

K-MEANS CLUSTERING UNTUK MENENTUKAN PENAMBAHAN BUKU DI PERPUSTAKAAN DEPAN SMKN 8 JAKARTA

Rifqi Arya Putra¹, Ni Wayan Parwati Septiani², Wanti Rahayu³

^{1,2,3}*Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Indraprasta PGRI
Jl. Raya Tengah No. 80, Kel. Gedong, Kec. Pasar Rebo, Jakarta Timur 13760
rifqiaptr11@gmail.com, wayan.parwati@gmail.com, wanti.unindra27@gmail.com*

ABSTRAK

Peningkatan kualitas layanan perpustakaan menuntut strategi yang tepat dalam pengelolaan koleksi buku, khususnya dalam memenuhi kebutuhan dan preferensi pemustaka. Salah satu permasalahan umum yang dihadapi adalah ketidaksesuaian antara koleksi yang tersedia dengan minat baca pengguna. Pemanfaatan teknologi data mining dalam pengelolaan perpustakaan menjadi penting untuk mendukung pengambilan keputusan berbasis data. Penelitian ini bertujuan untuk mengelompokkan data peminjaman buku menggunakan metode K-Means Clustering guna menentukan kebutuhan penambahan buku di Perpustakaan Depan SMK Negeri 8 Jakarta. Tahapan penelitian meliputi seleksi data, penentuan jumlah *cluster* dan centroid awal, serta proses iteratif hingga mencapai konvergensi. Hasil *clustering* menunjukkan dua kelompok utama, yaitu buku yang sering dipinjam dan buku yang jarang dipinjam. Buku dengan tingkat peminjaman tertinggi berdasarkan kategorinya direkomendasikan untuk ditambah dalam koleksi guna memenuhi kebutuhan pengguna. Proses *clustering* menghasilkan pemisahan *cluster* yang stabil dan mampu merepresentasikan pola peminjaman secara akurat, dengan nilai *Silhouette Coefficient* sebesar 0,63. Hal ini menunjukkan bahwa metode *clustering* yang diterapkan berjalan secara efektif dan menghasilkan kualitas pengelompokan yang baik. Implementasi metode ini terbukti mendukung perumusan strategi penambahan buku secara objektif dan efisien.

Kata Kunci: Perpustakaan, Data Mining, K-Means, Clustering, Java.

ABSTRACT

Improving the quality of library services requires appropriate strategies for managing book collections, particularly in meeting the needs and preferences of users. One common issue faced is the mismatch between the available collection and users' reading interests. The use of data mining technology in library management is essential to support data-driven decision-making. This study aims at clustering book borrowing data using the K Means Clustering method to determine the need for additional books at the Library of SMK Negeri 8 Jakarta. The research stages include data selection, determination of the number of clusters and initial centroids, and an iterative process until convergence is achieved. The clustering results identify two main groups, namely frequently borrowed books and rarely borrowed books. Books with the highest borrowing frequency in each category are recommended to be added to the collection to meet user demand. The clustering process produced stable cluster separation and accurately represented borrowing patterns, with a Silhouette Coefficient value of 0,63. This indicates that the applied clustering method performs effectively and results in high-quality grouping. The implementation of this method has proven useful in supporting the formulation of book procurement strategies in an objective and efficient manner.

Key Word: Library, Data Mining, K-Means, Clustering, Java.s

PENDAHULUAN

Peningkatan mutu layanan perpustakaan sekolah sangat bergantung pada kemampuan pengelola dalam memahami kebutuhan pemustaka serta memetakan koleksi buku secara tepat. Permasalahan umum yang sering ditemukan adalah ketidaksesuaian antara koleksi yang tersedia dengan minat baca siswa, sehingga banyak buku yang kurang dimanfaatkan secara optimal. Di sisi lain, buku-buku yang sering dicari justru jumlahnya terbatas. Oleh karena itu, diperlukan pendekatan berbasis data untuk

mendukung pengambilan keputusan dalam penambahan koleksi buku

Salah satu pendekatan yang relevan adalah pemanfaatan data mining, khususnya metode *unsupervised learning* seperti K-Means *clustering* yang mampu mengelompokkan data tanpa label atau target tertentu (Han et al., 2022). K-Means dipilih karena algoritma ini sederhana, efisien, dan mampu menangani data dalam jumlah besar secara cepat, meskipun memiliki keterbatasan dalam penentuan jumlah *cluster* awal (Jain, 2010).

Dengan metode ini, data peminjaman buku dapat dianalisis untuk mengidentifikasi pola penggunaan koleksi berdasarkan kategori dan frekuensi peminjaman.

Berbagai penelitian telah memanfaatkan algoritma K-Means dalam konteks perpustakaan. Misalnya, Afifah & Nurdiyanto (2023) berhasil mengelompokkan koleksi buku di perpustakaan berdasarkan frekuensi peminjaman, sehingga membantu dalam penataan dan rekomendasi koleksi. Penelitian serupa juga dilakukan oleh Hasanah & Purnomo (2022) yang menunjukkan bahwa pengelompokan buku dapat digunakan untuk mendukung pengambilan keputusan terkait pengadaan buku baru secara lebih objektif. Lebih lanjut, Fadlina et al., (2024) mengkategorikan buku ke dalam kelompok kurang diminati, diminati, dan paling diminati, yang menjadi dasar rekomendasi penambahan koleksi secara tepat sasaran. Penelitian lain oleh Firmansyah et al., (2022) menyoroti bahwa hasil *clustering* dapat membantu pustakawan menentukan jumlah penambahan buku sesuai dengan kebutuhan pengguna, sehingga memperbaiki layanan secara signifikan. Haryani et al., (2021) turut menyatakan bahwa penerapan data mining dengan K-Means mendukung pengelolaan koleksi secara objektif dan efisien karena keputusan terkait penambahan atau pengurangan buku didasarkan pada data peminjaman yang aktual.

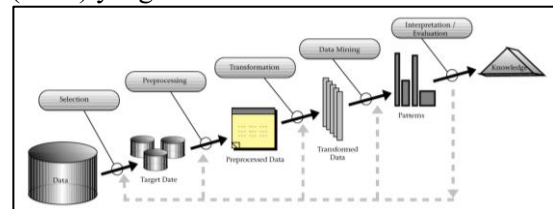
Tujuan dari penelitian ini adalah untuk menerapkan metode K-Means *clustering* pada data peminjaman buku di Perpustakaan Depan SMKN 8 Jakarta guna menentukan kebutuhan penambahan koleksi berdasarkan kategori dan frekuensi peminjaman. Hasil penelitian diharapkan dapat membantu pustakawan dalam merumuskan strategi penambahan buku secara objektif dan efisien.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining. Desain penelitian dirancang untuk menerapkan algoritma K-Means *clustering* dalam mengelompokkan data peminjaman buku berdasarkan frekuensi peminjaman per kategori selama satu tahun.

Penelitian dilakukan pada Perpustakaan Depan SMKN 8 Jakarta yang berlokasi di Pejaten Raya, RT.6/RW.6, Pejaten Bar., Ps. Minggu, Kota Jakarta Selatan. Proses penelitian berlangsung selama 4 bulan, terhitung sejak Februari hingga Mei 2025. Penelitian dilakukan secara langsung (*in-house*) dengan memanfaatkan sistem perpustakaan yang telah menyediakan data digital dalam bentuk *database*, sehingga memudahkan proses pengambilan dan analisis data.

Analisis data dalam penelitian ini mengacu pada tahapan proses *Knowledge Discovery in Databases* (KDD) menurut Fayyad et al., (1996) yang terdiri dari:



Gambar 1. Tahapan KDD

1. **Pembersihan Data (*Data Cleaning*)**
Menghilangkan data duplikat, data kosong, atau kesalahan input untuk memastikan kualitas data yang baik.
2. **Integrasi Data (*Data Integration*)**
Menggabungkan data dari beberapa tabel dalam *database* perpustakaan.
3. **Seleksi Data (*Data Selection*)**
Memilih data yang relevan, yaitu data peminjaman buku tahun 2024 dengan atribut kode buku, judul, kode kategori dan dipinjam.
4. **Transformasi Data (*Data Transformation*)**
Mengubah format data menjadi lebih sesuai untuk dianalisis, seperti menjumlahkan frekuensi peminjaman menjadi total tahunan pada setiap buku.
5. ***Data Mining***
Menerapkan algoritma K-Means *clustering* untuk mengelompokkan buku berdasarkan frekuensi peminjaman.
6. ***Pattern Evaluation***
Menilai kualitas *cluster* menggunakan *silhouette coefficient* untuk memastikan hasil pengelompokan valid dan representatif.
7. ***Knowledge Presentation***

Menyajikan hasil pengelompokan dalam bentuk grafik dan tabel sebagai dasar rekomendasi penambahan koleksi buku.

HASIL DAN PEMBAHASAN

Penelitian ini menerapkan metode K-Means *clustering* terhadap data peminjaman buku di Perpustakaan Depan SMKN 8 Jakarta selama periode Januari hingga Desember 2024. Sebanyak 770 data buku dianalisis setelah melalui proses pembersihan, integrasi, seleksi dan transformasi data.

Mengingat K-Means *clustering* merupakan algoritma *unsupervised learning*, seluruh dataset digunakan secara langsung tanpa perlu dibagi menjadi *data training* dan *data testing*. Hal ini disebabkan oleh karakteristik metode yang tidak memerlukan label atau target awal, melainkan mengelompokkan data berdasarkan kemiripan nilai atribut yang dimiliki.

Tabel 1. Dataset

Kode Buku	Judul	Kode Kategori	Dipinjam
BK00001	Zakat dan Infak: Salah Satu Solusi Mengatasi Problema Sosial di Indonesia	14	0
BK00002	Zakat dalam Perekonomian Modern	14	0
BK00003	Ya Allah, Sungguh Saya Tak Pantas di Surga, Tapi Juga Tak Kuat di Neraka	13	1
BK00004	Waspada Isis	14	0
BK00005	Wakaf dan Pemberdayaan Umat	14	2
BK00006	Virus Liberalisme di Perguruan Tinggi Islam	14	1
BK00007	The World Of Me	2	1
BK00008	The History Of The Sultanate Of Riau Lingga Viewed From A Legal and Cultural Perspective	20	0

BK00009	Teampil Mengemudi ke Negeri Akhirat	13	5
BK00010	Tegas dalam Prinsip, Lentur Dalam Cara	13	2
...
BK00070	100 Things I Wish I Knew Earlier	13	0

Proses *clustering* dilakukan dengan jumlah *cluster* ($k = 2$), yaitu buku yang sering dipinjam dan jarang dipinjam. Dua data dipilih sebagai centroid awal, masing-masing mewakili buku dengan frekuensi peminjaman tertinggi (C1) dan terendah (C0) dalam dataset.

- C0 menggunakan data:
BK00007 | THE WORLD OF ME | 2 | 1
- C1 menggunakan data:
BK00190 | BELENGGU | 15 | 9

Sebelum proses perhitungan, atribut kode kategori yang semula berupa kode numerik diubah terlebih dahulu menggunakan metode *One-Hot Encoding* (OHE). Konversi ini mengubah setiap kategori menjadi vektor biner agar dapat dihitung dalam proses jarak secara numerik.

Pada setiap iterasi, dilakukan perhitungan jarak antara tiap data dan kedua centroid menggunakan rumus *euclidean distance* sebagai berikut (Gan et al., 2007):

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

$d_{(x,y)}$ = Jarak antara 2 titik
 $x_i y_i$ = Nilai atribut ke- i
 n = Jumlah atribut atau dimensi data

Setiap buku akan dikelompokkan ke centroid terdekat berdasarkan nilai jarak tersebut. Iterasi dilanjutkan dengan memperbarui posisi centroid hingga tidak terjadi lagi perpindahan *cluster* atau centroid tetap (konvergen).

BK00001 nilai (14,0) dengan Jarak ke C0 (2,1)

Kategori 14 → Index 13 = 1

Kategori 2 → Index 1 = 1

$$C0 = \sqrt{(1 - 0)^2 + (0 - 1)^2 + (0 - 1)^2}$$

$$C0 = \sqrt{1 + 1 + 1} = 1,73$$

BK00001 nilai (14,0) dengan Jarak ke C1 (15,9)

Kategori 14 → Index 13 = 1

Kategori 15 → Index 14 = 1

$$C1 = \sqrt{(1 - 0)^2 + (0 - 1)^2 + (0 - 9)^2}$$

$$C1 = \sqrt{1 + 1 + 81} = 9,11 \text{ dst.}$$

Pada penelitian ini, proses iterasi dilakukan sebanyak 3 kali hingga posisi centroid tidak mengalami perubahan, yang menandakan bahwa proses telah mencapai konvergensi dan menghasilkan *clustering* akhir. Hasil perhitungan jarak antara setiap buku dengan centroid setelah iterasi ketiga dapat dilihat pada tabel berikut:

Tabel 2. Hasil *Clustering*

Kode Buku	Jarak C0	Jarak C1	Cluster
BK00001	1,41	7,07	C0
BK00002	1,41	7,07	C0
BK00003	1,00	6,08	C0
BK00004	1,41	7,07	C0
BK00005	1,41	5,10	C0
BK00006	1,00	6,08	C0
BK00007	1,00	6,08	C0
BK00008	1,41	7,07	C0
BK00009	4,12	2,24	C1
BK00010	1,41	5,10	C0
...
BK00770	1,41	7,07	C0

Dengan demikian, hasil *clustering* akhir menunjukkan bahwa, sebanyak 143 buku dikelompokkan ke dalam *cluster* C1 (sering dipinjam), sedangkan 627 buku lainnya masuk ke dalam *cluster* C0 (jarang dipinjam). Pembagian ini menjadi dasar dalam mengidentifikasi pola peminjaman yang dapat dimanfaatkan untuk pengambilan keputusan terkait pengelolaan koleksi buku di perpustakaan.

Untuk menilai kualitas hasil pengelompokan tersebut, dilakukan evaluasi menggunakan *silhouette coefficient*. Metode ini dipilih karena mampu mengukur seberapa baik suatu data berada dalam *cluster* yang tepat, dengan membandingkan rata-rata jarak data tersebut terhadap seluruh anggota dalam *cluster* yang sama ($a(i)$) dan rata-rata jaraknya terhadap anggota dari *cluster* terdekat lainnya ($b(i)$). Nilai *silhouette coefficient* dihitung menggunakan rumus berikut (Struyf et al., 1997):

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Nilai $s(i)$ antara -1 sampai 1:

Nilai > 0 = *cluster* cukup baik

Nilai ≈ 0 = data di batas dua *cluster*

Nilai < 0 = salah *cluster*

BK00001

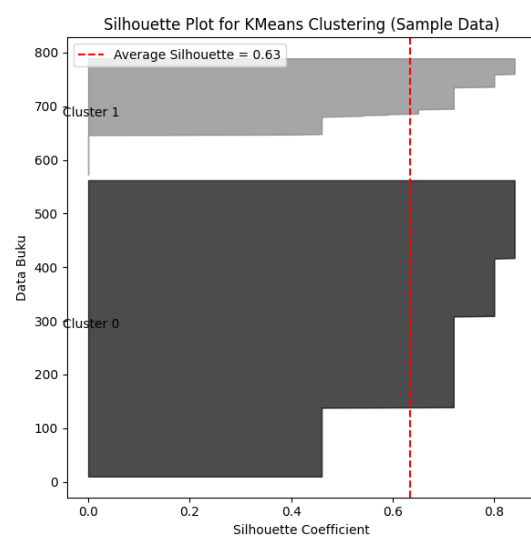
$$s(i) = \frac{7,07 - 1,41}{7,07} = \frac{5,66}{7,07} \approx 0,80 \text{ dst.}$$

Berdasarkan perhitungan menggunakan rumus tersebut, diperoleh nilai *silhouette coefficient* untuk masing-masing data yang telah dikelompokkan. Hasil perhitungan dari data *clustering* ditampilkan pada tabel berikut:

Tabel 3. Hasil *Silhouette Coefficient*

Kode Buku	Jarak C0	Jarak C1	Cluster	SC
BK00001	1,41	7,07	C0	0,80
BK00002	1,41	7,07	C0	0,80
BK00003	1,00	6,08	C0	0,84
BK00004	1,41	7,07	C0	0,80
BK00005	1,41	5,10	C0	0,72
BK00006	1,00	6,08	C0	0,84
BK00007	1,00	6,08	C0	0,84
BK00008	1,41	7,07	C0	0,80
BK00009	4,12	2,24	C1	0,46
BK00010	1,41	5,10	C0	0,72
...
BK00770	1,41	7,07	C0	0,80

Dari hasil tersebut, diperoleh rata-rata nilai *silhouette coefficient* sebesar 0,63 yang ditampilkan pada gambar di bawah ini:



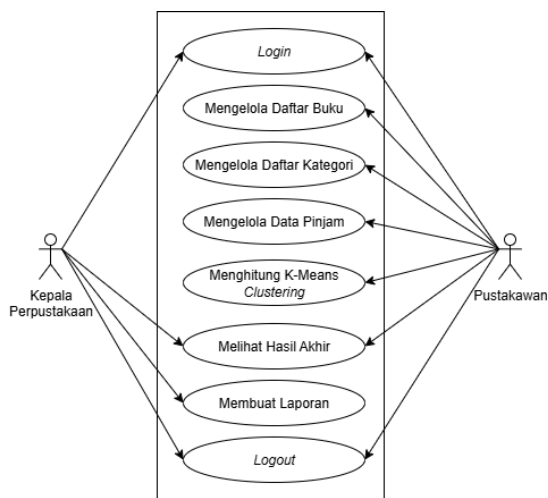
Gambar 2. Grafik Hasil *Silhouette Coefficient*

Gambar tersebut memperkuat kesimpulan bahwa proses *clustering* menggunakan

algoritma K-Means telah menghasilkan pemisahan *cluster* yang stabil dan representatif terhadap pola peminjaman buku, serta menunjukkan bahwa hasil pengelompokan tergolong baik.

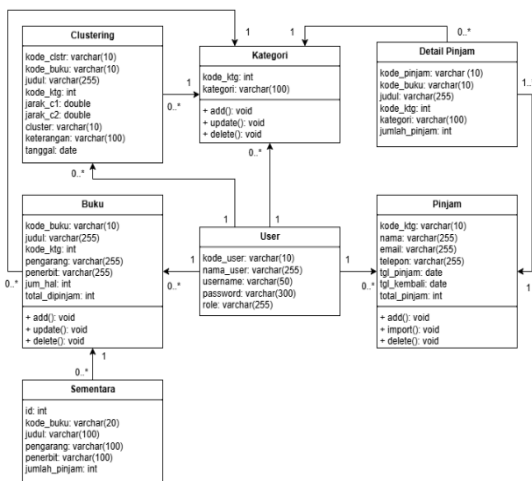
PEMODELAN PERANGKAT LUNAK

Untuk menggambarkan rancangan sistem yang dibangun menggunakan pemodelan perangkat lunak berbasis *Unified Modeling Language (UML)*. *Use Case Diagram* digunakan untuk memvisualisasikan interaksi antara aktor dan sistem, sehingga alur fungsi utama dapat dipahami dengan jelas, seperti gambar berikut:



Gambar 3. Use Case Diagram

Selanjutnya, *Class Diagram* digunakan untuk merepresentasikan struktur sistem secara statis, meliputi kelas, atribut, operasi, serta relasi antar kelas yang membentuk dasar perancangan aplikasi, seperti gambar berikut:



Gambar 4. Class Diagram

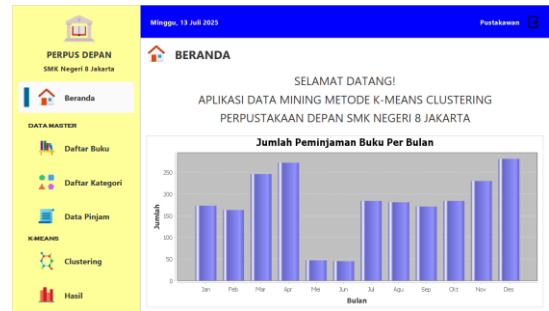
Menu Login



Gambar 5. Tampilan Menu Login

Saat program pertama kali dijalankan, pengguna akan langsung diarahkan ke tampilan form login. Pada tahap ini, pengguna diminta untuk memasukkan *username* dan *password* sesuai dengan akun yang dimiliki. Jika *username* dan *password* yang dimasukkan sudah sesuai, sistem akan menampilkan menu Beranda sebagai tampilan berikutnya.

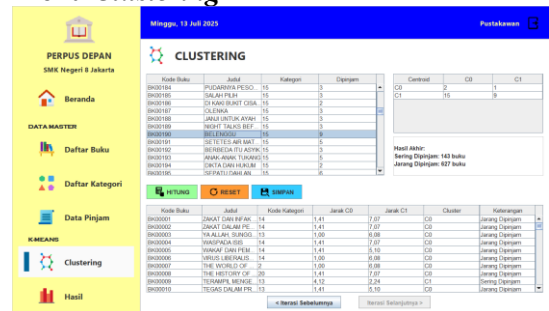
Menu Beranda



Gambar 6. Tampilan Menu Beranda

Setelah *login*, pengguna diarahkan ke menu Beranda yang menampilkan grafik jumlah peminjaman buku per bulan sebagai gambaran aktivitas peminjaman. Melalui halaman ini, pengguna dapat mengakses fitur lainnya lewat panel navigasi di kiri.

Menu Clustering

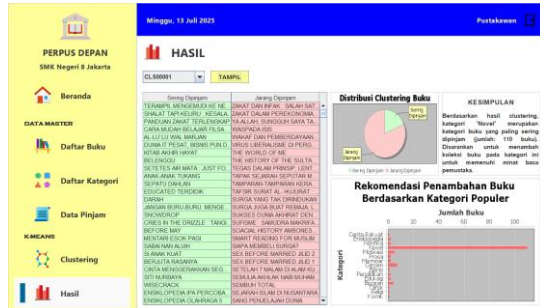


Gambar 7. Tampilan Menu Clustering

Halaman ini digunakan untuk mengelompokkan buku berdasarkan kategori dan frekuensi peminjaman menggunakan

algoritma K-Means. Pengguna memilih dua centroid awal, lalu sistem menghitung dan menampilkan hasil *cluster*, jarak antar data, serta ringkasan jumlah buku per *cluster*.

Menu Hasil



Gambar 8. Tampilan Menu Hasil

Halaman ini menampilkan hasil akhir *clustering* dalam bentuk tabel, grafik distribusi *cluster*, dan rekomendasi penambahan buku berdasarkan kategori populer. Disediakan juga kesimpulan ringkas untuk membantu pengambilan keputusan pengadaan buku.

SIMPULAN DAN SARAN

Penelitian ini berhasil mengelompokkan data peminjaman buku selama satu tahun menggunakan algoritma K-Means menjadi dua *cluster* utama, yaitu buku yang sering dipinjam dan jarang dipinjam. Proses *clustering* dilakukan berdasarkan frekuensi peminjaman, dan hasilnya divisualisasikan dalam bentuk tabel serta grafik yang mendukung analisis kebutuhan koleksi. Evaluasi performa model menggunakan *silhouette coefficient* dengan nilai sebesar 0,63 yang menunjukkan bahwa hasil pengelompokan berada pada kategori cukup baik. Untuk pengembangan lebih lanjut, disarankan menambahkan variable seperti tahun terbit atau umur buku, serta membandingkan dengan algoritma lain guna memperoleh hasil yang lebih akurat dan komprehensif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada SMK Negeri 8 Jakarta yang telah memberikan izin serta menyediakan data dan informasi pendukung yang sangat membantu dalam kelancaran pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada seluruh staf perpustakaan yang telah memberikan waktu,

bantuan, dan keterbukaan dalam memberikan data yang dibutuhkan. Tidak lupa, penulis menyampaikan apresiasi dan rasa hormat kepada dosen pembimbing yang dengan penuh kesabaran telah memberikan arahan, masukan, dan bimbingan selama proses penyusunan artikel ini, sehingga penulisan dapat diselesaikan dengan baik. Segala dukungan dan kontribusi yang diberikan menjadi bagian penting dalam keberhasilan penelitian ini.

DAFTAR PUSTAKA

- Afifah, I. A. N., & Nurdianto, H. (2023). Data Mining Clustering dalam Pengelompokan Buku Perpustakaan Menggunakan Algoritma K-Means. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 8(3), 802–814. <https://doi.org/10.29100/jipi.v8i3.3891>
- Fadlina, J., Utami, R., & Sitinjak, N. M. (2024). *Pengelompokan Buku dan Rekomendasi Buku Menggunakan K-Means Clustering Pada Dinas Perpustakaan dan Kearsipan Kota Medan*. 5(2). <https://doi.org/10.54593/awl.v5i2.288>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases* (© AAAI) (Vol. 17). www.ffly.com/
- Firmansyah, T., Poningih, & Andani, S. R. (2022). Analisis Clustering Algoritma K-Means Sebagai Rekomendasi Penambahan Koleksi Buku Di Perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun. *ZAHRA: Buletin Big Data, Data Science and Artificial Intelligence*, 1(1), 44–48. <https://ejurnal.pdsi.or.id/index.php/zahra/article/view/13%0Ahttps://ejurnal.pdsi.or.id/index.php/zahra/article/viewFile/13/11>
- Gan, Guojun., Ma, Chaoqun., & Wu, Jianhong. (2007). *Data Clustering: Theory, Algorithms, and Applications*. SIAM, Society for Industrial and Applied Mathematics; American Statistical Association.
- Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques*.

- Haryani, Nofriansyah, D., & Mariami, I. (2021). *Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering*. <https://ojs.trigunadharma.ac.id/index.php/jct/index>
- 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.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Struyf, A., Hubert, M., & Rousseeuw, P. J. (1997). Clustering in an Object-Oriented Environment. *Journal of Statistical Software*, 1(4), 1–30. <https://doi.org/10.18637/jss.v001.i04>