IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING DALAM PEMETAAN KONDISI PERPUSTAKAAN DI INDONESIA BERDASARKAN PROVINSI TAHUN 2023

Suci Amalia, Amalia Sabila, Nur Annisa Basulina, Dika Prasetia,
M Pandu Wilantara, Ken Ditha Tania, Ahmad Rifai
Sistem Informasi, Universitas Sriwijaya
Jl. Raya Palembang - Prabumulih No.KM. 32, Indralaya Indah,
Kec. Indralaya, Kabupaten Ogan Ilir, Sumatera Selatan 30862, Indonesia
09031182227028@student.unsri.ac.id

ABSTRAK

Perpustakaan memiliki peran strategis dalam peningkatan literasi dan penyediaan akses informasi bagi masyarakat di seluruh Indonesia. Namun, penyebaran dan pengelolaan perpustakaan yang belum merata menyebabkan kesenjangan dalam akses dan kualitas layanan, khususnya di tingkat kabupaten/kota. Permasalahan ini diperburuk oleh kurangnya pemanfaatan data yang ada untuk mendukung pengambilan keputusan berbasis bukti. Penelitian ini bertujuan untuk mengelompokkan data perpustakaan kabupaten/kota di Indonesia berdasarkan karakteristik tertentu guna mengidentifikasi pola distribusi dan kesenjangan layanan perpustakaan. Metode yang digunakan adalah *Knowledge Discovery in Databases* (KDD) dengan pendekatan algoritma K-Means Clustering. Data yang digunakan bersumber dari Perpustakaan Nasional dan mencakup atribut seperti jumlah perpustakaan, tenaga perpustakaan, koleksi, anggota, dan anggaran. Hasil klasterisasi menunjukkan bahwa data perpustakaan dapat dikelompokkan menjadi enam klaster utama yang mana setiap klaster memiliki karakteristik berbeda, dari provinsi dengan layanan perpustakaan sangat tertinggal (Papua Pegunungan) hingga yang sangat unggul seperti DKI Jakarta. Sebagian besar provinsi berada dalam klaster dengan kebutuhan peningkatan standar dan koleksi. Visualisasi pairplot membantu memahami pola hubungan antar variabel. Temuan ini diharapkan dapat menjadi dasar dalam perumusan kebijakan yang lebih tepat sasaran dalam pengembangan layanan perpustakaan di Indonesia.

Kata kunci : *K-Means clustering; KDD, perpustakaan, penambangan data, pemetaan perpustakaan, standar perpustakaan*

1. PENDAHULUAN

Pengembangan perpustakaan di Indonesia saat ini menjadi fokus utama dalam upaya meningkatkan literasi dan akses informasi bagi masyarakat. Meskipun masih menghadapi tantangan seperti keterbatasan teknologi dan sumber daya, perpustakaan memainkan peran penting dalam menyediakan sumber daya berkualitas tinggi, program literasi, dan pelatihan untuk membantu pemustaka mengakses, mengevaluasi, dan memanfaatkan informasi [1]. Berdasarkan catatan Perpustakaan Nasional Republik Indonesia (Perpusnas) pada tahun 2024, terdapat peningkatan yang signifikan dalam Indeks Pembangunan Literasi Masyarakat (IPLM) yang mencapai 73,52. Angka ini melampaui target awal yang telah ditetapkan, yaitu sebesar 71,4. Terdapat perbedaan yang signifikan terkait jumlah perpustakaan yang terakreditasi pada setiap provinsi. Provinsi Jawa Timur menempati posisi teratas dengan 2.691 perpustakaan yang terakreditasi, sementara beberapa provinsi seperti Papua Tengah dan Papua Pegunungan belum memiliki perpustakaan yang terakreditasi sama sekali (Badan Pusat Statistik, 2023). Selain itu menurut Laporan Akhir Kajian Indeks Pembangunan Literasi Masyarakat (IPLM) tahun 2023, tingkat kunjungan perpustakaan juga masih belum merata antar wilayah. Tingkat kunjungan perpustakaan secara nasional mencapai 1,91%.

Saat ini, Indonesia masih menghadapi tantangan dalam melakukan pemetaan komprehensif terkait kondisi perpustakaan di seluruh provinsi [2]. Data perpustakaan di Indonesia berdasarkan tingkat sektoral pada tahun 2022 telah dipublikasi oleh Badan Pusat Statistik (BPS). Mahasiswa yang mengunjungi perpustakaan konvensional berjumlah 12,65%, sedangkan yang mengunjungi perpustakaan digital berjumlah 4,27%. Selain itu, identifikasi provinsi dengan kondisi perpustakaan yang memerlukan perhatian khusus menjadi hal yang kompleks karena terbatasnya data yang dapat menggambarkan keadaan secara menyeluruh. Dari 11.000 lebih perpustakaan di Indonesia, lebih dari 3.000 di antaranya terletak di daerah dengan akses terbatas dan kualitas layanan yang rendah, terutama di provinsi-provinsi dengan angka literasi rendah seperti Papua, NTT, dan Kalimantan Barat (Badan Pusat Statistik, 2021). Hal ini menunjukkan perlunya metode yang efisien untuk menganalisis dan mengelompokkan data perpustakaan secara objektif.

Metode seperti K-Means Clustering dapat digunakan untuk mengidentifikasi provinsi-provinsi dengan karakteristik serupa, seperti tingkat akses, jumlah koleksi, dan kualitas layanan, sehingga dapat meningkatkan kemudahan dalam pengalokasian sumber daya dan pengambilan keputusan yang lebih tepat sasaran [3]. Data perpustakaan di Indonesia, seperti jumlah perpustakaan, koleksi buku, rasio

pustakawan, standar layanan, hingga anggaran, menjadi elemen penting dalam mengukur kualitas dan aksesibilitas layanan literasi Informasi ini nantinya akan berperan dalam mengevaluasi efektivitas program literasi, mengidentifikasi kesenjangan sumber daya, dan mengembangkan kebijakan untuk meningkatkan minat baca serta pemerataan fasilitas pendidikan.

Algoritma K-Means Clustering merupakan salah satu metode pengelompokan data yang paling populer dan banyak digunakan dalam berbagai bidang. Metode ini bekerja dengan membagi data ke dalam beberapa kelompok atau cluster berdasarkan kemiripan karakteristik data tersebut, dengan menggunakan jarak Euclidean sebagai ukuran kedekatan [4]. Kelebihan utama dari K-Means adalah kemampuannya untuk mengidentifikasi pola tersembunyi dalam data yang besar, serta efisiensinya dalam mengelompokkan data dalam jumlah besar dengan waktu komputasi yang relatif cepat [4] [5]. Selain itu, K-Means juga dikenal karena kesederhanaannya dalam implementasi, sehingga mudah diaplikasikan dalam berbagai konteks, termasuk dalam analisis data spasial dan pengelompokan berdasarkan wilayah [6]. Relevansi K-Means dalam memetakan kondisi perpustakaan di Indonesia berdasarkan provinsi terletak pada kemampuannya untuk mengelompokkan data berdasarkan karakteristik tertentu. Hal ini memungkinkan identifikasi perbedaan kondisi perpustakaan di tiap provinsi, sehingga dapat menjadi dasar dalam pengambilan kebijakan untuk meningkatkan kualitas layanan perpustakaan secara nasional [5] [6].

Penelitian ini menerapkan algoritma K-Means Clustering bertujuan untuk mengelompokkan kondisi perpustakaan di Indonesia dengan menggunakan data tahun 2023. Penelitian ini didasarkan pada beberapa indikator, yaitu jumlah perpustakaan di setiap provinsi, persentase perpustakaan dengan kategori memenuhi standar pelayanan, jumlah koleksi yang tersedia, tingkat kecukupan koleksi sesuai dengan standar IFLA (International Federation of Library Associations and Institution), rasio kecukupan pustakawan, jumlah kunjungan perpustakaan dibandingkan jumlah penduduk, dan anggaran dari perpustakaan. Nasir (2021) melakukan penelitian yang hasilnya menunjukkan bahwa clustering K-Means efektif dalam menemukan pola peminjaman buku dan pengelolaan koleksi perpustakaan [7]. Diperkuat oleh temuan TBM Karya Mulya (2023) tentang pengelolaan koleksi tidak sistematis yang berdampak pada ketidakpatuhan IFLA [8]. Dari penelitian ini diharapkan bisa memberikan gambaran kepada pemangku kebijakan terkait dengan kondisi perpustakaan di berbagai wilayah. Hal tersebut, dapat diidentifikasi provinsi mana yang membutuhkan perhatian lebih serta merancang strategi atau solusi yang tepat dalam meningkatkan kualitas perpustakaan. Tujuan utama dari penelitian ini sendiri adalah menyediakan landasan yang kuat dalam proses pengambilan keputusan untuk meningkatkan pemerataan dan juga kualitas perpustakaan di Indonesia.

2. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Beberapa penelitian sebelumnya menunjukkan efektivitas algoritma K-Means Clustering dalam pengelompokan data peminjaman buku di perpustakaan. Nasir [9] menerapkan metode K-Means untuk mengelompokkan buku berdasarkan frekuensi peminjaman, yang terbukti membantu dalam pengadaan dan pengelolaan koleksi. Baker [10] menggunakan algoritma serupa untuk membentuk dua klaster, yaitu buku yang sering dipinjam dan yang jarang dipinjam, sehingga memudahkan pemantauan serta penambahan koleksi secara tepat.

Penelitian di Universitas 17 Agustus 1945 Surabaya [9] menggunakan K-Means untuk membangun sistem rekomendasi berdasarkan pengelompokan pengguna dari NIM. Hasilnya terbentuk 16 klaster dengan masing-masing rekomendasi tiga judul buku terpopuler, menunjukkan efektivitas algoritma dalam memahami pola minat baca dan personalisasi layanan perpustakaan.

Studi lain di SMA 5 Muhammadiyah [10] menerapkan K-Means dalam tiga klaster berdasarkan data peminjaman, dengan hasil evaluasi Silhouette Score sebesar 0,791, Davies-Bouldin Score 0,626, dan Calinski-Harabasz Score 615,093. Klaster yang terbentuk mencerminkan minat pembaca terhadap genre tertentu: religi (tinggi), drama dan romance (sedang), serta umum (rendah), yang mengonfirmasi akurasi dan stabilitas algoritma dalam segmentasi preferensi pengguna.

2.2 Perpustakaan

Perpustakaan berperan sebagai pusat informasi yang mendukung pendidikan dan riset, menghadapi tantangan dalam mengelola koleksi seiring kemajuan teknologi informasi. Ahmad Fairus Zabidi menunjukkan bahwa pengelompokan koleksi berdasarkan pola peminjaman dapat dilakukan dengan algoritma K-Means Clustering sebagai teknik penambangan data [11]. Metode ini membantu perpustakaan mengenali preferensi pengguna, mengoptimalkan pengadaan koleksi, dan memberikan rekomendasi yang lebih sesuai.

I Wayan Nada dan Made Hery W. Griadhi menekankan peran perpustakaan perguruan tinggi sebagai pusat informasi strategis untuk aktivitas akademik, penelitian, dan pengembangan kewirausahaan [12]. Perpustakaan modern tidak hanya menyediakan bahan pustaka tetapi juga berperan sebagai lokasi belajar, pelatihan literasi informasi, dan konsultasi bagi mahasiswa yang ingin mengembangkan potensi akademik dan wirausaha mereka.

2.3 K-Means

Algoritma K-Means terbukti efisien untuk mengelompokkan pola peminjaman buku berdasarkan frekuensi sirkulasi. Studi dengan 200 judul buku berhasil membuat tiga kluster: tinggi (8 item), sedang (11 item), dan rendah (17 item), menunjukkan konsistensi antara perhitungan manual dan algoritma [13]. Di Perpustakaan Saintek UINSU Medan, K-Means digunakan untuk menganalisis 290 data peminjaman, menghasilkan tiga kategori yang memungkinkan pendekatan terencana dalam pengelolaan koleksi [14].

Implementasi algoritma ini mempercepat penyusunan buku fisik dan mendukung pengambilan keputusan dalam pengadaan koleksi. Studi di Perpustakaan UNSIKA dengan evaluasi Davies Bouldin menunjukkan bahwa K-Means berperan meningkatkan ketersediaan buku populer dan mengurangi kelebihan stok buku kurang diminati [15]. Hasil serupa di berbagai institusi menunjukkan metode ini dapat meningkatkan kepuasan pengguna dan memudahkan pengelolaan inventaris perpustakaan secara keseluruhan.

Dalam konteks pemetaan kondisi perpustakaan di Indonesia berdasarkan provinsi, pendekatan K-Means Clustering menawarkan potensi besar untuk menganalisis distribusi, ketersediaan, dan pemanfaatan sumber daya perpustakaan secara nasional. Penerapan algoritma ini dapat menghasilkan gambaran komprehensif tentang kondisi perpustakaan di berbagai wilayah, sehingga membantu perumusan kebijakan pengembangan perpustakaan yang lebih terarah dan berbasis data.

3. METODE PENELITIAN

Dalam melakukan penelitian, berikut adalah tahapan yang dilakukan peneliti.

3.1. Knowledge Discovery Data

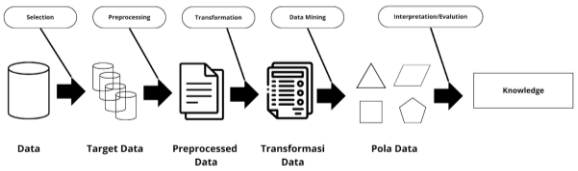
Penelitian ini menggunakan data dari Kaggle yang dikumpulkan pada Kajian Indeks Pembangunan Literasi Masyarakat (IPLM) 2023 oleh Perpustakaan Nasional Indonesia. Dataset ini menyajikan informasi mengenai keadaan serta statistik perpustakaan di seluruh provinsi Indonesia pada tahun 2023. Informasi yang diberikan telah diringkas agar lebih mudah diakses dan digunakan. Hal ini telah menghapus skor dan indikator IPLM yang memerlukan interpretasi yang rumit. Informasi atribut data yang akan digunakan dalam pemrosesan data mining ini meliputi beberapa yaitu sebagai berikut:

Tabel 1. Knowledge Discovery Atribut Data

Atribut Data	Informasi
Number of Libraries	Atribut data ini memberikan gambaran tentang distribusi perpustakaan di berbagai wilayah, yang dapat digunakan untuk menganalisis aksesibilitas layanan perpustakaan dan pemerataan sumber daya pendidikan
Percentage of Standardized Libraries	Atribut data ini menunjukkan proporsi perpustakaan di setiap provinsi yang memenuhi standar atau kriteria tertentu yang ditetapkan oleh lembaga Perpustakaan Nasional Republik Indonesia.
Number of Collections	Atribut data ini memberikan gambaran tentang ukuran dan kapasitas perpustakaan di masing-masing wilayah di Indonesia melalui jumlah buku atau materi perpustakaan yang dimiliki oleh perpustakaan di setiap provinsi di Indonesia
Library Collection Sufficiency	Atribut data menunjukkan rasio antara jumlah koleksi perpustakaan yang ada dengan perkiraan jumlah koleksi yang seharusnya dimiliki oleh perpustakaan di suatu provinsi, dinyatakan dalam persentase. Atribut ini mengukur seberapa memadai koleksi perpustakaan di suatu provinsi dibandingkan dengan kebutuhan idealnya. Semakin tinggi persentase, semakin memadai koleksi perpustakaan di provinsi tersebut
Librarian Sufficiency Ratio	Atribut data ini mengacu pada perbandingan antara jumlah pustakawan dengan jumlah penduduk atau pengguna yang mereka layani. Indikator ini digunakan untuk menilai apakah perpustakaan memiliki tenaga pustakawan yang cukup dalam memberikan layanan kepada masyarakat atau pengguna perpustakaan.
Library Visits	Atribut data ini mencerminkan seberapa sering perpustakaan dikunjungi oleh pengguna dalam periode tertentu. Indikator ini menunjukkan tingkat pemanfaatan perpustakaan oleh masyarakat dalam suatu wilayah, yang membantu dalam evaluasi efektivitas layanan, serta perencanaan peningkatan aksesibilitas dan fasilitas perpustakaan.
Library Budget	Atribut data ini memberikan informasi terkait alokasi dana operasional perpustakaan di setiap provinsi. Secara umum, data ini membantu mengevaluasi efektivitas pengelolaan perpustakaan dan mengidentifikasi daerah yang perlu ditingkatkan alokasi dananya.

Data dalam penelitian ini diperoleh melalui pendekatan sensus, di mana seluruh layanan perpustakaan daerah di Indonesia, mencakup 38 provinsi dan 514 kabupaten/kota, termasuk dalam penelitian ini. Perpustakaan Nasional (Perpusnas) Indonesia bisa memanfaatkan data yang telah dikumpulkan ini untuk merumuskan kebijakan strategis yang berhubungan dengan pengelolaan dan pengembangan perpustakaan. Sebelum melanjutkan ke tahap clustering untuk memproses informasi, langkah selanjutnya adalah melakukan proses *pra-processing* data. *Pra-processing* data meliputi serangkaian langkah untuk membersihkan, mengubah, dan menyiapkan data agar siap digunakan dalam

analisis clustering. Tahapan ini mencakup pembersihan, integrasi, transformasi, reduksi, dan diskritisasi data [22]. Lebih lanjut tahapan *Knowledge Discovery in Databases* (KDD) secara umum adalah pada gambar berikut [22] :



Gambar 1. Proses knowledge discovery data (KDD)

3.2. K-Means Clustering

Clustering adalah teknik yang digunakan untuk membedakan berbagai kumpulan data menjadi kelompok-kelompok yang dapat terlihat dari kecocokan yang diinginkan. Dalam data mining, Clustering adalah pengumpulan (grup) data tau objek dalam cluster yang memiliki kemungkinan mirip dengan aslinya serta setiap cluster data dapat dibedakan dari objek di cluster lain [16]. K-Means Clustering adalah algoritma yang dimanfaatkan untuk menjadikan sekumpulan data menjadi kelompok-kelompok dalam beberapa cluster berdasarkan kemiripan fitur atau atributnya. Proses metode ini dilakukan secara iteratif dengan menentukan sejumlah pusat cluster (centroid) dan mengelompokkan data tersebut berdasarkan kedekatannya dengan centroid tersebut [17]. Metode ini memiliki tujuan untuk meminimalkan jarak antara data dalam satu cluster dan memaksimalkan jarak antar cluster yang berbeda [18]. Tahapan dari algoritma K-Means adalah sebagai berikut:

1. Memilih secara acak sejumlah k titik data dari dataset sebagai centroid awal.
2. Pengisian: Menetapkan setiap titik data ke centroid terdekat dengan menggunakan rumus Euclidean Distance pada Persamaan 1.

$$d(x,c) = \sqrt{\sum_{i=1}^{n}(x_i - c_i)^2} \quad (1)$$

   Di sini d(*x,c)* menunjukkan jarak antara titik data *x* dan centroid *c*, *n* adalah jumlah dimensi (fitur) dalam dataset, *xi* adalah koordinat titik data *x* pada dimensi ke-*i* dan *ci* adalah nilai koordinat *centroid c pada dimensi ke-i*
3. Pembaruan *centroid:* Menghitung ulang centroid dengan mengambil nilai rata-rata dari semua titik data dalam setiap klaster menggunakan Persamaan 2.

$$c_i = \frac{1}{|C|}\sum_{x \in C} x_i \quad (2)$$

   Di sini *ci* adalah koordinat *centroid* pada dimensi ke-i, | C | adalah jumlah titik data dalam klaster *C*, dan *xi* adalah koordinat titik data *x* pada dimensi ke-*i*.
4. Perulangan: Mengulangi langkah 2 dan 3 hingga sampai terjadi konvergensi yaitu ketika *centroid* tidak mengalami perubahan signifikan.

3.3. Metode

Metode Elbow (siku) adalah teknik yang digunakan untuk menentukan jumlah klaster optimal dalam ruang data, khususnya dalam algoritma K-Means Clustering [14]. Tujuan utama metode ini adalah untuk menemukan titik di mana penambahan jumlah klaster tidak lagi memberikan penurunan yang signifikan dalam inersia (Within-Cluster Sum of Squares, WCSS) [15]. Secara formal, inersia atau WCSS dihitung sebagai jumlah kuadrat jarak antara setiap titik data dan centroidnya dalam klaster yang sesuai, dinyatakan pada Persamaan 3.

$$WCSS = \sum_{i=1}^{k}\sum_{x \in C_i} ||x - \mu_i||^2 \quad (3)$$

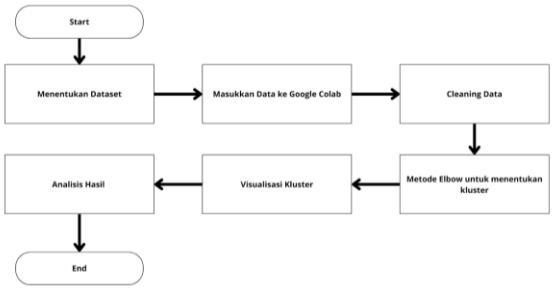
di mana:
K adalah jumlah klaster yang diuji,
Ci adalah klaster ke-*i*
x adalah titik data dalam klaster tersebut,
μi adalah centroid dari klaster ke-i,
|| *x* — | *μi* || ² adalah kuadrat jarak antara titik data *x* ke centroid *μi*.

3.4. Metode Evaluasi Performa Clustering

Setelah proses K-Means selesai, evaluasi dilakukan untuk mengukur kualitas clustering yang dihasilkan. Pada penelitian ini proses evaluasi dilakukan dengan menggunakan matrik penilaian Silhouette Score. Silhouette Score adalah metode evaluasi performa clustering yang digunakan untuk mengevaluasi seberapa baik setiap titik data cocok dengan klaster tempat mereka ditempatkan [21]. Skor ini mengukur seberapa dekat titik data terletak pada klaster yang ditugaskan dibandingkan dengan klaster lain. Semakin tinggi nilai Silhouette Score, semakin baik pembagian klaster. Bentuk Formal *Silhoute Score S* untuk data *i* dihitung pada Persamaan 4.

$$S(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \quad (4)$$

Di mana:
a(*i*) adalah rata-rata jarak antar titik data *i* dengan semua titik data lain dalam klaster yang sama.
b(*i*) adalah jarak rata-rata antara titi data *i* dengan semua titik data lain dengan klaster terdekat.
mmmmmm {*a*(*i*),*b*(*i*)} adalah nilai maksimum antara *a*(*i*) dan *b*(*i*) yang digunakan untuk memastikan nilai Silhouette Score berada dalam rentang -1 dan 1.
Secara detail setiap tahapan metode telah dijelaskan di atas, namun tahapan penelitian secara ringkas dapat dilihat juga pada gambar berikut [22]:



Gambar 2. Alur Penelitian

4. HASIL DAN PEMBAHASAN

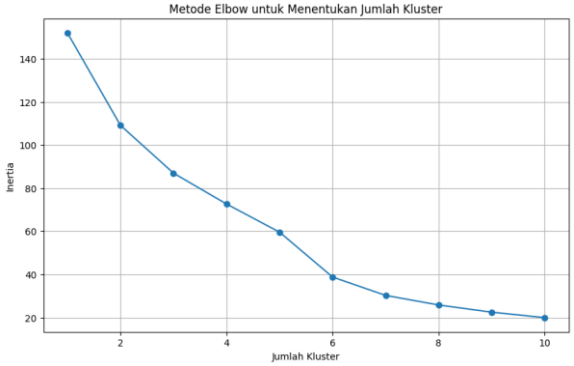
Setelah dilakukan proses penelitian, berikut adalah hasil dan pembahasan

4.1. *Data Processing*

Untuk mengoptimalkan penggunaan data dalam analisis klaster, beberapa langkah pengolahan dilaksanakan setelah data tersebut diperoleh. Tahap pertama adalah pemilihan fitur utama yang akan diterapkan dalam proses klasterisasi. Dari keseluruhan dataset, dipilih lima fitur yang dianggap paling relevan dalam menggambarkan kondisi perpustakaan di setiap provinsi. Peneliti memilih *Province, Number of Collections, Library Collection Sufficiency, Librarian Sufficiency Ratio, Library Visits, Percentage of Standardized Libraries* sebagai fitur yang akan digunakan. Hal ini dilakukan untuk dapat menggambarkan kondisi setiap perpustakaan di Indonesia berdasarkan informasi yang dibutuhkan.

Setelah fitur utama dipilih, dilakukan normalisasi menggunakan *StandardScaler* untuk menyamakan skala setiap variabel agar tidak ada fitur yang mendominasi hasil klasterisasi. Normalisasi ini memastikan bahwa bobot setiap fitur dalam perhitungan jarak antar data menjadi seimbang, sehingga hasil klasterisasi lebih akurat.

Setelah dilakukan proses pengoptimalan data, selanjutnya dilakukan penentuan jumlah cluster optimal dalam algoritma K-Means menggunakan metode elbow, dimana akan diproses klasterisasi pada berbagai jumlah klaster (1 hingga 10) dan menganalisis nilai inertia. Setelah dilakukan, hasilnya akan ditampilkan dalam grafik WCSS pada gambar 3 berikut:



Gambar 3. Hasil Grafik WCSS

Berdasarkan grafik WCSS pada data yang sudah dioptimalkan, peneliti menganalisis titik jumlah cluster optimal dimana terjadi menunjukkan penurunan *inertia* yang signifikan. Pada grafik tersebut, dapat dilihat bahwa jumlah klaster optimal berada pada 6 kluster. Setiap klaster memiliki rata-rata jumlah koleksi perpustakaan, kecukupan koleksi, rasio kecukupan pustakawan, jumlah kunjungan, dan persentase perpustakaan yang telah distandarkan.

4.2. Penerapan Algoritma K-Means

Setelah menetapkan jumlah cluster yang optimal, algoritma K-Means diterapkan untuk melaksanakan proses pengelompokan data. Dalam tahapan ini, setiap provinsi dikelompokkan ke dalam kluster yang memiliki ciri-ciri serupa berdasarkan atribut yang dianalisis. Analisis ringkasan dilakukan dengan mengelompokkan informasi sesuai dengan fungsi klaster guna memahami perbedaan antar klaster. Kemudian, dilakukan perhitungan agregasi statistik seperti rata-rata dan total data pada setiap klaster. Pada analisis ini, nilai rata-rata untuk fitur Persentase Perpustakaan Standar, Jumlah Koleksi, Kecukupan Koleksi Perpustakaan, Rasio Kecukupan Pustakawan, dan Kunjungan Perpustakaan dihitung. Jumlah data di setiap klaster dihitung berdasarkan nilai untuk fitur Persentase Perpustakaan Standarisasi. Setelah diproses menghasilkan hasil sebagai berikut:

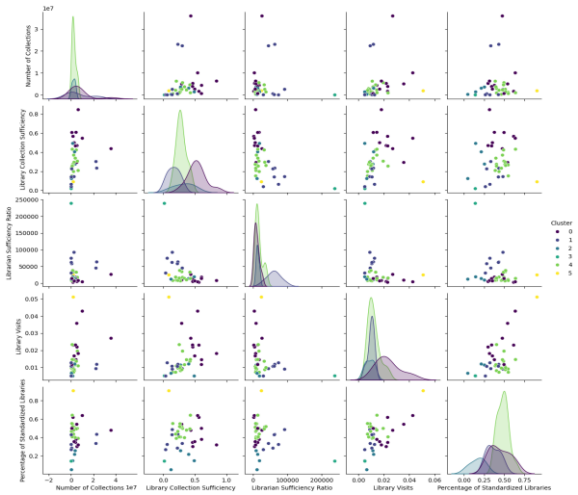
Tabel 2. Hasil Penerapan Algoritma K-Means

Parameter	Clus-ter 0	Clus-ter 1	Clus-ter 2	Clus-ter 3	Clus-ter 4	Clus-ter 5
% of Standardized Libraries (Mean)	446.240	364.057,00	165.425	140.800,00	482.600	906.800
Jumlah Provinsi (Count)	10	7,00	4	1,00	15	1
Number of Collections (Mean)	7,608,516	6,990,493	2,118,313	52,60	2,493,143	1,917,860
Library Collection Sufficiency (Mean)	533.220	163.786,00	323.500	18.400,00	285.093	89.800
Librarian Sufficiency Ratio (Mean)	10,906.6	60,988.3	13,689.8	238.41	16,585.3	24,838
Library Visits (Mean)	23.450	9.871,00	8.750	5.100,00	11.887	51.000

Berdasarkan hasil data pada tabel di atas, provinsi-provinsi di Indonesia telah dibagi menjadi enam kluster sesuai dengan karakteristik perpustakaannya. Persentase perpustakaan yang memenuhi kriteria, jumlah koleksi, kecukupan koleksi, rasio kecukupan pustakawan, serta kunjungan perpustakaan merupakan karakteristik yang dipakai. Masing-masing kluster menunjukkan tantangan dan kesempatan yang berbeda dalam pembangunan perpustakaan. Kemudian, dilakukan penelitian lebih lanjut terkait provinsi yang masuk pada setiap cluster dan akan dijelaskan sebagai berikut:

a. Kluster 0
   Kluster ini menunjukkan kinerja yang cukup baik dalam hal jumlah koleksi dan kecukupan koleksi. Namun, persentase perpustakaan yang memenuhi standar masih perlu ditingkatkan. Rasio pustakawan relatif rendah dibandingkan kluster lain, dan kunjungan perpustakaan sedang. Provinsi yang termasuk pada kluster ini adalah Sumatra Barat, Jambi, Sumatera Selatan, Kepulauan Bangka Belitung, DI Yogyakarta, Jawa Timur, Kalimantan Selatan, Sulawesi Selatan, Sulawesi Tenggara, Papua Barat.
b. Kluster 1
   Meskipun memiliki jumlah koleksi yang besar, kluster ini memiliki kecukupan koleksi yang sangat rendah, yang mengindikasikan bahwa koleksi mungkin tidak relevan atau usang. Rasio pustakawan sangat tinggi, tetapi tidak berkorelasi dengan kunjungan perpustakaan atau persentase perpustakaan standar. Provinsi yang termasuk pada kluster ini adalah Jawa Barat, Jawa Tengah, Banten, Maluku Utara, Papua Barat Daya, Papua Tengah, Papua Selatan.
c. Kluster 2
   Kluster ini menunjukkan kinerja yang rendah di semua area. Jumlah koleksi rendah, kecukupan koleksi sedang, rasio pustakawan sedang, dan kunjungan perpustakaan sangat rendah. Provinsi yang termasuk pada kluster ini adalah Riau, Kalimantan Timur, Sulawesi Barat, Papua.
d. Kluster 3
   Kluster ini sangat tertinggal. Jumlah koleksi dan kecukupan koleksi sangat rendah. Rasio pustakawan sangat tinggi, tetapi tidak berdampak positif pada kunjungan atau standar. Satu satunya provinsi yang termasuk pada kluster ini adalah Papua Pegunungan.
e. Kluster 4
   Kluster ini memiliki persentase perpustakaan yang memenuhi standar yang lumayan, tetapi jumlah koleksi dan kecukupan koleksi masih perlu ditingkatkan. Rasio pustakawan sedang, dan kunjungan perpustakaan rendah. Provinsi yang termasuk pada kluster ini adalah Aceh, Sumatera Utara, Bengkulu, Lampung, Kepulauan Riau, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Gorontalo, Maluku
f. Kluster 5
   Kluster ini sangat menonjol karena persentase perpustakaan yang memenuhi standar sangat tinggi dan kunjungan perpustakaan juga tinggi, meskipun jumlah koleksi dan kecukupan koleksi relatif rendah. Satu-satunya provinsi yang termasuk kedalam kluster ini adalah DKI Jakarta

Dalam proses analisis hasil klasterisasi ini, peneliti juga telah melakukan tahapan lanjutan untuk menampilkan hasil dalam bentuk visualisasi grafik pairplot seaborn.



Gambar 4. Hasil dalam Visualisasi Data Grafik Pairplot Seaborn

Gambar tersebut merupakan hasil visualisasi dari proses klasterisasi menggunakan algoritma K-Means yang bertujuan untuk memetakan kondisi perpustakaan di Indonesia berdasarkan provinsi pada tahun 2023. Visualisasi ini disajikan dalam bentuk pairplot menggunakan library Seaborn, yang memungkinkan untuk melihat hubungan antara berbagai metrik secara bersamaan. Dalam analisis ini, beberapa metrik yang digunakan antara lain Library Collection Sufficiency (kecukupan koleksi perpustakaan), Librarian Sufficiency Ratio (rasio kecukupan pustakawan), Library Visits (frekuensi kunjungan ke perpustakaan), dan Percentage of Standardized Libraries (persentase perpustakaan yang memenuhi standar). Setiap titik dalam grafik mewakili sebuah provinsi, dan warna atau bentuk titik tersebut menunjukkan klaster yang berbeda, yaitu kelompok provinsi dengan karakteristik serupa dalam hal kondisi perpustakaan.

Hasil klasterisasi ini memberikan gambaran tentang bagaimana provinsi-provinsi di Indonesia dapat dikelompokkan berdasarkan kondisi perpustakaan mereka. Misalnya, provinsi dengan koleksi perpustakaan yang lebih besar (Number of Collections 1e7) mungkin cenderung memiliki lebih banyak kunjungan atau rasio pustakawan yang lebih tinggi. Selain itu, visualisasi ini juga membantu mengidentifikasi pola atau korelasi antara metrik-metrik tersebut, seperti apakah provinsi dengan persentase perpustakaan standar yang tinggi juga memiliki kecukupan koleksi yang baik. Dengan demikian, analisis ini dapat menjadi dasar untuk pengambilan kebijakan dalam upaya meningkatkan kualitas dan aksesibilitas perpustakaan di berbagai provinsi di Indonesia. Judul penelitian, "Implementasi Algoritma K-Means Clustering dalam Pemetaan Kondisi Perpustakaan di Indonesia Berdasarkan Provinsi Tahun 2023", mencerminkan tujuan utama

dari studi ini, yaitu memberikan wawasan yang mendalam untuk perbaikan dan pengembangan perpustakaan di masa depan.

5. KESIMPULAN DAN SARAN

Tahapan penelitian yang dilakukan telah berhasil memetakan keadaan perpustakaan di Indonesia menurut provinsi dengan memanfaatkan algoritma K-Means clustering. Berdasarkan analisis yang dilakukan, terbentuk enam klaster yang masing-masing menunjukkan ciri-ciri khas perpustakaan di berbagai provinsi. Klaster 0 memperlihatkan kinerja baik dalam hal jumlah dan kecukupan koleksi, namun masih perlu perbaikan pada proporsi perpustakaan yang memenuhi standar. Klaster 1 memiliki koleksi yang luas tetapi relevansi koleksi tidak tinggi, meskipun jumlah pustakawannya banyak. Klaster 2 dan 3 menunjukkan performa yang sangat buruk di hampir semua segi, dengan Klaster 3 menjadi yang paling tertinggal. Klaster 4 memiliki proporsi perpustakaan yang layak, tetapi koleksi dan jumlah kunjungannya masih minim. Di sisi lain, Klaster 5 tampak mencolok dengan persentase standar perpustakaan dan kunjungan yang tinggi, meskipun koleksinya terbilang sedikit. Analisis ini memberikan wawasan penting bagi Lembaga Perpustakaan Nasional Indonesia (Perpusnas) dalam merumuskan kebijakan yang memperbaiki kualitas dan aksesibilitas perpustakaan nasional. Pada penelitian ini, peneliti melakukan analisis terhadap kluster yang menggambarkan faktor dan kondisi kondisi perpustakaan di Indonesia. Penelitian selanjutnya, diharapkan dapat melakukan penelitian mendalam terkait kluster tertentu untuk merumuskan strategi intervensi yang lebih spesifik dan efektif, serta memberikan wawasan lebih spesifik kepada Lembaga Perpusnas.

DAFTAR PUSTAKA

[1] P. A. Madya, "Peran Perpustakaan Dalam Meningkatkan Literasi Informasi Bagi Pemustaka I Nengah Sutrisna Jaya," *Msip*, Vol. 4, No. 2, Dec. 2024.

[2] I. N. N. Wasana, I. P. Suhartika, And R. T. Ginting, "Pemetaan Literasi Informasi Pustakawan Di Perpustakaan Nasional Republik Indonesia," *Jurnal Ilmiah Perpustakaan Dan Informasi (Jipus)*, Vol. 4, No. 1, 2024.

[3] M. R. Firmansyah And A. Habib, "Model Klasterisasi Dan Analisis Sistem Rekomendasi Kepada Pengguna Perpustakaan Universitas 17 Agustus 1945 Surabaya Menggunakan K-Means," *Jurnal Rekayasa Sistem Informasi Dan Teknologi*, Vol. 2, No. 2, Nov. 2024.

[4] B. Chong, "K-Means Clustering Algorithm: A Brief Review," *Academic Journal Of Computing & Information Science*, Vol. 4, No. 5, Pp. 37–40, 2021.

[5] L. Izzah And A. Jananto, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika," *Progresif: Jurnal Ilmiah Komputer*, Vol. 18, No. 1, Pp. 69–76, Feb. 2022.

[6] M. Rizki Nugroho, I. Edo Hendrawan, And Puwantoro, "Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit Asri," *Jurnal Nuansa Informatika*, Vol. 16, No. 1, Pp. 125–133, Jan. 2022.

[7] B. Krismarta And F. Septian, "Implementasi Algoritma K-Means Untuk Klasterisasi Minat Membaca Anak Pada Perpustakaan Sudut Baca Opera," *Oktal:Jurnal Ilmu Komputer Dan Science*, Vol. 3, No. 8, Aug. 2024.

[8] S. Yoanda And A. Gunaidi, "Pengolahan Koleksi Di Taman Baca Masyarakat Karya Mulya Kota Palembang Dalam Upaya Meningkatkan Temu Kembali Informasi," *Berkala Ilmu Perpustakaan Dan Informasi*, Vol. 19, No. 2, Pp. 195–207, Dec. 2023.

[9] M. Rizki Firmansyah And A. Habib, "Model Klasterisasi Dan Analisis Sistem Rekomendasi Kepada Pengguna Perpustakaan Universitas 17 Agustus 1945 Surabaya Menggunakan K-Means," *Jurnal Rekayasa Sistem Informasi Dan Teknologi*, Vol. 2, Pp. 744–750, Nov. 2024.

[10] Muhammad Sultan Reza Aditya Nurrahman, "Pengelompokan Data Peminjaman Buku Di Perpustakaan Sma 5 Muhammadiyah Dengan Menggunakan Metode K-Means Clustering," 2024.

[11] A. F. Zabidi, "Penerapan Algoritma K-Means Untuk Pengelompokan Koleksi Perpustakaan Dengan Data Mining," *Media Jurnal Informatika*, Vol. 16, No. 2, P. 233, Dec. 2024.

[12] I. Wayan Nada And M. W. Hery Griadhi, "Perpustakaan Sebagai Pusat Informasi Entrepreneurship," Msip, Vol. 4, No. 1, 2024.

[13] S. H. Haji, K. Jacksi, And R. M. Salah, "A Semantics-Based Clustering Approach For Online Laboratories Using K-Means And Hac Algorithms," *Mathematics*, Vol. 11, No. 3, Feb. 2023.

[14] M. Indah Tanjung, "Optimizing Library Book Circulation With K-Means Clustering: Insights From Uinsu Medan," *Journal Of Information Systems And Informatics*, Vol. 7, No. 1, 2025.

[15] S. H. Naibaho, Nailufar Farha Afifah, Yuyun Umaidah, And Nono Heryana, "Analysis Of Student Reading Interest In Unsika Library With K-Means Algorithm," *Antivirus : Jurnal Ilmiah Teknik Informatika*, Vol. 18, No. 1, Pp. 82–94, May 2024.

[16] N. A. Maori, "Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering," *Jurnal Simetris*, Vol. 14, Nov. 2023.

[17] Ramadhana, Islamiyah, And A. P. A. Masa, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara," *Adopsi Teknologi Dan Sistem*

Informasi (Atasi), Vol. 2, No. 1, Pp. 35–42, Jun. 2023.

[18] M. Sholeh And K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow Dan Silhouette Pada Model Clustering Dengan Menggunakan Algoritma K Means," Aug. 2023.

[19] M. R. Syahkur, D. Hartama, And Solikhun, "Evaluasi Jumlah Cluster Pada Algoritma K-Means++ Menggunakan Silhouette Dan Elbow Dengan Validasi Nilai Dbi Dalam Mengelompokkan Gizi Balita," *Jst (Jurnal Sains Dan Teknologi)*, Vol. 13, No. 3, Pp. 487–496, Oct. 2024.

[20] W. I. Permatasari And V. Tundjungsari, "Implementasi Algoritma K-Means Untuk Mengetahui Minat Siswa Sma Terhadap Mata Pelajaran Teknologi Informasi Dan Komunikasi," *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 5, No. 1, Pp. 168–174, Dec. 2024.

[21] S. P. Ananda, L. Muthoharoh, And S. D. P. Pratama, "Analisis K-Means Clustering Angkatan Kerja Berdasarkan Jenis Kelamin, Usia, Pendidikan, Dan Pendapatan Pada Data Satuan Kerja Angkatan Nasional Tahun 2021," *Prosiding Seminar Nasional Sains Dan Teknologi Seri Iii Fakultas Sains Dan Teknologi*, Vol. 2, No. 1, 2025.

[22] L. H. Annisa And D. Rusvinasari, "Segmentasi Pembelian Produk Menggunakan Algoritma K-Means Berdasarkan Clusterisasi Pada Pemilihan Menu Yang Ada Diumkm Kuliner," *Jurnal Informatika: Jurnal Pengembangan It*, Vol. 9, No. 3, Pp. 203–212, Dec. 2024.