



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 4 (2026) pp: 8394-8404

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Penerapan *Data Mining* Dalam Mengelompokan Buku Perpustakaan Menggunakan Algoritma *K-Means*

Yulia

Teknik Informatika, Universitas Islam Sumatera Barat

yuliaedwar2407@gmail.com

Abstrak

Perpustakaan sebagai sarana sumber informasi dan ilmu pengetahuan untuk menyimpan bahan pustaka yang dipakai oleh pemakai untuk menggali ilmu sumber informasi. Penelitian dilakukan pada salah satu Perpustakaan yang ada di kota Batam. Perpustakaan ini memiliki beragam koleksi buku seperti buku umum, karya ilmiah, bahasa, sejarah, dan lain sebagainya. Permasalahan yang sering terjadi adalah buku yang dipinjam kadang tidak tersedia, selain itu pihak perpustakaan juga mengalami kesulitan karena tidak tau berapa jumlah buku yang dipinjam sehingga pihak Perpustakaan melihat kembali catatan transaksi meminjam buku pada buku tamu. Untuk itu dibuatlah suatu sistem dengan pengolahan jumlah data yang besar dengan teknik data mining metode *k-means*. Data peminjam buku yang diolah dari tahun 2021-2024, dengan jumlah data sebanyak 36 diambil dari jumlah peminjam buku. Variabel yang dijadikan input yaitu buku karya umum, buku filsafat, buku agama, dan buku bahasa. Dari hasil yang diperoleh data peminjam buku yang telah diproses mendapatkan buku yang banyak dipinjam terdapat pada cluster 1 sebanyak 9 item, buku yang paling sedikit dipinjam terdapat pada cluster 2 sebanyak 15 item, buku yang cukup banyak dipinjam terdapat pada cluster 0 sebanyak 12 item. Hasil yang diperoleh antara hitungan manual dan RapidMiner dari data peminjam buku adalah sama dengan hasil yang telah diproses maka didapatkan jumlah buku yang dipinjam.

Kata kunci: Data mining, Buku, *K-Means*, Perpustakaan.

1. Latar Belakang

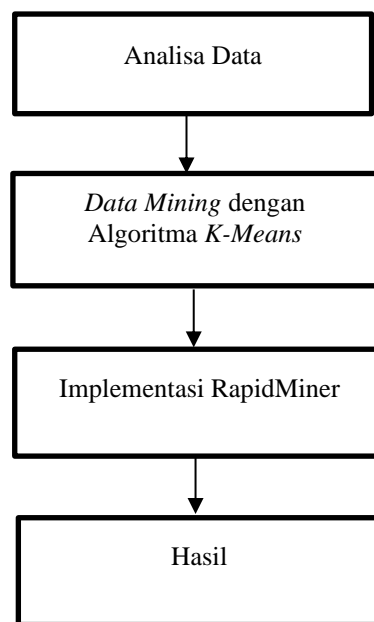
Pada perkembangan teknologi sekarang khususnya teknologi informasi dalam pengolahan data membuat masyarakat dituntut untuk menimba pengetahuan dan meningkatkan kemampuan dibidang teknologi dalam pengolahan data. Kebutuhan informasi menjadi suatu elemen penting dalam pengambilan keputusan [1]. Kadang perkembangan teknologi tersebut masyarakat harus dapat menyajikan informasi yang dibutuhkan harus cepat dan tepat. Permasalahan yang terjadi kebutuhan informasi tadi kadang tidak sesuai dengan yang diharapkan. Seperti ketika informasi yang dibutuhkan lambat dan perlu mencari file-file yang tersimpan dengan jumlahnya sangat banyak [2]. Penelitian ini dilakukan pada salah satu Perpustakaan yang ada di kota Batam, seperti yang terjadi pada sebuah perpustakaan di kota Batam, khususnya Perpustakaan Masjid Jabal Arafah kota Batam.

Perpustakaan sebagai wadah untuk menimba ilmu bagi semua kalangan baik muda maupun tua. Perpustakaan sebagai sarana sumber informasi dan ilmu pengetahuan untuk menyimpan bahan pustaka yang dipakai oleh pemakai untuk menggali ilmu sumber informasi [3]. Perpustakaan Masjid Jabal Arafah kota Batam ini memiliki luas area 50 m², dirancang agar pengunjung merasa nyaman saat membaca dan meminjam buku. Perpustakaan ini merupakan perpustakaan ke dua terbesar di kota Batam setelah perpustakaan daerah. Perpustakaan ini memiliki beragam koleksi buku yang tersedia dengan lengkap, seperti buku karya umum, filsafat, agama, bahasa, psikologi, buku islami dan lain sebagainya. Banyaknya masyarakat yang datang ke Perpustakaan ini untuk membaca dan meminjam buku dikarenakan memiliki koleksi buku yang cukup lengkap. Transaksi peminjam buku terjadi setiap hari dari buku yang dipinjam. Permasalahan yang sering terjadi adalah buku yang dipinjam kadang tidak ada, karena jumlah buku yang dipinjam tidak sebanding dengan stok buku yang ada. Hal ini akan berpengaruh kurangnya minat baca masyarakat untuk membaca. Selain itu pihak Perpustakaan juga mengalami kesulitan karena tidak tau berapa banyaknya jumlah buku yang dipinjam dalam tiap bulannya. Untuk mengetahui banyaknya jumlah buku yang dipinjam petugas Perpustakaan melihat kembali catatan transaksi meminjam pada daftar tamu pengunjung. Selain itu petugas Perpustakaan tidak dapat mengetahui pengelompokan buku mana yang banyak dipinjam, petugas perpustakaan hanya membuat catatan peminjam buku dalam pembuatan laporan bulanan, serta berapa banyak jumlah buku yang dipinjam dengan buku yang tersedia. Untuk itu dibuatlah suatu implementasi sistem dengan pengolahan jumlah data yang besar dengan teknik *data mining* dengan metode *k-means* agar bisa buku mengetahui buku mana yang banyak dipinjam dan mengelompokan buku.

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang dihasilkan dengan cara mengekstraksi dan mengenali pola yang terdapat dalam basis data. *Data mining* digunakan untuk mencari pengetahuan dalam jumlah data yang besar atau disebut dengan *Knowledge Discovery in Database (KDD)*, yang artinya KDD berhubungan dengan pola-pola sejumlah data yang besar dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi [4]. *Clustering* merupakan suatu metode dalam tambang data yang unsupervised, karena tidak ada satu atribut pun yang digunakan untuk memandu proses pembelajaran, jadi seluruh atribut input diperlakukan sama [5]. Teknik *clustering* meminimalisir suatu objek data yang mirip dan mengklaster data secara ilmiah [6]. Teknik *data mining* dengan metode *k-means* bertujuan agar pihak Perpustakaan dapat mengetahui pengelompokan buku antara buku yang dipinjam diperoleh dari transaksi peminjam buku dari daftar buku pengunjung agar dapat mengetahui buku yang banyak dipinjam sehingga pihak Perpustakaan dapat menambah jumlah koleksi buku tersebut. Dengan adanya persediaan buku yang lengkap diharapkan menambah minat baca masyarakat di Perpustakaan tersebut [7].

2. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode *clustering* algoritma *k-means*. Adapun tahapan dalam penelitian ini sebagai berikut :



Gambar 1. Tahapan Penelitian

Analisa data

Analisa data merupakan tahap dimana data yang diperoleh dari pengumpulan data. Data dikumpulkan dari transaksi peminjam buku tahun 2021-2024. Data tersebut diseleksi dari banyaknya buku yang dipinjam dari jenis buku karya umum, filsafat, agama dan bahasa dengan jumlah data sebanyak 36 diambil dari jumlah peminjam tiap bulanya.

Data mining dengan Algoritma *K-Means*

Data mining merupakan sebuah tambang data dalam mengolah data yang cukup besar. Dalam penelitian ini menggunakan metode *clustering* dengan algoritma *k-means* sebagai pengelompokan data [8]. Tahapan dalam algoritma *k-means* yaitu menentukan jumlah *clusternya* terlebih dahulu. Dalam penelitian ini ditentukan sebanyak 3 *cluster*.

Implementasi RapidMiner

Implementasi sistem merupakan tahap setelah data diolah dan dianalisa dengan algoritma *k-means* maka tahap diimplementasikan dengan aplikasi RapidMiner. Implementasi RapidMiner ini mencari kecocokan hasil yang diperoleh dengan algoritma *k-means* dengan implementasi RapidMiner, sehingga diketahui hasil pengelompokan hasil *cluster* yang diperoleh sama [9].

3. Hasil dan Diskusi

Adapun tahapan dalam data mining menggunakan algoritma *K-Means* [10]:

1. Seleksi Data

Tahapa awal yang dilakukan adalah seleksi data. Data peminjaman buku yang akan diolah tahun 2021-2024, dengan jumlah data sebanyak 36 diambil dari jumlah peminjam tiap bulanya.

2. *Cleaning Data*

Setelah data diseleksi tahap selanjutnya adalah pembersihan data, hasil dari seleksi data tersebut berjumlah 36 data, dimana nama peminjam, judul buku yang dipinjam dihilangkan karena yang akan menjadi variabel input adalah karya umum, filsafat, agama dan bahasa.

3. Perhitungan Algoritma

Algoritma *K-Means* adalah algoritma yang digunakan oleh peneliti dalam pengelompokan data. Berikut adalah proses dari algoritma *k-means* [11] :

a. Menentukan Jumlah *Cluster*

Dalam pengelompokan ini, jumlah *cluster* yang telah ditentukan sebanyak 3 *cluster*, sehingga dapat ditentukan untuk nilai *k* adalah $k = 3$.

b. Menentukan titik pusat *Cluster*

Tabel 1. Titik Pusat *Cluster*

c0	Ferbruari 2018	14	20	35	15
c1	Mei 2019	35	24	21	9
c2	Maret 2020	19	31	12	10

Menentukan titik pusat cluster secara acak dari data/objek yang tersedia sebanyak jumlah *cluster* *k* yang telah ditentukan.

c. Hitung jarak setiap data yang ada terhadap setiap pusat *cluster*

Rumus yang digunakan untuk menghitung jarak setiap data terhadap setiap pusat *cluster* adalah *Euclidian Distance* untuk melakukan perhitungan jarak setiap data terhadap titik pusat *cluster* hingga ditemukan jarak paling dekat dari setiap data centroid. Berikut adalah persamaan *Euclidian Distance* [12].

$$d(x_i, u_i) = \sqrt{(x_i - u_i)^2} \quad (1)$$

Keterangan:

$d(x,y)$ yaitu jarak data x dengan y

$x_i = (x1,x2,\dots,xi)$ yaitu variabel data

$y_i = (y1,y2,\dots,yi)$ yaitu variabel pada *centroid*

d. Alokasikan masing-masing data ke dalam *centroid* yang paling terdekat. Setelah menghitung jarak setiap data yang ada terhadap setiap pusat *cluster*, maka langkah selanjutnya adalah alokasikan masing-masing objek ke *centroid* terdekat, dengan cara mengelompokkan berdasarkan jarak minimum objek ke pusat *cluster* [13], yaitu sebagai berikut:

Tabel 2. Perhitungan Jarak Data Iterasi 1

	$c0$	$c1$	$c2$		
k1	11	19	25	1	
k2	0	26	26	1	
k3	23	28	20		1
k4	16	28	28	1	
k5	13	30	22	1	
k6	8	27	22	1	
k7	18	23	12		1
k8	19	26	15		1
k9	16	21	11		1
k10	14	28	23	1	
k11	25	21	17		1
k12	14	18	12		1
k13	31	12	19	1	
k14	20	16	8		1
k15	15	22	13		1
k16	22	11	9		1
k17	24	16	17	1	
k18	18	8	20	1	
k19	14	29	24	1	
k20	19	18	10		1
k21	17	12	14	1	
k22	27	32	14		1
k23	18	24	24	1	
k24	16	22	12		1
k25	22	29	16		1
k26	24	23	22		1
k27	26	20	0		1
k28	8	20	19	1	
k29	26	0	20	1	
k30	21	21	22	1	
k31	15	27	20	1	
k32	25	15	25	1	
k33	24	26	20		1
k34	15	22	16	1	
k35	11	28	27	1	
k36	8	21	23	1	

- e. Lakukan iterasi dan kemudian tentukan posisi $C0$ baru dengan cara menghitung rata-rata dari data-data yang berada pada *centroid* yang sama [14]. Kemudian kita tentukan lagi pusat *cluster* dari data yang baru, caranya dengan menjumlahkan semua nilai yang merupakan anggota dari *cluster* dan dibagi total jumlah anggota *cluster* [15].

Cluster0 = 1,2,4,5,6,10,19,23,28,31,34,35,36

Cluster1 = 13,16,17,18,21,26,29,30,32

Cluster2 = 3,7,8,9,11,12,14,15,20,22,24,25,27,33,34

Berikut adalah perhitungan C0 baru sebagai berikut:

$C0,1 = (22+14+19+8+10+12+10+18+17+22+14+12+17)/13 = 14$
 $C0,2 = (15+20+33+30+23+15+16+11+22+18+15+11+21)/13 = 19$
 $C0,3 = (30+35+37+31+29+23+24+21+28+21+20+29+32)/13 = 28$
 $C0,4 = (14+15+23+12+17+21+21+19+12+20+13+12+8)/13 = 16$

$C1,1 = (34+26+30+24+35+23+33)/7 = 29$

$C1,2 = (21+21+23+28+24+16+18)/7 = 22$

$C1,3 = (12+15+27+25+21+19+21)/7 = 20$

$C1,4 = (16+21+12+10+9+24+23)/7 = 16$

$C2,1 = (10+12+9+15+19+17+21+13+25+18+5+14+10+21+19+13)/16 = 15$

$C2,2 = (14+23+21+27+14+26+31+24+27+22+29+30+38+12+31+21)/16 = 24$

$C2,3 = (15+18+17+21+12+23+20+21+18+17+12+23+23+14+12+13)/16 = 17$

$C2,4 = (7+12+10+14+10+12+12+13+10+9+8+12+12+20+10+8)/16 = 11$

Tabel 3. Cluster Baru Iterasi ke-1

c0	14	19	28	16
c1	29	22	20	16
c2	15	24	17	11

- f. Ulangi langkah ke-3 hingga nilai *centroid* tidak mengalami perubahan. Karena proses yang dilakukan baru pada iterasi ke-0, maka perlu dilakukan beberapa iterasi lagi untuk dapat membandingkan nilai dari dua iterasi terakhir. Jika nilai dua iterasi terakhir tersebut telah sama, maka proses iterasi telah selesai, dan jika tidak maka ulangi lagi langkah berikutnya [16].

Tabel 4. Hasil Perhitungan Jarak Dan Pengelompokan Data Iterasi ke-2

	c0	c1	c2	
k1	10	14	17	1
k2	7	22	18	1
k3	17	23	12	1
k4	19	24	25	1
k5	13	26	17	1
k6	6	21	14	1
k7	11	18	3	1
k8	13	21	7	1
k9	10	16	6	1
k10	9	19	15	1
k11	18	16	12	1
k12	9	14	6	1

DOI: <https://doi.org/10.31004/riggs.v4i4.4772>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

k13	25	9	20		1	
k14	16	13	10			1
k15	9	17	5			1
k16	17	10	11		1	
k17	18	8	15		1	
k18	17	8	18		1	
k19	8	21	15	1		
k20	14	14	4			1
k21	15	12	12		1	
k22	22	28	13			1
k23	12	16	16	1		
k24	12	18	8			1
k25	20	26	16			1
k26	18	14	16		1	
k27	21	17	10			1
k28	5	15	11	1		
k29	23	10	20		1	
k30	15	11	17		1	
k31	9	19	12	1		
k32	21	8	23		1	
k33	18	22	13			1
k34	8	16	7			1
k35	10	23	18	1		
k36	10	19	15	1		

Pada tabel iterasi-1 dan iterasi-2 masih belum terlihat kesamaan pada setiap nilai *cluster* yang diperoleh, oleh sebab itu harus kembali mengulangi langkah ke-3 hingga nilai *centroid* tidak mengalami perubahan. Maka *C0* barunya adalah sebagai berikut :

$$Cluster0 = 1,2,4,5,6,10,8,23,28,31,35,36$$

$$Cluster1 = 13,16,17,18,21,26,29,30,32$$

$$Cluster2 = 3,7,8,9,11,12,14,15,20,22,24,25,27,33,34$$

Berikut adalah perhitungan *C0* baru sebagai berikut:

$$C0,1 = (22+14+19+8+10+12+10+18+17+11+12+17)/12 = 14$$

$$C0,2 = (15+20+33+30+23+15+16+18+22+18+11+21)/12 = 20$$

$$C0,3 = (30+35+37+31+29+23+24+21+28+21+29+32)/12 = 28$$

$$C0,4 = (14+15+23+12+17+21+21+19+12+20+12+8)/12 = 16$$

$$C1,1 = (34+25+26+30+24+35+23+33+21)/9 = 28$$

$$C1,2 = (21+21+23+28+24+16+18+27+12)/9 = 21$$

$$C1,3 = (12+15+27+25+21+19+21+18+14)/9 = 19$$

$$C1,4 = (16+21+12+10+9+24+23+10+20)/9 = 16$$

$$C2,1 = (10+12+9+15+19+17+21+13+18+5+14+10+21+19+13)/15 = 14$$

$$C2,2 = (14+23+21+27+14+26+31+24+22+29+30+38+12+31+21)/15 = 24$$

$$C2,3 = (15+18+17+21+12+23+20+21+17+12+23+23+14+12+13)/15 = 18$$

$$C2,4 = (7+12+10+14+10+12+12+13+9+8+12+12+20+10+8)/15 = 11$$

DOI: <https://doi.org/10.31004/riggs.v4i4.4772>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Dari perhitungan pusat cluster yang telah diperoleh di atas maka akan mendapatkan *centroid* baru untuk iterasi-2 yaitu sebagai berikut:

Tabel 5. *Centroid* Baru iterasi ke -2

c0	14	20	28	16
c1	28	21	19	16
c2	14	24	18	11

Karena proses yang dilakukan pada iterasi ke-2 masih belum mendapat hasil yang sama dengan iterasi sebelumnya (iterasi-1), maka perlu melakukan beberapa iterasi lagi untuk dapat membandingkan nilai dari dua iterasi terakhir. Jika nilai dua iterasi terakhir tersebut telah sama, maka proses iterasi dinyatakan telah selesai, tetapi jika masih belum mendapatkan hasil yang sama dengan sebelumnya maka ulangi lagi langkah kembali [17]. Berikut adalah tabel hasil perhitungan jarak pengelompokan data dengan iterasi-3 adalah:

Tabel 6. Hasil Perhitungan Jarak Dan Pengelompokan Data Iterasi ke-2

k1	9	14	17	1		
k2	7	21	18	1		
k3	18	22	12	1		
k4	18	24	25	1		
k5	13	25	16	1		
k6	5	21	13	1		
k7	12	17	2			1
k8	14	20	6			1
k9	11	14	6			1
k10	9	18	15	1		
k11	19	15	13			1
k12	10	13	6			1
k13	26	9	22		1	
k14	16	13	10			1
k15	9	16	4			1
k16	18	9	11		1	
k17	19	7	16		1	
k18	17	9	19		1	
k19	9	20	15	1		
k20	14	12	5			1
k21	15	12	13		1	
k22	22	27	12			1
k23	12	15	16	1		
k24	12	17	8			1
k25	20	25	15			1
k26	18	13	17		1	
k27	21	16	10			1
k28	6	15	11	1		
k29	24	11	21		1	

1		
1		
		1
1		
1		
1		
		1
		1
		1
1		
		1
		1
	1	
		1
		1
	1	
1		
		1
	1	
		1
1		
		1
		1
	1	
1		
	1	
	1	
1		

Tabel 8. Hasil pengelompokan Data Dari Iterasi Ke-3

1		
1		
		1
1		
1		
1		
		1
		1
		1

- Memiliki Fasilitas Sekolah Menurut Provinsi Berdasarkan Tingkat Pendidikan,” 2018, pp. 52–57.
- [4] D. Siburian, S. R. Andani, and I. P. Sari, “Implementasi Algoritma K-Means untuk Pengelompokan Peminjaman Buku Pada Perpustakaan Sekolah,” *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 2, pp. 115–124, 2022, doi: 10.55123/jomlai.v1i2.725.
- [5] H. L. Sari and I. Y. Beti, “Penerapan Data Mining Dalam Pengelompokan Buku Yang Dipinjam Menggunakan Algoritma K-Means,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, pp. 925–933, 2023, doi: 10.30865/klik.v3i6.826.
- [6] D. Nanda, A. Pratama, and A. Hidayat, “Penerapan Text Mining Dalam Klasterisasi Buku Di Perpustakaan Dengan Algoritma K-Means Clustering (Studi Kasus : Perpustakaan Universitas Nahdlatul Ulama Kalimantan Barat),” *J. Komput. Multidipliner*, vol. 7, no. 3, pp. 35–43, 2024.
- [7] A. H. Kusna, I. H. Santi, and F. Febrinita, “Pengelompokan Data Kelas Desa Berdasarkan Data Letter - C Menggunakan Algoritma K-Means Clustering,” *J. Inform. Polinema*, vol. 11, no. 2, pp. 127–136, 2025.
- [8] B. Krismarta and F. Septian, “Implementasi Algoritma K-Means Untuk Klasterisasi Minat Membaca Anak Pada Perpustakaan Sudut Baca Opera,” *OKTAL J. Ilmu Komput. dan Sci.*, vol. 3, no. 8, pp. 2022–2031, 2024.
- [9] F. Handayani, “Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar,” *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.
- [10] V. Darsono, A. Amroni, and A. Andrianti, “Penerapan Data Mining Algoritma K-Means Untuk Rekomendasi Pemilihan Bidang Studi Perguruan Tinggi Pada Siswa SMKN 1 Kota Jambi Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM),” *J. Inform. dan Rekayasa Komput.*, vol. 2, no. September, pp. 161–171, 2022.
- [11] N. Nuur Hasanah and A. Sidiq Purnomo, “Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus : Perpustakaan Politeknik LPP Yogyakarta),” *J. Teknol. dan Sist. Inf. Bisnis*, vol. 4, no. 2, pp. 300–311, 2022.
- [12] M. O. Harefa and S. Aripin, “Penerapan Metode K-Means Dalam Pengelompokan Buku Untuk Menentukan Minat Baca Pada Perpustakaan Daerah Kota Medan,” *Bul. Artif. Intell.*, vol. 3, no. 1, pp. 26–34, 2024.
- [13] J. Fadilina, R. Utami, and N. Marsan Sitinjak, “Pengelompokan Buku Dan Rekomendasi Buku Menggunakan K-Means Clustering Pada Dinas Perpustakaan Dan Kearsipan Kota Medan,” *J. Widya*, vol. 5, no. 2, pp. 1045–1058, 2024.
- [14] U. Surapati and M. Jannah, “Penerapan Data Mining Menggunakan Metode K-Means Untuk Mengetahui Minat Customer Dalam Pembelian Merchandise Kpop,” *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 791–800, 2024.
- [15] S. Surya, M. Sholeh, and D. Andayati, “Penerapan Metode Clustering dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa,” *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 6, no. 1, pp. 51–60, 2023.
- [16] M. B. Uddin and Z. Fatah, “Gudang Jurnal Multidisiplin Ilmu Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas,” *Gudang J. Multidisiplin Ilmu*, vol. 2, no. November, pp. 86–94, 2024.
- [17] I. Ajeng Nur Afifah and H. Nurdianto, “Data Mining Clustering Dalam Pengelompokan Buku Perpustakaan Menggunakan Algoritma K-Means,” *JUPI*, vol. 8, no. 3, pp. 802–814, 2023.