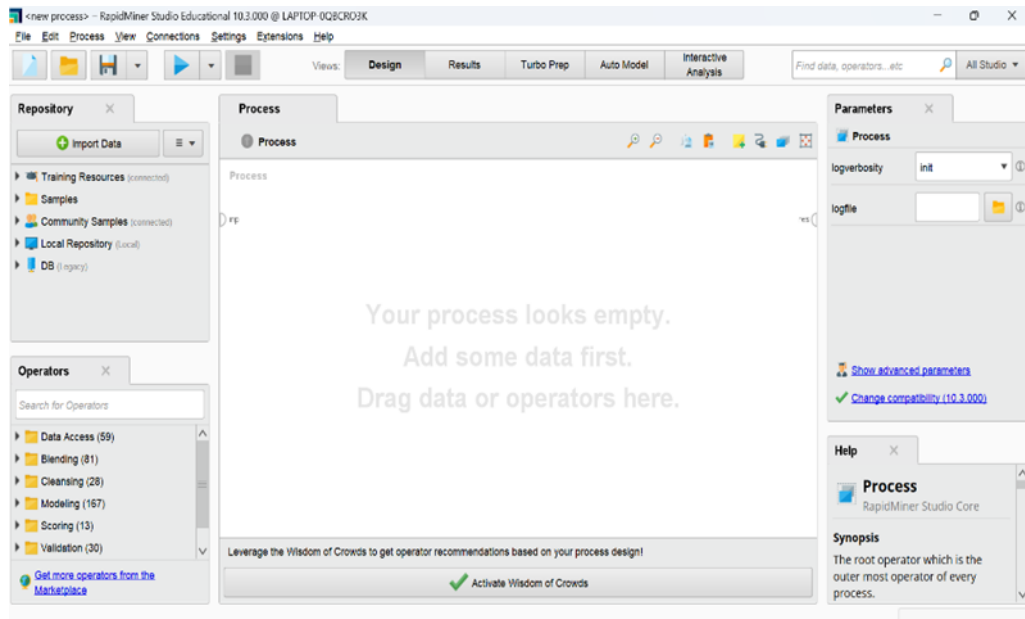
2.6.1. RapidMiner

RapidMiner adalah sebuah *software* untuk pengolahan data mining. RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. Pekerjaan yang dilakukan oleh rapidminer *text mining* adalah berkisar dengan analisis teks, mengekstrak pola-pola dari dataset yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan *database* [25]. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik [26]. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang diintegrasikan pada produknya sendiri. RapidMiner ditulis menggunakan bahasa java sehingga dapat bekerja disemua sistem operasi

RapidMiner adalah *platform* perangkat lunak yang digunakan untuk analisis data dan pembelajaran mesin. Menurut para ahli, RapidMiner memungkinkan pengguna untuk melakukan eksplorasi data, pemodelan, dan evaluasi hasil dengan antarmuka yang intuitif, sehingga memudahkan pengguna yang tidak memiliki latar belakang teknis untuk menerapkan teknik analisis data yang kompleks. RapidMiner juga mendukung berbagai metode analisis, termasuk analisis prediktif dan analisis

statistik, yang dapat diintegrasikan dengan berbagai sumber data. Lebih lanjut, RapidMiner dikenal karena kemampuannya dalam mengotomatisasi proses analisis data, yang membantu dalam pengambilan keputusan berbasis data. Dengan demikian, RapidMiner menjadi alat yang sangat berharga dalam berbagai bidang, termasuk bisnis, kesehatan, dan penelitian akademis, karena dapat meningkatkan efisiensi dan efektivitas analisis data.

RapidMiner adalah perangkat lunak yang digunakan untuk analisis data dan pengolahan data mining. Perangkat ini memungkinkan pengguna untuk melakukan analisis teks, ekstraksi pola dari dataset, dan menggabungkan metode statistik serta kecerdasan buatan untuk menghasilkan informasi yang berkualitas tinggi dari data yang diolah. RapidMiner juga mendukung berbagai algoritma *machine learning*, termasuk algoritma klasifikasi seperti *Naïve Bayes* dan *Decision Tree*, yang sering digunakan dalam penelitian untuk memprediksi dan mengklasifikasikan data. Keunggulan RapidMiner terletak pada antarmuka grafisnya yang memudahkan pengguna dalam merancang alur kerja analisis. Selain itu, RapidMiner dapat digunakan dalam berbagai aplikasi, mulai dari analisis sentimen di media sosial hingga pengembangan refrigeran ramah lingkungan.



Gambar 2. 5. Tampilan RapidMiner Versi 10.3

2.7. Penelitian Terdahulu

Penelitian terdahulu adalah bagian penting yang berfungsi untuk mengulas dan menganalisis hasil-hasil penelitian yang sudah dilakukan sebelumnya yang relevan dengan topik penelitian ini. Beberapa penelitian terdahulu yang berasal dari artikel pada jurnal akan disajikan dan diulas pada bagian ini.

Tabel 2. 1. Penelitian Terdahulu

Referensi Penelitian	1
Judul	Analisis Clustering Buku Sebagai Upaya Untuk Meningkatkan Minat Baca Siswa Pada Perpustakaan SMA Negeri 3 Pekanbaru
Nama Penulis	Wahyuni Melati Putri, Elvira Asril
Tahun	2023
Data	<ul style="list-style-type: none"> Jenis Data: Data peminjaman buku di perpustakaan SMA Negeri 3 Pekanbaru pada tahun 2022.

	<ul style="list-style-type: none"> • Bentuk Data: Hardfile data peminjaman buku yang disusun berdasarkan kelas. • Jumlah Data: Total ada 153 data dengan atribut sebanyak 5 jenis. • Atribut Data: Jenis buku seperti Buku Umum A, Buku Umum B, Buku Peminatan IPA, Buku Peminatan IPS, dan Buku Fiksi.
Hasil	<p>Algoritma K-Means efektif dalam mengelompokkan jenis buku berdasarkan pola peminjaman sehingga dapat membantu meningkatkan minat baca siswa melalui manajemen koleksi perpustakaan yang lebih baik. Perhitungan manual dibandingkan dengan hasil dari software RapidMiner menunjukkan perbedaan kecil kemungkinan karena pengambilan sampel acak saat pengujian tetapi secara umum mendukung hasil clustering tersebut.</p>

Referensi Penelitian	2
Judul	Penerapan Metode K-Means Dalam Pengelompokkan Buku Untuk Menentukan Minat Baca Pada Perpustakaan Daerah Kota Medan
Nama Penulis	Misael Oktavianda Harefa, Soeb Aripin
Tahun	2024

Data	<ul style="list-style-type: none"> • Jenis Data: Data kategori buku di Perpustakaan Daerah Kota Medan. • Jumlah Sampel: Sebanyak 100 sampel data kategori buku. • Atribut Data: Jumlah peminjam buku (K1), Stok buku tersedia (K2) • Kategori Buku: Beragam kategori seperti Publikasi Umum, Bibliografi, Ensiklopedia, Biografi, Majalah dan Jurnal, Media Massa dan lain-lain.
Hasil	Penggunaan metode K-Means clustering sangat membantu dalam pengklasifikasian jenis buku sesuai tingkat minat baca masyarakat di perpustakaan daerah Kota Medan dengan hasil yang valid setelah dua kali iterasi perhitungan centroid baru.

Referensi Penelitian	3
Judul	Klasterisasi Data Kegemaran Membaca Menggunakan Algoritma K-Means di SMA Al-Islam Cirebon
Nama Penulis	Luthfia Fahmi Az Zahroh, Nining Rahaningsih, Raditya Danar Dana
Tahun	2024
Data	<ul style="list-style-type: none"> • Jenis Data: Data kegemaran membaca buku siswi SMA Al-Islam Boarding School Cirebon.

	<ul style="list-style-type: none"> • Jumlah Sampel: Sekitar 215 siswi dari populasi 220 orang dengan margin error 1%. • Atribut Data: Nama (nominal), Kelas (nominal) Jurusan (integer setelah inisialisasi; IPA=1, IPS=2), Jumlah buku yang dibaca dalam sebulan (integer), Genre buku (nominal), Alasan membaca (nominal), Skala kegemaran membaca (integer), Durasi membaca buku dalam menit (integer). • Metode Pengumpulan: Survei langsung kepada siswa. • Pengolahan Data: Menggunakan aplikasi RapidMiner Studio untuk proses seleksi data, preprocessing, transformasi data menjadi numerik agar dapat diproses algoritma K-Means..
<p>Hasil</p>	<p>Algoritma K-Means efektif digunakan untuk mengelompokkan pola kegemaran membaca siswa berdasarkan atribut kuantitatif seperti jumlah bacaan dan durasi waktu baca sehingga memberikan insight penting bagi pengembangan program literasi di sekolah. Berdasarkan nilai Davies-Bouldin Index terkecil yaitu pada k=2 dengan nilai DBI =0.970.</p>

Referensi Penelitian	4
Judul	Analisis Pengaruh Jenis Buku Terhadap Minat Baca Mahasiswa di Perpustakaan Ibrahimy dengan Algoritma K-Means Clustering
Nama Penulis	Mahmudi, Zaehol Fatah
Tahun	2025
Data	<ul style="list-style-type: none"> • Jenis Data: Dataset berupa data peminjaman buku oleh mahasiswa di Perpustakaan Ibrahimy. • Atribut Data: Jenis buku (kategori seperti fiksi, non-fiksi, referensi, akademik), Frekuensi membaca (berapa sering mahasiswa meminjam/membaca buku), Durasi membaca (rata-rata waktu membaca per sesi), Motivasi membaca (alasan mahasiswa membaca buku), Umur mahasiswa. • Jumlah Sampel: Survei melibatkan sekitar 100 mahasiswa. • Metode Pengumpulan Data: Kuesioner, observasi, dan wawancara dengan mahasiswa, dosen, serta pustakawan.
Hasil	Penelitian ini membuktikan bahwa algoritma K-Means Clustering dapat digunakan secara efektif untuk menganalisis pola minat baca berdasarkan jenis buku di kalangan mahasiswa Perpustakaan Ibrahimy. Mean Accuracy mencapai sekitar 92.8%, menunjukkan bahwa

	<p>model clustering efektif dalam mengelompokkan minat baca berdasarkan jenis buku. Standard Deviation rendah yaitu sekitar 0.0373, menandakan stabilitas hasil antar pengujian.</p>
--	--

Referensi Penelitian	5
Judul	Implementasi Algoritma K-Means untuk Mengetahui Minat Siswa SMA Terhadap Mata Pelajaran Teknologi Informasi dan Komunikasi
Nama Penulis	Widya Indah Permatasari, Vitri Tundjungsari
Tahun	2025
Data	<ul style="list-style-type: none"> • Jenis Data: Tanggapan siswa kelas X semester akhir di SMAN 1 Tarumajaya terhadap minat mereka pada mata pelajaran Teknologi Informasi dan Komunikasi (TIK). • Jumlah Sampel: Dari total populasi 431 siswa kelas X, diperoleh data sebanyak 214 siswa melalui penyebaran Google Form. • Teknik Sampling: Teknik Slovin digunakan untuk menentukan ukuran sampel minimum yaitu sebanyak 207 siswa; data aktual melebihi jumlah ini sehingga valid.

	<ul style="list-style-type: none"> • Bentuk Data: Data berupa jawaban kuesioner dengan kategori minat atau tidak minat terhadap beberapa pertanyaan terkait TIK.
Hasil	<p>Algoritma K-Means efektif dalam mengelompokkan tingkat minat siswa SMA pada mata pelajaran TIK secara sistematis. Mayoritas siswa memiliki tingkat ketertarikan yang cukup tinggi dibandingkan kelompok yang kurang berminat. Visualisasi menunjukkan distribusi dua cluster tersebut berdasarkan koordinat nilai pada sumbu X (minat teori) dan Y (minat kegiatan). Metode Elbow digunakan untuk menentukan jumlah cluster optimal; hasil menunjukkan k=2 adalah pilihan terbaik karena penurunan Within-cluster Sum of Squares paling signifikan sebelum grafik melandai.</p>

Referensi Penelitian	6
Judul	Implementasi Algoritma K-Means untuk Mengelompokkan Pola Membaca Siswa di Era Digital
Nama Penulis	Yongky Permana Putra, Reflan Nuari
Tahun	2025
Data	<ul style="list-style-type: none"> • Data pola membaca siswa SMK Negeri 4 Bandar Lampung. • Metode Pengumpulan Data: Kuesioner disebarkan kepada siswa kelas 10 dan 11 dari 10 jurusan.

	<ul style="list-style-type: none"> • Jumlah Responden: Awalnya terdapat 224 data kuesioner, setelah preprocessing tersisa 214 data valid (10 data dianggap outlier dan dikeluarkan). • Isi Kuesioner: Durasi membaca (berapa lama waktu membaca), Jenis bacaan (buku cetak, e-book, artikel online, media sosial), Frekuensi membaca (harian atau mingguan), Perangkat yang digunakan untuk membaca (buku cetak, smartphone, tablet, laptop).
<p>Hasil</p>	<p>Algoritma K-Means efektif membantu mengelompokkan pola baca siswa di era digital sehingga dapat mendukung pengambilan keputusan pendidikan berbasis data secara tepat sasaran. Algoritma K-Means diterapkan dengan jumlah cluster sebanyak empat kluster. Nilai indeks Davies-Bouldin sebesar -2.224, menunjukkan kualitas kluster sangat baik karena jarak antar elemen dalam satu kluster kecil dan perbedaan antar kluster besar.</p>