

Implementasi Algoritma K-Means Untuk Rekomendasi Pengadaan Buku

Aan Ardianto^{1*}, Dwi Hartanti², Joni Maulindar³

^{1,2,3}Teknik Informatika, Universitas Duta Bangsa Surakarta

*aanardh@gmail.com

Abstrak

Salah satu tantangan utama yang dihadapi oleh perpustakaan adalah menentukan pengadaan koleksi buku agar sesuai dengan kebutuhan dan minat peminjam. Di Dinas Arsip dan Perpustakaan Kabupaten Sragen, pengadaan buku seringkali dilakukan berdasarkan intuisi atau permintaan yang tidak terstruktur, akibatnya banyak buku yang tersedia kurang diminati oleh pengunjung. Hal ini menyebabkan rendahnya angka peminjaman buku, sehingga perpustakaan tidak dapat memberikan layanan secara optimal. Berdasarkan permasalahan di atas, penulis mencoba untuk mengelompokkan data peminjaman buku pada tahun 2023 dari Dinas Arsip dan Perpustakaan berdasarkan kelompok umur, kategori buku, jumlah peminjaman, jumlah judul dan jumlah eksemplar menggunakan teknik data mining dengan algoritma k-means clustering. Untuk proses data awal penulis menggunakan metode normalisasi Min Max, setelah dilakukan normalisasi, menghitung k-means dengan 3 klaster, setelah itu mencari klaster optimal dengan metode elbow, silhouette, dan gap statistik. Hasil dari klaster optimal dibandingkan dengan hasil dari metode Dunn Index. Hasil penelitian mengidentifikasi 3 cluster, yaitu cluster 1 kelompok buku kurang diminati berjumlah 7 kategori yaitu Karya Umum, Ilmu Sosial, Bahasa, Ilmu Murni, Ilmu Terapan, Kesenian dan Olahraga, Sejarah dan Geografi, untuk cluster 2 kelompok buku yang diminati berjumlah 2 kategori yaitu Filsafat dan Psikologi, Agama, sedangkan cluster 3 kelompok buku yang paling banyak diminati berjumlah 1 kategori yaitu Kesusasteraan.

Kata kunci : Algoritma K-Means, Data Mining, Koleksi Buku, Pengadaan Buku

Abstract

One of the main challenges faced by libraries is determining the procurement of book collections that align with the needs and interests of borrowers. At the Sragen Regency Archives and Library Service, book procurement is often based on intuition or unstructured requests, resulting in many books that are less popular among visitors. This leads to a low number of book borrowings, meaning the library cannot provide optimal services. Based on this issue, the author attempts to cluster the book borrowing data for the year 2023 from the Sragen Regency Archives and Library Service by age group, book categories, number of borrowings, number of titles, and number of copies using data mining techniques with the k-means clustering algorithm. For the initial data processing, the author uses the Min-Max normalization method. After normalization, k-means is calculated with 3 clusters, followed by finding the optimal cluster using the elbow method, silhouette, and gap statistics. The results of the optimal cluster are compared with the results from the Dunn Index method. The research identifies three clusters: Cluster 1 contains book groups with low interest, consisting of 7 categories: General Works, Social Sciences, Language, Pure Sciences, Applied Sciences, Arts and Sports, History and Geography; Cluster 2 contains book groups with moderate interest, consisting of 2 categories: Philosophy and Psychology, Religion; and Cluster 3 contains the book group with the highest interest, consisting of 1 category: Literature.

Keywords : K-Means Algorithm, Data Mining, Book Collection, Book Acquisition

1. Pendahuluan

Perpustakaan merupakan pusat informasi dan pengetahuan yang sangat penting bagi

masyarakat. Salah satu tantangan utama yang dihadapi oleh perpustakaan adalah menentukan pengadaan koleksi buku agar sesuai dengan

kebutuhan dan minat peminjam. Di Dinas Arsip dan Perpustakaan Kabupaten Sragen, pengadaan buku seringkali dilakukan berdasarkan intuisi atau permintaan yang tidak terstruktur, sehingga buku-buku hasil pengadaan kurang menarik bagi pengunjung. Hal ini menyebabkan rendahnya angka peminjaman buku, sehingga perpustakaan tidak dapat memberikan layanan secara optimal. Dari laporan yang didapatkan penulis, hanya ada sekitar 123.344 pengunjung yang mengunjungi perpustakaan. Angka tersebut masih dianggap kecil untuk skala Kabupaten.

Berdasarkan permasalahan di atas, penulis mencoba untuk mengelompokkan data peminjaman buku pada tahun 2023 berdasarkan kelompok umur, kategori buku, jumlah peminjaman, jumlah judul dan jumlah eksemplar menggunakan teknik data mining dengan algoritma K-Means *clustering*, sehingga data tersebut nantinya dikelompokkan menjadi beberapa *cluster* berdasarkan kemiripan data.

Algoritma K-Means adalah salah satu metode *clusterisasi* yang efektif dan efisien, akan membantu mengidentifikasi pola peminjaman yang tidak terlihat secara kasat mata, serta memberikan rekomendasi pengadaan buku yang lebih akurat. K-Means memiliki keunggulan dalam kemampuannya untuk memproses data berukuran besar serta melakukan *clusterisasi*

dengan cepat^[1], sehingga cocok digunakan dalam analisis data perpustakaan yang biasanya memiliki volume data yang besar. Algoritma ini juga bersifat interaktif dan mudah diperbarui ketika data baru ditambahkan, sehingga sangat fleksibel dalam pengelolaan data dinamis seperti di perpustakaan.

Berdasarkan efektivitas yang telah dibuktikan oleh penggunaan metode K-Means dalam penelitian-penelitian sebelumnya, studi ini akan menerapkan konsep data mining dengan algoritma *K-Means Clustering* untuk menentukan kategori mana saja yang dapat dipertimbangkan untuk penambahan buku. Dengan menggunakan data historis peminjaman buku, sistem ini akan mampu memberikan rekomendasi buku yang lebih sesuai dengan minat dan kebutuhan masyarakat.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Dalam menjalankan penelitian ini, penulis merujuk pada penelitian terdahulu yang berkaitan dengan topik penelitian ini, antara lain:

- Penelitian dengan judul “Penerapan Algoritma Naive Bayes Dalam Pengadaan Buku Referensi Pada Perpustakaan SMA Negeri 1 Trimurjo Berbasis Web”. Penelitian ini dilatarbelakangi oleh kebutuhan untuk meningkatkan efisiensi pengadaan buku di perpustakaan SMA Negeri 1 Trimurjo, karena

pengadaan buku sebelumnya dilakukan tanpa dasar yang kuat, sehingga banyak buku yang tidak diminati. Penelitian ini bertujuan untuk menyediakan sistem rekomendasi pengadaan buku berbasis web menggunakan algoritma Naive Bayes untuk membantu petugas perpustakaan^[2].

- Penelitian dengan judul “Rancang Bangun Sistem Informasi Manajemen dan Klasterisasi Pengadaan Buku Perpustakaan Kabupaten Subang (SIPERSU) Menggunakan Algoritma K-Means”. Penelitian ini dilatar belakangi oleh pencatatan transaksi di Perpustakaan Kabupaten Subang yang masih manual dan tidak efektif, menyebabkan banyak buku yang tidak sesuai dengan minat pengunjung. Penelitian ini bertujuan untuk membangun sistem informasi manajemen perpustakaan dengan klasterisasi buku menggunakan algoritma K-Means untuk meningkatkan efisiensi pengadaan buku^[3].
- Penelitian dengan judul “Penerapan Data Mining Menggunakan Algoritma K-Means untuk Pengelolaan Koleksi Buku di Perpustakaan Universitas Timor”. Penelitian ini dilatarbelakangi oleh masalah dalam mengidentifikasi buku yang paling sering dipinjam di Perpustakaan Universitas Timor, sehingga pengelolaan koleksi buku kurang optimal. Penelitian ini bertujuan untuk menggunakan algoritma K-Means untuk

mengelompokkan data peminjaman buku dan membantu perpustakaan dalam menentukan penambahan koleksi buku dengan lebih efektif dan efisien^[4].

- Penelitian dengan judul “Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi”. Penelitian ini membahas penggunaan algoritma K-Means untuk membantu penjual smartphone dalam memberikan rekomendasi produk berdasarkan spesifikasi dan harga^[5].

2.2. Landasan Teori

1. Normalisasi MinMax

Normalisasi data diperlukan untuk alat proses data yang digunakan untuk perhitungan, juga untuk mempersempit cakupan suatu data. Normalisasi Min Max mengkonversi nilai dari sebuah data dengan rentang nilai minimum 0 (nol) dan nilai maksimum 1 (satu)^[7]. Untuk rumus dari normalisasi Min Max adalah sebagai berikut :

$$x^I = \frac{x - \min(x)}{\max(x) - \min(x)}$$

2. Metode K-Means

Metode K-Means merupakan metode pengelompokkan yang berkaitan dengan metode K-Medoids yang melakukan partisi set sebuah data ke sejumlah cluster yang sudah ditetapkan langsung di awal. Metode K-Means sangat sederhana untuk di jalankan dan diterapkan,

cukup cepat, mudah di pergunakan, dan mudah beradaptasi umum digunakan dalam berbagai aplikasi kecil hingga menengah. Secara ilmiah, K-Means menjadi salah satu metode yang paling penting dalam penggunaan data mining. Metode k-means clustering merupakan salah satu dari pengelompokan metode nonhirarki yang tujuannya mengelompokkan objek yang diawali dengan mengidentifikasi data yang akan di cluster^[8].

3. Penentuan Jumlah Cluster

Ada beberapa metode yang dapat digunakan untuk menentukan jumlah cluster optimal pada proses clustering. Dalam penelitian ini, metode yang akan digunakan adalah silhouette method, elbow method dan gap statistic. Semua metode ini akan menampilkan plot untuk menentukan jumlah cluster optimal^[10].

a) Silhouette Method

Metode Silhouette digunakan untuk memilih jumlah cluster optimal dengan menggunakan data skala rasio. Ketika diterapkan, algoritma silhouette akan mengukur jarak rata-rata dari suatu objek terhadap seluruh objek yang terdapat pada cluster yang sama dengan objek di clusterlainnya. Nilai Silhouette yang mendekati 1 menunjukkan jumlah cluster yang optimal.

b) Elbow Method

Metode Elbow digunakan untuk memilih jumlah cluster berdasarkan siku yang terbentuk pada suatu titik di grafik SSE dan didasarkan pada

penurunan SSE yang besar. Jika nilai cluster sebelumnya (k-1) dengan nilai cluster selanjutnya (k) mengalami penurunan terbesar maka jumlah cluster tersebut yang tepat (k). Metode ini menggunakan nilai Sum of Square Error (SSE) dari masing-masing jumlah cluster. Semakin besar jumlah cluster, maka SSE akan terus mengecil, sehingga jumlah cluster terbaik adalah jumlah cluster yang mengalami penurunan terbesar. Rumus SSE dapat dituliskan sebagai :

$$SSE = \sum_{k=1}^K \sum_{x_i} |x_i - c_k|^2$$

c) Gap Statistik

Gap statistik merupakan ukuran yang paling konstan untuk menentukan jumlah cluster jika dibandingkan ukuran yang lain. Jarak antara objek berpasangan di dalam cluster dirumuskan sebagai :

$$D_r = \sum_{i, i' \in C_r} d_{ii'}$$

Dimana d adalah kuadrat dari jarak euclidean. Jumlah kuadrat di dalam cluster dirumuskan sebagai berikut.

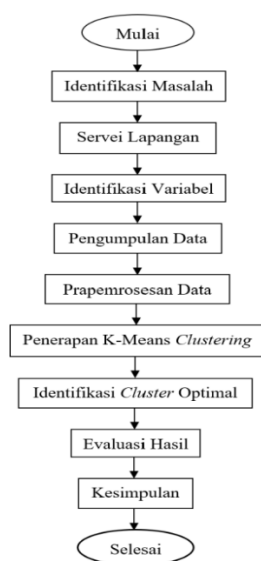
$$W_k = \sum_{r=1}^k \frac{1}{2n_r} D_r$$
$$Gap_n(k) = E_n^* \{ \log(W_k) \} - \log(W_k)$$

Nilai gap merupakan hasil estimasi jumlah cluster optimum dengan menggunakan pendekatan standarisasi W_k . Dimana E_n^* adalah ekspektasi dari distribusi jumlah sampel. Kriteria jumlah cluster optimal merupakan jumlah cluster yang memiliki nilai gap statistik tertinggi atau jika nilai gap selalu naik maka jumlah cluster optimum

adalah nilai yang mengindikasikan kenaikan gap minimum.

3. Metode Penelitian

Jenis penelitian yang digunakan dalam penelitian ini termasuk dalam penelitian kuantitatif, sebab data yang dipakai pada penelitian ini berupa jumlah peminjaman buku berdasarkan kategori dan kelompok umur di Dinas Arsip dan Perpustakaan Kabupaten Sragen. Data yang diperoleh dalam penelitian ini berasal dari data sekunder, yaitu data yang sudah ada dan dikumpulkan oleh pihak lain, kemudian digunakan sebagai bahan analisis dalam penelitian ini. Pada penelitian ini objek yang diteliti adalah jenis buku yang dipinjam, sedangkan atribut yang digunakan seperti total peminjaman buku dari perpustakaan. Adapun tahapan penelitian digambarkan pada Gambar 1 berikut :



Gambar 1. Alur penelitian pengadaan buku

Berdasarkan Gambar 1 alur penelitian dapat dijelaskan sebagai berikut :

1. Tahap identifikasi masalah, merupakan langkah awal dalam proses penelitian ini. Berfokus untuk mengevaluasi objek penelitian, menetapkan cakupan penelitian, dan mengidentifikasi permasalahan yang perlu dipecahkan^[1]. Dalam konteks perpustakaan Dinas Arsip dan Perpustakaan Kabupaten Sragen, tahap ini akan mencakup identifikasi dan pemetaan permasalahan terkait dengan proses pengadaan buku yang kurang optimal. Dengan menargetkan perbaikan pada proses pengadaan buku, penelitian ini bertujuan untuk mengatasi pengadaan buku agar lebih akurat dan sesuai dengan kebutuhan pembaca.
2. Tahapan survei lapangan dalam penelitian ini bertujuan untuk memperoleh informasi terkait bidang penelitian serta menjadi langkah dalam pengumpulan data yang diperlukan.
3. Identifikasi variabel, pada tahapan identifikasi variabel berdasarkan tahapan identifikasi masalah dan survei lapangan maka didapatkan solusi dari permasalahan yang diketahui dengan atribut yang telah ditentukan yaitu kelompok umur, kategori buku, jumlah peminjaman, jumlah judul dan jumlah eksemplar.
4. Proses pengumpulan data pada penelitian ini dilakukan dengan meminta informasi dari

pengelola perpustakaan untuk dijadikan sebagai bahan analisis. Data yang telah dikumpulkan selanjutnya dilakukan proses pengolahan data guna memperoleh informasi terkait kelompok umur, kategori buku, jumlah peminjaman, jumlah judul dan jumlah eksemplar dari data peminjaman buku di perpustakaan.

5. Pada tahap prapemrosesan data, dilakukan serangkaian proses termasuk pembersihan, penambahan, pengurangan, dan penyusunan data guna memastikan data yang digunakan dalam proses analisis mining sesuai dengan kebutuhan.
6. Implementasi K-Means *Clustering*, pada tahap ini dilakukan proses normalisasi data terlebih dahulu. Normalisasi data diperlukan untuk memperkecil rentan data. Dalam penelitian ini digunakan normalisasi Min Max, yaitu mengonversi nilai suatu data ke dalam rentan nilai minimum 0 dan nilai maksimum 1 Untuk persamaannya sebagai berikut^[12]:
7. Identifikasi *cluster* optimal, pada tahapan ini akan ditentukan banyaknya *cluster* optimal yang terbentuk dengan menggunakan metode Elbow, metode Shilhouette, dan Gap Statistik.
8. Tahap akhir dalam proses data mining adalah evaluasi hasil yang didapatkan dari hasil *cluster optimal* dibandingkan dengan hasil dari Dunn Index. Dunn Index mengukur rasio

antara jarak minimum antar *cluster* sebagai indikator *separation* (pemisah) dan ukuran maksimum *cluster* sebagai *compactness* (kepadatan)^[14]. Jumlah *cluster* optimal ditentukan dari nilai Dunn Index yang semakin besar. Rumus perhitungan Dunn Index adalah sebagai berikut:

$$Dunn\ Index = \frac{\min d(C_i, C_j)}{\max(diam(C_k))}$$

Dimana $\min d(C_i, C_j)$ mengukur jarak minimum antar *cluster* dan $\max(diam(C_k))$ menunjukkan diameter *cluster* terbesar

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

Dalam penerapan algoritma K-Means, penulis melakukan pengujian dengan menetapkan nilai k sebesar 3, sehingga data akan dikelompokkan ke dalam tiga *cluster* yang berbeda. Setelah melalui proses iterasi dan perhitungan menggunakan Rstudio, diperoleh *cluster* 1 dengan 7 data, *cluster* 2 dengan 2 data dan *cluster* 3 dengan 1 data.

K-means clustering with 3 clusters of sizes 7, 2, 1

```

Cluster means:
Masa.Balita..0.sd.5.tahun. Masa.Dewasa.Akhir..36.45.tahun. Masa.Dewasa.Awal..26.35.tahun.
1 0.148768473 0.05467372 0.2020906
2 0.003448276 0.59567901 0.7926829
3 0.275862069 0.00000000 0.00000000
Masa.Kanak.Kanak..6.sd.11.tahun. Masa.Manula..56.200.tahun. Masa.Remaja.Akhir..17.25.tahun.
1 0.1462585 0.007142857 0.2964670
2 0.1488095 0.50000000 0.7903226
3 1.00000000 0.314285714 0.00000000
Masa.Remaja.Awal..12.16.tahun. Masa.Lansia.Awal..46.55.tahun. Jumlah.Peminjaman
1 0.04035513 0.03747073 0.08322102
2 0.08192090 0.08196721 0.59316038
3 1.00000000 1.00000000 1.00000000
Jumlah.Judul Jumlah.Eksemplar
1 0.04915515 0.08308004
2 0.57634409 0.58510638
3 1.00000000 1.00000000

Clustering vector:
000 - Karya Umum 100 - Filsafat dan Psikologi 200 - Agama
1 1 2
300 - Ilmu Sosial 400 - Bahasa 500 - Ilmu Murni
1 1 1
600 - Ilmu Terapan 700 - Kesenian dan Olahraga 800 - Kesusasteraan
1 1 3
900 - Sejarah dan Geografi
1 1

```

Gambar 2. Hasil *clusterisasi* dengan 3 *cluster*

Pada klasifikasi ini diketahui "*Within-Cluster Sum of Squares*" (WCSS) dan "*Between-Cluster Sum of Squares*" (BCSS) sebagai berikut :

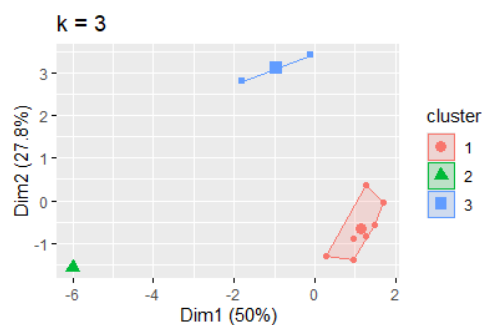
```
within cluster sum of squares by cluster:
[1] 1.573464 1.058685 0.000000
(between_SS / total_SS = 73.8 %)
```

Gambar 3. WCSS dan BCSS dari 3 *cluster*

Nilai [1.573464, 1.058685, 0.000000] menunjukkan WCSS untuk masing-masing dari tiga *cluster*, WCSS mengukur variasi titik data dalam setiap *cluster*, di mana nilai yang lebih rendah menunjukkan bahwa titik-titik dalam *cluster* tersebut lebih dekat dengan pusat *cluster* [15]. *Cluster* 1 memiliki WCSS sebesar 1.573464, yang menunjukkan bahwa titik-titik dalam *cluster* ini memiliki penyebaran yang moderat di sekitar pusat *cluster*. *Cluster* 2 memiliki WCSS sebesar 1.058685, yang menunjukkan bahwa *cluster* ini sedikit lebih kompak dibandingkan *Cluster* 1. *Cluster* 3 memiliki WCSS sebesar 0.000000, yang berarti semua titik dalam *Cluster* 3 identik atau terpusat sempurna pada centroid *cluster* tersebut. Rasio *Between-Cluster Sum of Squares* (BCSS) terhadap *Total Sum of Squares* (TSS) menunjukkan bahwa 73.8% dari variasi total data dijelaskan oleh pemisahan antar-*cluster*. Rasio yang tinggi (mendekati 100%) mengindikasikan bahwa *cluster-cluster* tersebut memiliki pemisah yang baik, sedangkan rasio yang rendah menunjukkan bahwa *cluster-cluster* tersebut tidak memiliki pemisahan yang jelas. Dalam hal ini,

rasio 73.8% menunjukkan pemisahan yang relatif baik antara *cluster-cluster* tersebut.

Plot hasil pembagian data ke dalam tiga *cluster* dapat dilihat pada gambar di bawah ini. Plot ini menampilkan representasi visual mengenai bagaimana data dikelompokkan berdasarkan karakteristik yang serupa di dalam dataset. Setiap *cluster* direpresentasikan dengan warna yang berbeda untuk memudahkan identifikasi dan membedakan antara kelompok-kelompok data.



Gambar 4. Visualisasi Plot dari 3 *cluster*

Jika disimpulkan hasil *clustering* dari 3 *cluster* ini terdiri dari:

Cluster 1 : Karya Umum, Ilmu Sosial, Bahasa, Ilmu Murni, Ilmu Terapan, Kesenian dan Olahraga, Sejarah dan Geografi

Cluster 2 : Filsafat dan Psikologi, Agama

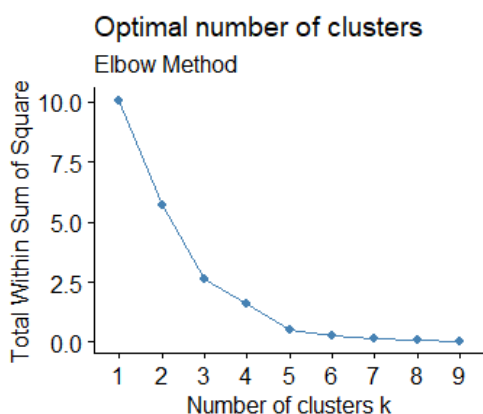
Cluster 3 : Kesusastaan

4.2. Pembahasan

Dari hasil klastering di atas, diperlukan evaluasi untuk menentukan jumlah klaster yang paling sesuai. Tanpa penentuan klaster optimal, ada risiko terjadinya jumlah klaster terlalu banyak atau jumlah klaster terlalu sedikit, yang dapat

mengurangi kualitas dan kegunaan hasil klustering. Oleh karena itu, penulis menggunakan beberapa metode evaluasi untuk memastikan bahwa jumlah kluster yang dipilih benar-benar optimal.

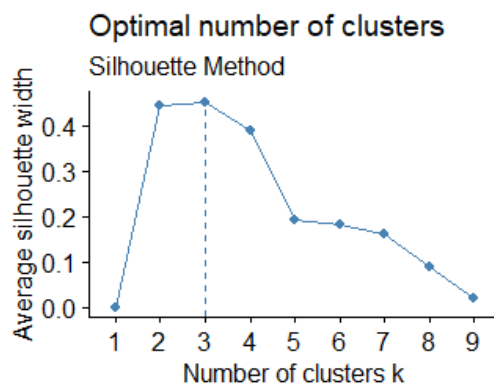
a) Metode Elbow



Gambar 5. Metode Elbow

Berdasarkan plot di atas, terlihat bahwa titik yang menyebabkan grafik membentuk siku atau "elbow" adalah titik ketiga. Hal ini menunjukkan adanya perubahan signifikan dalam penurunan *Within-Cluster Sum of Squares (WCSS)* hingga titik tersebut, dan setelahnya, penurunan WCSS cenderung melambat. Metode Elbow ini bekerja dengan prinsip mengidentifikasi titik di mana penambahan jumlah *cluster* tidak lagi memberikan pengurangan WCSS yang signifikan, sehingga titik ini dianggap sebagai jumlah *cluster* optimal. Dalam kasus ini, titik siku pada posisi ke tiga mengindikasikan bahwa tiga *cluster* merupakan pilihan optimal.

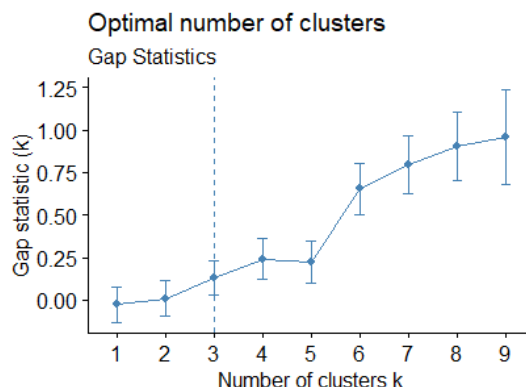
b) Metode Silhouette



Gambar 6. Metode Silhouette

Berdasarkan plot diatas, titik yang berada di puncak garis yaitu pada saat titik ke tiga. Dalam kasus ini, titik puncak pada grafik Silhouette menunjukkan bahwa pembagian data ke dalam tiga *cluster* menghasilkan kualitas *clusterisasi* yang terbaik dibandingkan jumlah *cluster* lainnya. Hal ini mengindikasikan bahwa dengan tiga *cluster*, data dalam masing-masing kelompok lebih homogen (serupa), dan jarak antara *cluster* yang berbeda cukup besar, menciptakan pemisahan yang jelas.

c) Gap Statistik



Gambar 7. Metode Gap Statistik

Berdasarkan plot yang ditampilkan di atas, terlihat bahwa garis putus-putus yang menjadi penanda optimalitas *clusterisasi* terdapat pada titik ketiga. Hal ini menunjukkan bahwa nilai Gap Statistic mencapai nilai tertinggi atau signifikan pada saat jumlah *cluster* adalah tiga. Pada kasus ini, garis putus-putus yang berada di titik ketiga menandakan bahwa jumlah *cluster* tiga memberikan pemisahan terbaik antara *cluster* pada data asli dibandingkan dengan data acak. Ini berarti bahwa dengan tiga *cluster*, model *clusterisasi* dapat menjelaskan struktur data secara signifikan lebih baik dibandingkan jumlah *cluster* lainnya.

d) Dunn Index

Dari hasil proses Rstudio didapatkan hasil nilai dunn index sebagai berikut

Tabel 1. Nilai Dunn Index

K	Dunn Index
2	1.1204818
3	0.9157707
4	1.1065620

Dalam kasus ini, nilai Dunn Index tertinggi dicapai ketika jumlah *cluster* adalah dua *cluster*. Hal ini mengindikasikan bahwa dengan dua *cluster*, struktur *clusterisasi* mampu memberikan keseimbangan terbaik antara pemisahan antar *cluster* (*inter-cluster separation*) dan kepadatan di dalam *cluster* (*intra-cluster compactness*).

Meskipun terdapat perbedaan antara ketiga metode pertama (Elbow, Silhouette, Gap Statistic) dengan Dunn Index, tiga klaster dianggap sebagai

jumlah yang lebih optimal untuk menggambarkan tingkat minat peminjam terhadap kategori buku. Hal ini didasarkan pada konsistensi hasil dari ketiga metode yang saling memperkuat, sementara Dunn Index memberikan perspektif tambahan yang dapat dipertimbangkan dalam analisis lebih lanjut. Hasil klastering tersebut dapat dirangkum sebagai berikut:

Tabel 2. Hasil *Clustering*

No	Kategori	Cluster		
		1	2	3
1	Karya Umum	KD		
2	Filsafat dan Psikologi		D	
3	Agama		D	
4	Ilmu Sosial	KD		
5	Bahasa	KD		
6	Ilmu Murni	KD		
7	Ilmu Terapan	KD		
8	Kesenian dan Olahraga	KD		
9	Kesusasteraan			BD
10	Sejarah dan Geografi	KD		

Keterangan:

KD : Kurang Diminati

D : Diminati

BD : Banyak Diminati

Klaster 1, yang terdiri dari kategori Karya Umum, Ilmu Sosial, Bahasa, Ilmu Murni, Ilmu Terapan, Kesenian dan Olahraga, serta Sejarah dan Geografi, dikategorikan sebagai kurang diminati. Hal ini kemungkinan disebabkan oleh sifat buku-buku dalam kategori ini yang dianggap terlalu umum, teknis, atau kurang relevan dengan kebutuhan sehari-hari masyarakat. Misalnya, buku-buku ilmu sosial dan ilmu murni mungkin lebih menarik bagi kalangan akademisi, tetapi

kurang diminati oleh peminjam umum. Untuk meningkatkan minat terhadap kategori ini, perpustakaan dapat melakukan promosi lebih gencar, menambah variasi judul yang lebih menarik, dan mengadakan acara seperti diskusi buku atau workshop yang mengaitkan topik-topik ini dengan tren terkini.

Klaster 2, yang mencakup kategori Filsafat dan Psikologi serta Agama, dikategorikan sebagai diminati. Kategori ini menarik minat peminjam karena topiknya yang berkaitan dengan pengembangan diri, pemahaman diri, dan spiritualitas. Buku-buku filsafat dan psikologi sering dicari oleh mereka yang ingin memahami diri sendiri atau mencari solusi untuk masalah pribadi, sementara buku agama selalu diminati karena relevansinya dengan kehidupan beragama. Untuk mempertahankan minat ini, perpustakaan dapat menambah koleksi buku terbaru dalam kategori ini dan mengadakan acara seperti seminar atau kajian keagamaan yang melibatkan buku-buku tersebut.

Klaster 3, yang hanya terdiri dari kategori Kesusasteraan, merupakan klaster yang paling banyak diminati. Kategori ini menarik minat peminjam karena sifatnya yang menghibur dan mudah dinikmati oleh berbagai kalangan. Novel, cerpen, dan puisi sering kali menjadi pilihan utama pembaca karena alur cerita yang menarik dan tema yang relevan dengan kehidupan sehari-hari. Untuk mempertahankan minat yang tinggi

terhadap kategori ini, perpustakaan dapat menambah koleksi buku sastra terbaru, mengadakan acara literasi seperti klub buku atau bedah novel, serta mempromosikan buku-buku sastra melalui media sosial atau display khusus di perpustakaan.

5. Kesimpulan

Pada *clustering* dengan nilai $K = 3$, hasil menunjukkan adanya tiga kelompok minat terhadap kategori buku di Dinas Arsip dan Perpustakaan Kabupaten Sragen: kategori yang "Kurang Diminati" (KD) seperti Karya Umum, Ilmu Sosial, Bahasa, Ilmu Murni, Ilmu Terapan, Kesenian dan Olahraga, serta Sejarah dan Geografi; kategori yang "Diminati" (D) seperti Filsafat dan Psikologi serta Agama; dan kategori yang "Banyak Diminati" (BD) yaitu Kesusasteraan. Dengan segmentasi ini, pengelola perpustakaan dapat lebih tepat dalam merencanakan pengadaan buku berdasarkan minat pemustaka yang beragam, sehingga alokasi anggaran dapat difokuskan pada kategori buku yang memiliki tingkat permintaan tinggi. Penerapan algoritma K-Means terbukti efektif dalam menyajikan hasil segmentasi yang akurat, dan metode evaluasi *clusterisasi* seperti Elbow, Silhouette, Gap Statistic, dan Dunn Index memperkuat pemilihan jumlah *cluster* optimal. Namun, untuk meningkatkan akurasi dan relevansi *clusterisasi* ini dalam jangka panjang,

penting untuk mempertimbangkan data tambahan, seperti tren peminjaman terbaru. Dengan demikian, pengelola dapat terus menyesuaikan pengadaan buku secara dinamis dan memenuhi kebutuhan informasi masyarakat secara lebih optimal.

6. Daftar Pustaka

- [1] N. Badriyah dan M. Walid, "Penentuan Bidang Minat Tugas akhir Mahasiswa Teknik Informatika Universitas Islam Madura Menggunakan Metode K-Means," *JINTEKS (Jurnal Informatika Teknologi dan Sains)*, vol. 5, no. 4, hlm. 566–572, 2023.
- [2] A. Saputra Dinata, "Penerapan Algoritma Naive Bayes Dalam Pengadaan Buku Referensi Pada Perpustakaan SMA Negeri 1 Trimurjo Berbasis Web," *Indonesian Journal of Science, Technology and Humanities*, vol. 1, no. 2, hlm. 80–90, 2023.
- [3] M. Iqbal, N. Khoirunnisa, dan M. I. Ramdhani, "Rancang Bangun Sistem Informasi Manajemen dan Klasterisasi Pengadaan Buku Perpustakaan Kabupaten Subang (SIPERSU) Menggunakan Algoritma K-Means," *Jurnal Review Pendidikan dan Pengajaran*, vol. 7, no. 3, hlm. 8508–8518, 2024.
- [4] P. Y. Mau, D. Chrisinta, dan E. Binsasi, "Implementasi Data Mining Dalam Menentukan Penambahan Koleksi Buku di Perpustakaan: Algoritma K-Means Clustering," *Journal of Mathematics, Computations, and Statistics*, vol. 7, no. 1, hlm. 122–132, 2024.
- [5] A. M. Nur, M. Saiful, H. Bahtiar, dan M. T. Hidayat, "Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi," *Infotek: Jurnal Informatika dan Teknologi*, vol. 7, no. 2, hlm. 478–488, Jul 2024, doi: 10.29408/jit.v7i2.26283.
- [6] Suhartini, L. K. Wijaya, dan N. A. Pratiwi, "Penerapan Algoritma K-Means Untuk Pendataan Obat Berdasarkan Laporan Bulanan Pada Dinas Kesehatan Kabupaten Lombok Timur," *Infotek : Jurnal Informatika dan Teknologi*, vol. 3, no. 2, hlm. 147–156, Jul 2020.
- [7] V. Putri, "Normalisasi Data Dengan Menggunakan Model Min Max Untuk Klasifikasi Nasabah Potensial Pada Bidang Pembelian Properti Menggunakan Algoritma K-Nearest Neighbor," *Informasi dan Teknologi Ilmiah (INTI)*, vol. 11, no. 3, hlm. 111–119, 2024.
- [8] F. A. Tanjung, A. P. Windarto, M. Fauzan, M. P. Studi, S. Informasi, dan S. Tunas Bangsa, "Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, vol. 6, no. 1, hlm. 61–74, 2021, [Daring]. Tersedia pada: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [9] A. Atira dan B. Nurina Sari, "Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Cluster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 17, hlm. 76–86, 2023, doi: 10.5281/zenodo.8282638.
- [10] N. Thamrin dan A. W. Wijayanto, "Analisis Cluster dengan Menggunakan Hard Clustering dan Soft Clustering untuk Pengelompokan Tingkat Kesejahteraan Kabupaten/Kota di Pulau Jawa," *Indonesian Journal of Statistics and Its Applications*, vol. 5, no. 1, hlm. 141–160, Mar 2021, doi: 10.29244/ijsa.v5i1p141-160.
- [11] M. S. 'Priadana dan D. 'Sunarsi, *Metode Penelitian Kuantitatif*. Tangerang: Pascal Books, 2021.

- [12] D. J. Sitanggang, E. Buulolo, dan S. Aripin, "Normalisasi Data Dengan Model Min Max Untuk Klasifikasi Calon Mahasiswa Yang Layak Memperoleh KIP Kuliah Dengan Algoritma K-Nearest Neighbor," *Informasi dan Teknologi Ilmiah (INTI)*, vol. 11, no. 3, hlm. 88–95, 2024.
- [13] Hidayatullah, S. Martha, dan S. Aprizkiyandari, "Analisis K-Means Menggunakan Metode Dunn Index Dalam Menentukan Jumlah Cluster Optimal (Studi Kasus: Indikator Pendidikan SMA di Indonesia Tahun 2022)," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, vol. 13, no. 3, hlm. 303–310, 2024.
- [14] M. T. Jatipaningrum, S. E. Azhari, dan K. Suryowati, "Pengelompokan Kabupaten Dan Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kesejahteraan Dengan Metode K-Means Dan Density-Based Spatial Clustering Of Applications With Noise," *Jurnal Derivat*, vol. 9, no. 1, hlm. 70–81, 2022.
- [15] D. Prasetyo, W. Lestati, dan V. Atina, "Penerapan Clustering Dengan K-Means Untuk Pemilihan Menu Favorit Di Tetra Coffeeshop," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 11, no. 3, hlm. 201–218, 2024..