

BAB II

LANDASAN TEORI

2.1. DATA MINING

Data mining adalah proses menganalisis sejumlah data yang besar dari berbagai sudut pandang dan menghubungkannya menjadi informasi bermanfaat. Informasi tersebut dapat diubah menjadi pengetahuan tentang tren historis. Teknik-teknik yang digunakan dalam data mining meliputi statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengenali informasi berharga serta pengetahuan terkait dari berbagai basis data yang besar (Panaungi et al., 2025). Data mining juga dapat diartikan sebagai gabungan beberapa nilai, informasi, atau pengetahuan, serta pengalaman yang menghasilkan sebuah kerangka kerja (framework) yang akan dievaluasi dan menghasilkan informasi atau pengetahuan yang baru (Perpustakaan & Kearsipan, 2024).

Data mining memiliki berbagai teknik atau tipe yang digunakan untuk mengidentifikasi pola dan hubungan dalam data. Tipe-tipe data mining yang umum meliputi:

1. **Klasifikasi (Classification):** Teknik yang digunakan untuk memprediksi kelas atau kategori suatu data berdasarkan atribut-atribut yang ada. Contohnya adalah klasifikasi kelulusan mahasiswa (Hasanah & Purnomo, 2022).
2. **Clustering (Clustering):** Merupakan salah satu metode data mining yang bersifat tanpa arahan (unsupervised). Proses ini bertujuan untuk mengelompokkan data atau objek ke dalam kelompok-kelompok tertentu (disebut cluster), di mana objek-objek dalam satu kelompok memiliki tingkat kesamaan yang tinggi di antara mereka, namun berbeda secara signifikan dengan objek-objek di kelompok lain (Nur Afifah & Nurdiyanto, 2023). Ini memungkinkan identifikasi pola atau relasi antar data yang mungkin tidak terlihat secara langsung, membantu memahami struktur data, dan mempermudah pengambilan keputusan.

3. Asosiasi (Association): Teknik untuk menemukan hubungan antar item dalam kumpulan data, seperti pola pembelian produk yang sering dibeli bersamaan.
4. Regresi (Regression): Metode yang digunakan untuk memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independen untuk tujuan prediksi nilai numerik.

Penelitian ini secara khusus akan berfokus pada teknik clustering menggunakan algoritma K-Means untuk mengelompokkan buku sesuai dengan minat peminjam.

Data mining memiliki tujuan utama, seperti menghasilkan pengetahuan baru yang memiliki nilai tambah yang signifikan (Nur et al., 2024). Proses ini melibatkan identifikasi karakteristik penting dalam data, menemukan model yang menggambarkan kelas atau konsep, serta mengungkap pola dan hubungan antar atribut yang sebelumnya tidak terlihat. Dalam konteks pengelolaan informasi perpustakaan, data mining dapat membantu pustakawan menganalisis data peminjaman buku untuk mengidentifikasi tren, preferensi peminjam, dan kebutuhan koleksi untuk pengadaan pembelian buku (Febriyanto et al., 2021). Penerapan data mining dalam pengelolaan data akademik terbukti efisien dalam mengelompokkan data yang memiliki karakteristik serupa. Informasi yang diperoleh dari proses data mining mampu menjadi dasar untuk pengambilan keputusan yang berkualitas. Hal ini krusial bagi Perpustakaan UMG dalam mengatasi tantangan ketersediaan buku yang tidak merata sesuai minat peminjam dan memastikan koleksi yang tersedia lebih relevan dan sesuai dengan minat spesifik peminjam.

2.2. PENGELOMPOKKAN (CLUSTERING)

Pengelompokan (clustering) merupakan teknik yang dapat diterapkan untuk menyelesaikan tantangan dalam mengelola koleksi yang luas dan beragam di Perpustakaan Muhammadiyah Gresik, khususnya dalam mengidentifikasi minat peminjaman buku peminjam. Pengelompokan, atau dikenal juga sebagai segmentasi, adalah salah satu teknik utama dalam unsupervised learning yang

bertujuan untuk mengelompokkan data atau objek ke dalam cluster-cluster tertentu.(Nur Afifah & Nurdianto, 2023). Tujuannya adalah mengelompokkan data yang serupa tanpa label yang telah ditentukan sebelumnya. Dalam sebuah cluster, objek-objek memiliki tingkat kesamaan yang tinggi satu sama lain, namun sangat berbeda dengan objek-objek di cluster lain. Clustering membantu mengidentifikasi struktur tersembunyi dalam data yang tidak memiliki label, sehingga mempermudah proses pengambilan keputusan dan pemahaman terhadap data yang kompleks(Mau et al., 2024).

Teknik pengelompokan dalam konteks manajemen perpustakaan memiliki potensi aplikasi yang luas, termasuk dalam penyesuaian pengelolaan koleksi (Zabidi, 2024). khususnya di institusi pendidikan tinggi seperti Universitas Muhammadiyah Gresik, perpustakaan sering menghadapi tantangan dalam memastikan ketersediaan literatur yang relevan dan sesuai dengan minat serta kebutuhan spesifik peminjamnya. Data peminjaman buku dari sistem informasi perpustakaan (seperti OPAC dan e-Perpus UMG) bisa dianalisis pakai algoritma K-Means untuk mengelompokkan buku sesuai dengan minat peminjam. Hasilnya dapat menjadi prioritas pengadaan buku dan rekomendasi koleksi yang sesuai minat peminjam.

Penelitian ini mengimplementasikan teknik clustering untuk mengelompokkan buku sesuai dengan minat peminjam. Data yang digunakan adalah rekapitulasi jumlah peminjaman buku untuk periode 01 Januari 2023 hingga 17 April 2025. Clusterisasi ini diharapkan dapat membantu pengelolaan koleksi dan penyediaan layanan perpustakaan agar lebih terarah, efisien, dan responsif terhadap kebutuhan peminjam.

2.3. NORMALISASI MIN MAX

. Normalisasi Min Max adalah sebuah metode prapemrosesan data yang penting dalam analisis data. Tujuannya adalah untuk mempersempit cakupan suatu data dengan mengkonversi nilai-nilai dari sebuah data ke dalam rentang nilai minimum 0 (nol) dan nilai maksimum 1 (satu) (Zabidi, 2024). Proses ini krusial untuk memastikan bahwa semua variabel dalam dataset memiliki skala

yang seragam, sehingga mencegah dominasi fitur dengan rentang nilai yang lebih besar yang dapat mempengaruhi perhitungan dalam algoritma clustering seperti K-Means, yang sangat bergantung pada jarak antar titik. Normalisasi Min Max membantu data diproses dengan benar dalam perhitungan. Normalisasi Min-Max dapat dihitung dengan persamaan (2.1)

$$X_{norm} = \frac{X_1 - X_{min}}{X_{max} - X_{min}} \quad (2.1)$$

Dimana:

X_{norm} = nilai setelah dinormalisasi.

X_1 = nilai sebelum dinormalisasi.

X_{min} = nilai minimal setiap variable.

X_{max} = nilai maksimal setiap variable.

2.4. ALGORITMA K-MEANS CLUSTERING

Algoritma K-Means pertama kali digagas oleh Stuart Lloyd pada tahun 1957 dan dipublikasikan oleh E.W. Forgy pada tahun 1965, merupakan algoritma pengelompokan data non-hierarkis yang mempartisi data ke dalam sejumlah cluster yang telah ditentukan di awal. Setiap data akan masuk dalam cluster tertentu dan dapat berpindah cluster pada tahapan proses berikutnya (Padiku & Lahinta, 2022). K-Means bekerja dengan membagi data ke dalam beberapa cluster berdasarkan jarak antara elemen data dan pusat cluster, yang disebut centroid (Andi Syahrul Ramdana et al., 2024).

Tujuan utamanya adalah untuk menciptakan kelompok data yang sangat mirip secara internal dan berbeda secara signifikan antar kelompok. Konsep ini dicapai dengan mengelompokkan objek berdasarkan kemiripan karakteristik di sekitar titik pusat (centroid), yang prosesnya berulang hingga cluster stabil. K-Means sangat sederhana, cepat, mudah digunakan, dan adaptif, menjadikannya metode penting dalam berbagai aplikasi data mining, terutama untuk memproses data besar secara efisien (Nur Afifah & Nurdiyanto, 2023).

Algoritma K-Means dikenal efisien. Prinsip algoritma K-Means melibatkan proses iteratif untuk mencapai pengelompokan data yang optimal. Tahapan dalam algoritma K-Means secara detail adalah sebagai berikut:

1. Langkah awal adalah menentukan Jumlah Cluster (K) yang diinginkan atau akan dibentuk dari dataset (Padiku & Lahinta, 2022a). Setelah nilai K ditentukan, algoritma membangkitkan K buah centroid (titik pusat cluster) awal secara acak atau ditentukan secara default dari dataset.
2. Untuk setiap data input, jaraknya ke masing-masing centroid dihitung menggunakan rumus jarak Euclidean. Data kemudian akan dialokasikan atau dikelompokkan ke dalam cluster yang memiliki centroid terdekat (jarak minimum) (Asminah, 2022). Rumus jarak Euclidean ($D(x,y)$) yang umum digunakan untuk menentukan kemiripan antar data adalah sebagai berikut:

$$D(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.2)$$

Di mana:

$D(x,y)$ = jarak antara data x dan y.

x_i = nilai x pada atribut ke-i

y_i = nilai y pada atribut ke-i

n = jumlah atribut (dimensi data).

3. Data kemudian akan dialokasikan atau dikelompokkan ke dalam cluster yang memiliki centroid terdekat (jarak minimum).
4. Setelah semua data dialokasikan ke cluster masing-masing, posisi centroid baru dihitung. Centroid baru ini diperoleh dari nilai rata-rata (mean) seluruh data yang menjadi anggota cluster tersebut. Rumus untuk memperbarui centroid baru ($\mu_j^{(t+1)}$) adalah sebagai berikut:

$$\mu_j^{(t+1)} = \frac{1}{N_{S_j}} \sum_{x \in S_j} x \quad (2.3)$$

Di mana :

$\mu_j^{(t+1)}$ = centroid baru untuk cluster j pada iterasi ke-(t+1)

N_{S_j} = jumlah data pada cluster S_j

$\sum_{x \in S_j} x$ = jumlah total nilai semua data x yang berada dalam cluster S_j

5. Ulangi Iterasi: Langkah 3 hingga 5 diulang terus-menerus sampai tidak ada lagi data yang berpindah cluster (posisi centroid baru sama dengan centroid sebelumnya) atau telah mencapai kondisi konvergensi.

K-Means sangat sederhana untuk dijalankan dan diterapkan, cukup cepat, mudah digunakan, dan mudah beradaptasi, sehingga umum digunakan dalam berbagai aplikasi data mining. Kelemahannya adalah K-Means bekerja dengan asumsi bahwa data yang dikelompokkan memiliki bentuk yang relatif bulat dan simetris, sehingga kurang efektif untuk menangani data yang lebih kompleks atau memiliki distribusi yang tidak teratur (Zabidi, 2024). Disisi lain, K-Means juga sensitif terhadap outlier, yang dapat mengganggu proses pengelompokan dan menghasilkan cluster yang tidak representatif (Ani et al., 2021).

Penentuan jumlah cluster optimal (k) merupakan tahap krusial dalam proses clustering, karena pemilihan nilai K yang tepat akan menghasilkan partisi cluster yang lebih representatif dan bermakna secara substansi (Suraya et al., 2023). Sebaliknya, penentuan K yang tidak tepat dapat menghasilkan pengelompokan yang kurang relevan. Oleh karena itu, evaluasi clustering menggunakan metrik kuantitatif seperti Davies-Bouldin Index (untuk meminimalkan rasio dispersi intra-cluster terhadap separasi inter-cluster) dan Silhouette Coefficient (untuk mengukur kepadatan dan pemisahan cluster) menjadi esensial dalam validasi hasil.

2.5. DAVIES-BOULDIN INDEX (BDI)

Davies-Bouldin Index (DBI) adalah metrik evaluasi kualitas cluster yang mengukur seberapa baik cluster-cluster yang terbentuk terpisah dan seberapa padat data di dalam masing-masing cluster. Metrik ini bekerja dengan menghitung rasio jarak antar cluster terhadap jarak di dalam cluster itu sendiri (Learning et al., 2025). Secara umum, nilai DBI yang lebih rendah mengindikasikan hasil clustering yang lebih optimal. Semakin kecil nilai DBI, maka definisi dan pemisahan cluster dapat dikatakan semakin baik, yang pada akhirnya menandakan efektivitas yang tinggi dari model clustering yang telah diterapkan. Sebagai

contoh, dalam sebuah penelitian, nilai DBI sebesar 0,541 untuk K=3 cluster menunjukkan kualitas clustering yang baik (Ulfah et al., 2024).

Berikut ialah tahapan pada evaluasi cluster melalui penggunaan metode Davies Bouldin Index:

1. Sum of square within cluster (SSW) yakni persamaan guna mengetahui matrik kohesi dalam suatu cluster ke-i

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{c_i} d(X_j, C_j) \quad (2.4)$$

Dimana :

m_i = jumlah data dalam cluster ke-i

c_i = centroid cluster ke-i

$d(x_j, c_i)$ = jarak euclidean antara setiap data dengan centroid

2. Sum of square between cluster (SSB) yakni rumus yang digunakan guna memperoleh pengetahuan terkait nilai pemisah diantara cluster.

$$SSB_{i,j} = d(c_i, c_j) \quad (2.5)$$

Dimana :

$d(c_i, c_j)$ = jarak antara centroid

3. Setelah nilai pemisah dan kohesi diperoleh, perhitungan rasio ($R_{i,j}$) dilakukan untuk memperoleh pengetahuan terkait perbandingan diantara cluster ke-i dan cluster ke-j, guna menentukan nilai rasio di setiap cluster.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_i} \quad (2.6)$$

Dimana :

SSW_i = Sum of Square Within Cluster pada centroid i

SSB_i = Sum of Square Between Cluster antara data ke-i dengan j pada cluster yang berbeda

4. Davies Bouldin Index (DBI) Nilai rasio yang diperoleh melalui penghitungan rasio dipergunakan dalam menghitung nilai DBI dengan mempergunakan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i, \dots, k}) \quad (2.7)$$

Berdasarkan perhitungan Davies Bouldin Index (DBI) tersebut, mampu dihasilkan simpulan bahwasanya makin kecil nilai DBI yang didapat (dengan nilai non negatif ≥ 0) makin baik kualitas cluster yang tercipta.

2.6. SILHOUETTE COEFFICIENT

Silhouette Coefficient adalah metrik yang digunakan untuk memilih jumlah cluster optimal dengan menggunakan data skala rasio. Silhouette Coefficient mengukur seberapa baik setiap objek ditempatkan dalam cluster-nya dengan menilai dua aspek: Kepadatan Intra-Cluster (seberapa dekat objek dalam satu cluster) dan Pemisahan Antar-Cluster (Inter-Cluster Separation) (seberapa jauh cluster satu dengan lainnya).

Tujuannya adalah untuk memvalidasi bahwa pengelompokan data tidak hanya optimal secara visual (misalnya dengan Elbow Method) tetapi juga secara matematis. Nilai Silhouette Coefficient berkisar antara -1 (clustering buruk) hingga 1 (clustering sangat baik) (Panaungi et al., 2025).

- Nilai mendekati 1 menunjukkan cluster yang terbentuk sempurna dan sangat baik.
- Nilai antara 0.50-0.70 menunjukkan struktur cluster yang cukup baik.
- Nilai di bawah 0.50 menunjukkan cluster yang tumpang tindih atau tidak terdefinisi dengan kurang baik

Tahapan perhitungan Silhouette Coefficient:

1. Hitung rata-rata jarak pada sebuah data misal i bersama seluruh data lain yang berada dalam satu cluster yang sama mempergunakan persamaan 2.8.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.8)$$

Dimana:

$a(i)$ = Perbedaan rata-rata pada data i terhadap seluruh data lainnya dalam cluster A

i, j = Indeks dari data

$d(i, j)$ = Jarak antara data i dan data j

- Melakukan penghitungan akan rata-rata jarak antara suatu bersama seluruh data di cluster lain mempergunakan persamaan 2.9.

$$D(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (2.9)$$

Dimana:

$d(i, C)$ = rata-rata jarak pada data (i) ke seluruh data C

C = Cluster selain A

- Pilih nilai jarak minimum mempergunakan persamaan 2.10.

$$B(i) = \frac{\min_{c \neq A} d(i, j)}{c \neq A} \quad (2.10)$$

Dimana:

$b(i)$ = Nilai jarak rata-rata minimum antara data ke- i dengan seluruh data dalam cluster yang berbeda.

- Hitung nilai Silhouette Coefficient dengan mempergunakan persamaan 2.11.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.11)$$

Dimana :

$a(i)$ = jarak rata-rata data point i ke semua titik dalam cluster yang sama.

$b(i)$ = jarak rata-rata data point i ke semua titik di cluster terdekat (bukan cluster asal).

$s(i)$ = nilai silhouette coefficient

Metrik evaluasi seperti Silhouette Score sangat penting untuk memvalidasi bahwa pengelompokan data tidak hanya optimal secara visual (misalnya dengan Elbow Method) tetapi juga secara matematis, sehingga memastikan kualitas

clustering yang baik. Kualitas ini ditandai dengan pemisahan antar-cluster yang jelas dan kesamaan intra-cluster yang tinggi. Skor akhir dari metrik ini adalah rata-rata nilai $s(i)$ untuk semua data. Sebagai ilustrasi, dalam sebuah studi, nilai Silhouette Score sebesar 0.82 untuk $k=3$ menunjukkan struktur cluster yang sangat baik, dibuktikan oleh pemisahan antar-cluster yang jelas dan kepadatan intra-cluster yang tinggi (Mau et al., 2024).

2.7. PENELITIAN TERDAHULU

Tabel 2.1 Hasil Penelitian Terdahulu

No	Judul Penelitian	Permasalahan	Hasil	Referensi
1.	Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering	Perpustakaan memiliki ± 2000 judul buku dalam beberapa kategori, namun belum ada sistem yang mengatur koleksi buku secara maksimal agar tidak terjadi penumpukan stok. Dibutuhkan sistem untuk menentukan buku yang harus ditambah atau dikurangi stoknya.	Data kategori buku dikelompokkan menjadi tiga cluster: paling diminati, diminati, dan kurang diminati. Hasil ini menjadi bahan evaluasi pustakawan dalam meningkatkan koleksi buku. Implementasi sistem data mining berbasis desktop dengan K-Means terbukti sangat baik dan efektif dalam pengelompokan buku. Dinyatakan ada 10 kategori buku paling diminati (jumlah peminjam lebih besar, stok lebih sedikit), 6	Haryani, Nofriansyah, & Mariami. (2021). Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering. <i>Jurnal Cyber Tech</i> , 1(1), 1–12.

			kategori buku diminati (peminjam sama dengan stok), dan 14 kategori buku kurang diminati (peminjam lebih sedikit, stok lebih besar).	
2.	Implementasi Data Mining Dalam Menentukan Penambahan Koleksi Buku di Perpustakaan: Algoritma K-Means Clustering	Penentuan jenis koleksi buku yang perlu diperbanyak atau tidak menjadi informasi krusial, karena berkaitan dengan kebutuhan pengunjung dan efisiensi pengadaan. Kesalahan pengadaan koleksi dapat dihindari dengan analisis data transaksi peminjaman buku.	Cluster optimal yang terbentuk adalah dua cluster: kategori buku paling diminati dan kategori buku kurang diminati. Berdasarkan stok buku dan kategori diminati, penambahan koleksi buku diperlukan untuk jenis buku pertanian dan ekonomi. Algoritma K-Means efektif untuk analisis pola minat baca dan preferensi mahasiswa.	Mau, P. Y., Chrisinta, D., & Binsasi, E. (2024). Implementasi Data Mining Dalam Menentukan Penambahan Koleksi Buku di Perpustakaan: Algoritma K-Means Clustering. <i>Journal of Mathematics, Computations, and Statistics</i> , 7(1), 122–132.
3.	Penerapan Algoritma K-Means untuk Pengelompokan Koleksi Perpustakaan dengan Data Mining	Pengelolaan koleksi yang terus berkembang dengan jumlah item sangat banyak menyulitkan pengelompokan	Koleksi perpustakaan dapat dikelompokkan berdasarkan kesamaan karakteristik, membantu pengelola	Zabidi, A. F. (2024). Penerapan Algoritma K-Means untuk Pengelompokan Koleksi

		<p>informasi relevan. Perlu peningkatan pengalaman pengguna, optimasi pencarian informasi, dan penyediaan layanan personal. Algoritma K-Means mengharuskan penentuan jumlah cluster (K) terlebih dahulu, sensitif terhadap outlier, dan kurang efektif untuk data yang tidak teratur.</p>	<p>memahami kebutuhan pengguna dan mengoptimalkan koleksi. Ditemukan 4 cluster optimal: buku sering dipinjam dengan durasi pendek (referensi cepat), buku jarang dipinjam dengan durasi panjang (riset mendalam), buku frekuensi sedang pada topik teknologi, dan buku sangat jarang dipinjam dengan nilai referensial tinggi jangka panjang. Evaluasi menunjukkan Silhouette Score 0,68 (kualitas clustering baik), penurunan WCSS tajam setelah cluster keempat, dan Purity 0,85 (data sesuai kategori). Hasil ini memberikan informasi yang berguna bagi pengelola perpustakaan dalam</p>	<p>Perpustakaan dengan Data Mining. Media Jurnal Informatika, 16(2), 233–242.</p>
--	--	---	--	---

			menyusun koleksi dan memberikan rekomendasi buku.	
4.	PENERAPAN METODE K-MEANS UNTUK CLUSTERING PENGUNJUNG PERPUSTAKAAN ITN MALANG	Data transaksi peminjaman selama ini tidak dimanfaatkan lebih lanjut untuk menggali pola informasi yang mungkin berharga dalam mengevaluasi koleksi buku maupun dokumentasi lain. Penentuan pengadaan buku masih dilakukan secara manual.	Aplikasi yang dirancang dapat mengelompokkan pengunjung dan peminjam menggunakan metode K-Means. Pengelompokan data pengunjung bertujuan untuk mengetahui jenis pengunjung yang sering berkunjung, dan data peminjam untuk mengetahui buku yang banyak dipinjam. Hasil pengelompokan dibagi menjadi 3 cluster: sangat sering dipinjam, sering dipinjam, dan jarang dipinjam untuk buku; serta sangat sering berkunjung, sering berkunjung, dan jarang berkunjung untuk pengunjung. Pengujian fungsional sistem berhasil dan	Febrianto, A., Achmadi, S., & Sasmito, A. P. (2021). Penerapan Metode K-Means untuk Clustering Pengunjung Perpustakaan ITN Malang. JATI (Jurnal Mahasiswa Teknik Informatika), 5(1), 61–70.

			berjalan baik pada browser Mozilla Firefox dan Google Chrome. Mayoritas pengguna menilai sangat baik (50% Sangat Baik, 48% Baik, 2% Kurang Baik).	
5.	PENERAPAN ALGORITMA K-MEANS UNTUK MANAJEMEN PERSEDIAAN DI PERPUSTAKAAN	Kesulitan dalam mengidentifikasi dan meramalkan kebutuhan stok dapat mengakibatkan kelangkaan bahan bacaan atau sumber lainnya, mengganggu pelayanan perpustakaan secara keseluruhan.	Dengan penerapan K-Means dan pendekatan RFM (Recency, Frequency, Monetary), akurasi pengelompokan barang meningkat hingga 70%. Analisis menunjukkan efektivitas algoritma K-Means dalam memprediksi kebutuhan stok. Metode ini dapat mendukung keputusan pengelolaan stok di perpustakaan dengan lebih akurat.	Ramdana, A. S., Kusri, & Pramono, E. (2024). Penerapan Algoritma K-Means Untuk Manajemen Persediaan di Perpustakaan. <i>Jurnal Informatika Teknologi dan Sains (JINTEKS)</i> , 6(1), 109–114.
6.	DATA MINING CLUSTERING DALAM PENGELOMPOKA	Buku yang dipinjam kadang tidak ada karena jumlah buku yang dipinjam tidak	Hasil perhitungan manual dan RapidMiner menunjukkan bahwa	Afifah, I. A. N., & Nurdianto, H. (2023). Data Mining Clustering

	<p>N BUKU PERPUSTAKAAN MENGGUNAKAN ALGORITMA K-MEANS</p>	<p>sebanding dengan stok yang ada, menyebabkan kurangnya minat baca. Petugas kesulitan mengetahui banyaknya jumlah buku yang dipinjam dan tingkat peminjaman buku untuk laporan bulanan.</p>	<p>buku yang banyak dipinjam terdapat pada cluster 0 (8 item), buku sedikit dipinjam pada cluster 1 (17 item), dan buku cukup banyak dipinjam pada cluster 2 (11 item). Penerapan data mining dengan K-Means clustering dapat mengelompokkan buku, membantu pihak perpustakaan mengetahui buku mana yang sering dipinjam. Pengelompokan ini dapat membantu pustakawan dalam mengelola peletakan buku.</p>	<p>Dalam Pengelompokan Buku Perpustakaan Menggunakan Algoritma K-Means. JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), 8(3), 802–814.</p>
7.	<p>Implementasi Data Mining Clustering Data Perpustakaan Menggunakan Algoritma K-Means untuk Menentukan Penambahan Koleksi Buku di Perpustakaan UPY</p>	<p>Kesulitan dalam menentukan jenis koleksi yang seharusnya menjadi prioritas untuk diperbanyak atau tidak, karena pengelola belum memiliki pengetahuan</p>	<p>Sistem implementasi data mining clustering data perpustakaan menggunakan algoritma K-Means berbasis website diharapkan dapat membantu pihak</p>	<p>Bakker, E. (2020). Implementasi Data Mining Clustering Data Perpustakaan Menggunakan Algoritma K-Means untuk Menentukan Penambahan</p>

		berdasarkan data pengunjung. Diperlukan sistem untuk membantu menentukan keputusan pengadaan koleksi yang tepat dan cepat.	perpustakaan UPY dalam membentuk clustering peminjaman buku. Pengujian aplikasi menunjukkan sistem berjalan baik: tampilan menarik (56% responden), mudah dijalankan (70%), kinerja sangat baik (53%), dan bermanfaat (53%).	Koleksi Buku di Perpustakaan UPY. Seri Prosiding Seminar Nasional Dinamika Informatika, 4(1).
8.	IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKAN BUKU MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING (STUDI KASUS: PERPUSTAKAAN POLITEKNIK LPP YOGYAKARTA)	Rasio ketersediaan buku dengan jumlah peminjam tidak seimbang, sehingga berpotensi terjadi kekosongan buku. Buku baru atau penambahan eksemplar berasal dari sumbangan, sehingga pengelola tidak bisa mengontrol judul buku yang seharusnya diperbanyak. Beberapa judul buku yang peminatnya sedikit justru terlalu banyak eksemplarnya.	K-Means Clustering dapat membantu pengelompokan buku menjadi paling diminati, cukup diminati, dan sedikit diminati. Hasil ini diharapkan menjadi bahan pertimbangan keputusan untuk memperbanyak judul buku yang paling diminati agar rasio ketersediaan dan peminat seimbang. Dari 30 judul buku yang diuji, 3 judul merupakan paling diminati.	Hasanah, N. N., & Purnomo, A. S. (2022). Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik LPP Yogyakarta). Jurnal Teknologi Dan Sistem Informasi Bisnis, 4(2), 300–311.
9.	Penerapan Clustering	Penambahan koleksi	Metode clustering K-	Padiku, I. R., &

	K-Means Untuk Mendukung Pengelolaan Koleksi Pada Perpustakaan Fakultas Teknik Universitas Negeri Gorontalo	buku hanya berdasarkan acuan buku terlaris, rekomendasi dosen, atau buku terkini, tanpa mempertimbangkan koleksi yang diminati mahasiswa. Hal ini berdampak negatif jika ada penambahan koleksi buku yang kurang diminati, menyebabkan masalah pelayanan dan kerugian finansial.	Means dapat digunakan untuk mengelompokkan minat baca mahasiswa dan mendukung pengelolaan penambahan koleksi buku. Dari 101 buku yang dipinjam, Cluster A (1 buku) paling sedikit diminati, Cluster B (91 buku) paling banyak diminati, dan Cluster C (9 buku) rata-rata diminati. Nilai Davies Bouldin Index (DBI) rata-rata 0,38 menunjukkan hasil evaluasi yang cukup baik.	Lahinta, A. (2022). Penerapan Clustering K-Means Untuk Mendukung Pengelolaan Koleksi Pada Perpustakaan Fakultas Teknik Universitas Negeri Gorontalo. JURNAL TEKNIK, 20(1), 54–62.(Padiku & Lahinta, 2022b)
10.	K-Means Clustering untuk Analisis Tren Peminjaman Buku di Perpustakaan	Pustakawan menghadapi tantangan dalam memahami pola peminjaman buku yang semakin berkembang. Sebagian besar penelitian sebelumnya masih terbatas pada aspek frekuensi peminjaman.	Algoritma K-Means Clustering efektif untuk analisis tren peminjaman buku. Ditemukan tiga cluster utama (C1: pendidikan umum/moralitas; C2: kewarganegaraan/nasionalisme; C3: bisnis/akuntansi/ekon	Rosiana, L., & Yuadi, I. (2025). K-Means Clustering untuk Analisis Tren Peminjaman Buku di Perpustakaan. Journal of Technology and Informatics (JOTI), 7(1), 1–10.

			<p>omi) yang mencerminkan kecenderungan peminjaman mahasiswa, sangat berkorelasi dengan relevansi konten buku terhadap kurikulum dan kebutuhan akademik. Cluster C1 memiliki dominasi paling besar. Perpustakaan disarankan untuk mengelola koleksi berdasarkan cluster, melakukan evaluasi berkala, dan mempromosikan buku yang kurang populer.</p>	
11.	Implementasi Algoritma K-Means Untuk Rekomendasi Pengadaan Buku	Pengadaan buku seringkali berdasarkan intuisi atau permintaan tidak terstruktur, menyebabkan banyak buku kurang diminati dan rendahnya angka peminjaman.	<p>Penelitian mengidentifikasi 3 cluster: Cluster 1 (7 kategori) kurang diminati, Cluster 2 (2 kategori) diminati, dan Cluster 3 (1 kategori) banyak diminati. Cluster 3 (Kesusasteraan) adalah yang paling banyak diminati.</p>	<p>Ardhianto, A., Hartanti, D., & Maulindar, J. (2025). Implementasi Algoritma K-Means Untuk Rekomendasi Pengadaan Buku. Infotek: Jurnal Informatika dan Teknologi, 8(1),</p>

			<p>Algoritma K-Means efektif menyajikan segmentasi akurat. Hasil ini membantu pengelola merencanakan pengadaan buku berdasarkan minat pemustaka.</p>	160–171.
12.	<p>Penerapan Algoritma K-Means Clustering dalam Menganalisis Pola Peminjaman Buku di Perpustakaan</p>	<p>Manajemen koleksi buku dan pola peminjaman seringkali kompleks, membutuhkan wawasan untuk meningkatkan efisiensi dan pelayanan.</p>	<p>Buku dikelompokkan menjadi lima kategori: buku lama dengan halaman sedikit, buku modern dengan peminjaman singkat, buku klasik dengan durasi peminjaman lama, buku dengan halaman banyak dan peminjaman moderat, dan buku baru dengan variasi durasi peminjaman. Metode Elbow menunjukkan lima cluster adalah yang paling efektif. Nilai Davies-Bouldin Index (DBI) sebesar 0.50 menunjukkan kualitas clustering yang baik.</p>	<p>Sigit, R. A., Akbar, V. R., Rahmaddeni, Aziz, S., & Pebriansyah, E. R. (2024). Penerapan Algoritma K-Means Clustering dalam Menganalisis Pola Peminjaman Buku di Perpustakaan. <i>The Indonesian Journal of Computer Science</i>, 13(5), 8125–8135.</p>

13.	PENGELOMPOKAN BUKU DAN REKOMENDASI BUKU MENGGUNAKAN K-MEANS CLUSTERING PADA DINAS PERPUSTAKAAN DAN KEARSIPAN KOTA MEDAN	Dinas Perpustakaan dan Kearsipan Kota Medan belum mempunyai data akurat mengenai kelompok buku yang paling banyak diminati dan judul buku yang dapat direkomendasikan kepada pengunjung, menyebabkan kesulitan dalam pengadaan koleksi buku.	Penelitian menghasilkan 3 cluster: cluster 0 (kurang diminati) berjumlah 362 buku, cluster 1 (diminati) berjumlah 43 buku, dan cluster 2 (paling diminati) berjumlah 8 buku. Hasil cluster 2 adalah yang direkomendasikan dalam perencanaan penambahan koleksi buku.	Fadlina, J., Utami, R., & Sitinjak, N. M. (2024). Pengelompokan Buku Dan Rekomendasi Buku Menggunakan K-Means Clustering Pada Dinas Perpustakaan Dan Kearsipan Kota Medan. <i>Jurnal Widya</i> , 5(2), 1045–1058.
14.	Prediksi Persediaan Buku Berdasarkan Pola Peminjaman Mahasiswa di Perpustakaan Universitas Ivet Menggunakan Metode K-Means Clustering	Koleksi buku dan referensi terbatas, serta minat baca dan peminjaman buku yang rendah di Perpustakaan Universitas Ivet Semarang. Terdapat kurangnya sinkronisasi antara database kemahasiswaan dengan perpustakaan, menyebabkan pendataan absensi dan peminjaman buku masih manual dan belum terorganisir.	Data buku dikelompokkan menjadi tiga cluster: prioritas (cluster 0), dipertimbangkan (cluster 1), dan tidak diprioritaskan (cluster 2) dalam perencanaan penambahan koleksi buku. Cluster dengan nilai terbesar adalah cluster yang direkomendasikan. Cluster 0 memiliki 51 item data, cluster 1 memiliki 19 item data, dan cluster 2	Dedi Aprivian, M. I., & Aini, N. Q. (2023). Prediksi Persediaan Buku Berdasarkan Pola Peminjaman Mahasiswa di Perpustakaan Universitas Ivet Menggunakan Metode K-Means Clustering. <i>Journal of Systems, Information Technology, and Electronics Engineering</i> , 3(1), 6–11.

			memiliki 27 item data. Hasil ini dapat dijadikan bahan acuan untuk penambahan koleksi buku.	
15.	Rekomendasi Pemilihan Bahan Bacaan Pengunjung Perpustakaan Menggunakan Metode K-Means Clustering	Terdapat masalah dalam menentukan kelompok pengunjung yang memiliki minat baca yang berbeda-beda.	Kmeans membagi responden menjadi 3 kelompok minat baca: rendah (0), sedang (1), dan tinggi (2). Sebanyak 78% responden berpendidikan S1/S2 berada di cluster tinggi, dengan 85%-nya memilih literatur ilmiah. Hasil visualisasi memperlihatkan distribusi yang jelas antar kelompok.	Panaungi, F., Harnelia, H., Abbas, M. A., & Saputra, R. A. (2025). Rekomendasi Pemilihan Bahan Bacaan Pengunjung Perpustakaan Menggunakan Metode K-Means Clustering. <i>Komputika: Jurnal Sistem Komputer</i> , 14(1), 105–113.