

# DATA MINING CLUSTERING DALAM PENGELOMPOKAN BUKU PERPUSTAKAAN MENGGUNAKAN ALGORITMA K-MEANS

Inas Ajeng Nur Afifah\*<sup>1)</sup>, Heri Nurdiyanto<sup>2)</sup>

1. STMIK Dharmawacana Metro, Indonesia
2. STMIK Dharmawacana Metro, Indonesia

## Article Info

**Kata Kunci:** clustering; data mining; K-Means; perpustakaan

**Keywords:** clustering; datamining; K-Means; libraries

## Article history:

Received 7 April 2023

Revised 21 April 2023

Accepted 5 May 2023

Available online 1 September 2023

## DOI :

<https://doi.org/10.29100/jipi.v8i3.3891>

\* Corresponding author.

Corresponding Author

E-mail address:

[inasajeng1101@gmail.com](mailto:inasajeng1101@gmail.com)

## ABSTRAK

Kebutuhan informasi menjadi suatu elemen penting dalam pengambilan keputusan. Perkembangan teknologi tersebut masyarakat harus dapat menyajikan informasi yang dibutuhkan harus cepat dan tepat. Masalah yang terjadi kebutuhan informasi tadi kadang tidak sesuai dengan yang diharapkan. Perpustakaan sebagai wadah untuk menimba ilmu bagi semua kalangan baik muda atau tua. Perpustakaan sebagai sarana sumber informasi dan ilmu pengetahuan untuk menyimpan bahan pustaka yang dipakai oleh pemakai untuk menggali ilmu sumber informasi. Tujuan utama dari penelitian ini adalah penggunaan metode data mining dalam membantu proses pengelolaan letak buku di perpustakaan. Penambangan data adalah kumpulan metode yang digunakan untuk mendapatkan informasi dari data dengan mempelajari pola dari data tersebut. Selain itu, data mining sering juga disebut Knowledge Discovery in Databases (KDD), yaitu kegiatan yang mencakup pengumpulan penggunaan data historis untuk menemukan keteraturan, pola atau hubungan dalam kumpulan data berukuran besar. K-means adalah suatu teknik pengelompokan data yang mana keberadaan tiap-tiap titik data dalam suatu cluster yang ditentukan oleh derajat keanggotan. Data yang akan digunakan dalam teknik K-Means ini adalah sampel data yang hanya diambil kolom judul buku. Hasil yang diperoleh hitungan manual dan RapidMiner dari peminjaman data buku adalah sama dengan hasil yang telah diproses maka didapatkan jumlah buku yang banyak dipinjamkan terdapat pada cluster0 sebanyak 8 item, jumlah buku yang sedikit dipinjam terdapat pada cluster1 sebanyak 17 item, dan jumlah buku yang cukup banyak dipinjam terdapat pada cluster 2 sebanyak 11 item.”

## ABSTRACT

*Information needs become an important aspect of decision-making. As these technologies develop, the company needs to be able to present the required information quickly and accurately. The earlier problem of information needs is not always what was anticipated. The library is a place to learn for all ages. The library is a medium of information and knowledge sources for storing library documents used by users to explore knowledge of information sources. Information needs are becoming a significant aspect of decision-making. As these technologies evolve, the business must be able to present the necessary information quickly and accurately. The previous problem of information requirements is not always as anticipated. The library offers learning opportunities for all ages. The library is a medium of information and knowledge sources for storing library documents used by users to explore knowledge of information sources. The main focus of this research is the use of data mining methods to assist in managing latex books in the library. Data mining is a collection of methodologies used to obtain information from data by studying models from data. In addition, data mining is often also called Knowledge Discovery in Databases (KDD), which is an activity that includes collecting the use of historical data to find regular patterns or relationships in large data sets. K-means is a data aggregation technique in which the existence of every data point in a cluster is determined by the degree of adhesion. The data to be used in the K-averaging technique is a sample of data that is taken only from the title column of the book. The results obtained by manual and RapidMiner calculations from borrowing book data are*

*the same as the results that have been processed, so it is obtained that the number of books that are mostly loaned is in cluster 0 of 8 items, the number of books that were borrowed a little was in cluster 1 as many as 17 items, and the number of books that were borrowed was quite a lot in cluster 2 as many as 11 items.*

## I. PENDAHULUAN

PADAera sekarang perkembangan teknologi khususnya 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. 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. Perpustakaan sebagai wadah untuk menimba ilmu bagi semua kalangan baik muda atau tua. Perpustakaan sebagai sarana sumber informasi dan ilmu pengetahuan untuk menyimpan bahan pustaka yang dipakai oleh pemakai untuk menggali ilmu sumber informasi[1]. Penelitian ini dilakukan pada salah satu Perpustakaan Daerah Kota Metro. Perpustakaan ini memiliki beragam koleksi buku yang tersedia, seperti buku pendidikan, novel, buku karya umum, filsafat, dan lain sebagainya.

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 petugas perpustakaan melihat kembali catatan transaksi meminjam pada daftar tamu pengunjung[2]. Selain itu petugas perpustakaan tidak dapat mengetahui tingkat peminjaman buku dalam pembuatan laporan bulanan, serta berapa banyak jumlah buku yang dipinjam dengan buku yang tersedia. Untuk itu dibuatlah suatu sistem dengan pengolahan jumlah data yang besar dengan teknik data mining dengan metode k-means.

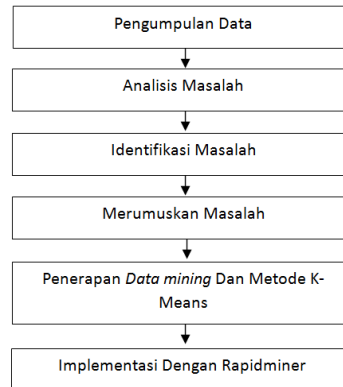
Tujuan utama dari penelitian ini adalah penggunaan metode *data mining* dalam membantu proses pengelolaan letak buku di perpustakaan. Hasil dari penelitian tersebut adalah terbentuk cluster yang di dalamnya berisi buku yang berada dalam kategori dimana terdapat koleksi buku dengan judul yang serupa dan hal tersebut dapat membantu pustakawan dalam mengelola peletakan buku.

Sehingga dalam penelitian ini fokus pada metode data mining dengan kasus pengelompokan data (*clustering*). Adanya data dalam skala besar memungkinkan metode data mining dengan teknik *clustering* yang dapat mengelompokkan data ke dalam beberapa kelompok yang diinginkan. Teknik *clustering* yang digunakan yaitu K-means. Metode K-means ini diterapkan untuk pengelolaan terkait jumlah koleksi buku karena hal tersebut menjadi hal utama dalam peningkatan kualitas pelayanan perpustakaan. Teknik data mining dengan metode K-Means bertujuan agar pihak perpustakaan dapat mengetahui pengelompokan buku antara buku yang dipinjam diperoleh dari transaksi peminjaman buku dari daftar buku pengunjung agar dapat mengetahui buku yang banyak dipinjam sehingga pihak perpustakaan dapat menambah koleksi buku tersebut. Dengan adanya persediaan buku yang lengkap diharapkan menambah minat baca masyarakat di Perpustakaan tersebut.

K-means adalah suatu teknik pengelompokan data yang mana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotan[3]. Data yang akan digunakan dalam teknik K-Means ini ialah sampel data yang hanya diambil kolom judul buku.

## II. METODE PENELITIAN

Pada penelitian ini, dilakukan beberapa tahapan yang saling berkaitan antara satu tahap dengan tahap lainnya. *Flowchart* tahapan penelitian yang dilakukan dapat dilihat di pada *gambar 1 Flowchart Penelitian*



Gambar 1 Flowchart Penelitian

1. Tahap awal dilakukan pengumpulan data yang bertujuan untuk menyediakan bahan agar proses selanjutnya yaitu menganalisa permasalahan yang terjadi pada proses kegiatan pengolahan data buku di perpustakaan[10].
2. Tahap analisa permasalahan dilakukan untuk dapat melihat seberapa banyak permasalahan yang diperoleh berdasarkan tingkat kepentingan dari permasalahan yang akan diselesaikan
3. Tahap identifikasi masalah adalah mengelompokkan permasalahan yang diperoleh yang sesuai dengan tujuan penelitian.
4. Tahap merumuskan masalah adalah tahap pemilihan masalah yang telah teridentifikasi untuk dapat dirumuskan agar dapat menyelesaikan permasalahan yang sesuai dengan tujuan penelitian[11].
5. Proses pengolahan data dilakukan terlebih dahulu melakukan identifikasi sering dihadapi oleh pihak perpustakaan, kemudian mendeskripsikan masalah-masalah tersebut untuk diperoleh solusinya. Tahap selanjutnya dilakukan analisa masalah menggunakan teknik data mining dengan algoritma apriori untuk mendapatkan hasil sebagai tujuan yang akan dicapai[12].
6. Pada tahap ini dilakukan analisis hasil berdasarkan perhitungan algoritma K-Means yang dapat digunakan untuk pengambilan keputusan dalam penentuan pola pengelompokkan buku.
7. Tahap akhir adalah menyimpulkan hasil penelitian apakah telah memberikan solusi dari permasalahan yang telah diselesaikan.

### 2.1 Algoritma K-Means Clustering

K-means termasuk dalam partioning clustering yaitu setiap data harus masuk dalam cluster tertentu dan memungkinkan bagi setiap data yang termasuk dalam cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain. K-means memisahkan data ke k daerah bagian yang terpisah, dimana k adalah bilangan integer positif. Algoritma K-Means sangat terkenal karena kemudahan dan kemampuannya untuk mengklasterkan data besar dan outlier dengan sangat cepat.

Secara sederhana algoritma K-Means adalah sebagai berikut :

1. Tentukan k sebagai jumlah cluster yang ingin dibentuk.
2. Bangkitkan k centroid (titik pusat cluster) awal secara random atau ditentukan secara default.
3. Hitung masing-masing jarak setiap data ke masing-masing centroid.
4. Setiap data memilih centroid terdekat
5. Tentukan posisi centroid baru dengan cara menghitung nilai rata-rata dari data-data yang memilih pada centroid yang sama.

6. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.

K-Means merupakan algoritma *clustering* yang berulang-ulang[4]. Algoritma k-means menetapkan nilai-nilai *cluster* (k) secara acak, dimana nilai tersebut menjadi pusat dari *cluster* atau disebut sebagai *centroid*, mean atau means. Algoritma k-means dalam implementasinya sangat mudah, cepat, mudah beradaptasi sederhana untuk diimplementasikan dan dijalankan, relatif cepat, dan mudah beradaptasi serta mempunyai kemampuan yang besar dalam mengolah data yang cukup besar dan waktu yang efisien. Yang menjadi kelemahan dalam algoritma k-means saat menentukan *cluster* awal, karena bergantung pada inisial data yang diberikan[5].

Pengelompokan data dengan metode K-Means dilakukan dengan algoritma :

1. Tentukan jumlah kelompok
2. Hitung pusat kelompok (*centroid*/rata-rata) dari data yang ada di masing-masing kelompok. Lokasi *centroid* setiap kelompok diambil dari rata-rata (mean) semua nilai data pada setiap fiturnya. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke-i dalam sebuah kelompok dan p menyatakan dimensi data, maka persamaan untuk menghitung *centroid* fitur ke-i digunakan Algoritma 1.

Algoritma 1.....

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j$$

Alokasikan data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*, diantaranya adalah Euclidean. Pengukuran jarak pada ruang jarak (distance space) Euclidean dapat dicari menggunakan Algoritma 2.

Algoritma 2.....

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

## 2.2 Teknik Clustering

*Clustering* disebut sebagai *segmentation*. *Clustering* adalah salah satu sub kategori *data mining* dan merupakan salah satu proses di mana sampel yang sama dibagi menjadi kelompok-kelompok yang disebut *cluster*[6]. Setiap *cluster* termasuk sampel dimana anggota yang mirip satu sama lain dan berbeda dengan sampel yang tersedia dari kelompok lain.

## 2.3 RapidMiner

*RapidMiner* dikenal dengan sebutan YALE (*Yet Another Learning Environment*), dikemukakan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer merupakan Unit Kecerdasan Buatan Universitas Teknik Dortmund tahun 2001. Pada tahun 2006, perkembangannya didorong oleh Rapid-I, sebuah perusahaan yang didirikan oleh Ingo Mierswa dan Ralf Klinkenberg pada tahun yang sama. Pada tahun 2007, nama perangkat lunak itu berubah dari YALE menjadi *RapidMiner*[7]. Pada tahun 2013, perusahaan melakukan *rebranding* dari *Rapid-I* menjadi *RapidMiner*. *RapidMiner* sebuah perangkat lunak ilmu yang dikembangkan oleh perusahaan yang bersama-sama menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam penambangan teks dan analisis prediktif[8]. *RapidMiner* digunakan dalam bidang bisnis, selain itu juga dalam bidang pendidikan, penelitian, pelatihan, *prototype* yang mendukung proses pembelajaran mesin seperti visualisasi, model, persiapan data dan optimasi[9].

### III. HASIL DAN PEMBAHASAN

#### 4.1 Perhitungan Algoritma

Ada sebanyak 200 buku yang akan dianalisis tingkat peminatnya berdasarkan jumlah peminjaman pada masing-masing judul bukunya. Data tersebut akan menjadi sumber data yang selanjutnya akan di terapkan metode clustering K-Means. Perhitungan dalam modul k-means menggunakan data seperti tabel I Data Clustering.

TABLE I  
 DATA CLUSTERING

No	Bulan	Filsafat dan Psikologi	Agama	Ilmu Sosial	Bahasa
1	Januari	135	89	80	95
2	Februari	156	97	86	79
3	Maret	90	108	78	85
4	April	89	89	80	56
5	Mei	80	78	83	76
6	Juni	56	56	79	98
7	Juli	78	73	74	78
8	Agustus	67	72	57	76
9	September	100	89	90	93
10	Oktober	78	86	87	90
11	Nopember	87	78	56	79
12	Desember	76	60	69	87

Algoritma K-Means adalah algoritma yang digunakan oleh peneliti dalam pengelompokkan data. Berikut adalah proses dari algoritma K-Means :

(1) Menentukan Jumlah *Cluster*

Pengelompokan akan dibagi dalam 3 kategori, dimana dalam metode clustering K-Means berarti akan menghasilkan 3 cluster yaitu prioritas tinggi (buku yang banyak dipinjam), dipertimbangkan (buku yang sering dipinjam), dan bukan prioritas (buku sedikit dipinjam) dengan nilai  $k = 3$

(2) Menentukan Titik pusat *Cluster*

TABLE II  
 TITIK PUSAT CLUSTER

c0	Februari	2019	156	97	86	79
c1	April	2020	90	60	59	65
c2	Juni	2021	60	57	78	100

Menentukan titik pusat *cluster* secara acak dari data/objek yang tersedia sebanyak jumlah *clusterk* yang telah ditentukan pada *Table II Titik Pusat Cluster*.

(3) Hitung jarak setiap data yang ada terhadap setiap pusat *cluster*

Rumus yang digunakan untu menghitung jarak setiap data terhadap pusat-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.

Berikut adalah perhitungan jarak setiap data untuk iterasi 1 adalah sebagai berikut :

$$(c0, k1) = \sqrt{(156 - 135)^2 + (97 - 89)^2 + (86 - 80)^2 + (95 - 79)^2}$$

$$\begin{aligned}
 &= \sqrt{441 + 64 + 36 + 256} \\
 &= 28 \\
 (c1, k1) &= \sqrt{(90 - 135)^2 + (60 - 89)^2 + (59 - 80)^2 + (65 - 79)^2} \\
 &= \sqrt{2.025 + 841 + 441 + 196} \\
 &= 55 \\
 (c2, k1) &= \sqrt{(60 - 135)^2 + (57 - 89)^2 + (78 - 80)^2 + (100 - 79)^2} \\
 &= \sqrt{5.625 + 1.024 + 4 + 441} \\
 &= 84 \\
 (c0, k2) &= \sqrt{(156 - 156)^2 + (97 - 97)^2 + (86 - 86)^2 + (95 - 95)^2} \\
 &= 0 \\
 (c1, k2) &= \sqrt{(90 - 156)^2 + (60 - 97)^2 + (59 - 86)^2 + (65 - 95)^2} \\
 &= \sqrt{4.356 + 1.369 + 729 + 900} \\
 &= 86 \\
 (c2, k2) &= \sqrt{(60 - 156)^2 + (57 - 97)^2 + (78 - 86)^2 + (100 - 95)^2} \\
 &= \sqrt{9.216 + 1600 + 64 + 25} \\
 &= 104 \\
 (c0, k3) &= \sqrt{(156 - 90)^2 + (97 - 108)^2 + (86 - 78)^2 + (95 - 85)^2} \\
 &= \sqrt{4.356 + 121 + 64 + 100} \\
 &= 68 \\
 (c1, k3) &= \sqrt{(90 - 90)^2 + (60 - 108)^2 + (59 - 78)^2 + (65 - 85)^2} \\
 &= \sqrt{0 + 2.304 + 361 + 400} \\
 &= 55 \\
 (c2, k3) &= \sqrt{(60 - 90)^2 + (57 - 108)^2 + (78 - 78)^2 + (100 - 79)^2} \\
 &= \sqrt{900 + 2.601 + 0 + 441} \\
 &= 63 \\
 (c0, k4) &= \sqrt{(156 - 89)^2 + (97 - 89)^2 + (86 - 80)^2 + (95 - 50)^2} \\
 &= \sqrt{4.489 + 64 + 36 + 2.025} \\
 &= 81 \\
 (c1, k4) &= \sqrt{(90 - 89)^2 + (60 - 89)^2 + (59 - 80)^2 + (65 - 50)^2} \\
 &= \sqrt{1 + 841 + 441 + 225} \\
 &= 39 \\
 (c2, k4) &= \sqrt{(60 - 89)^2 + (57 - 89)^2 + (78 - 80)^2 + (100 - 50)^2} \\
 &= \sqrt{841 + 1.024 + 4 + 2.500} \\
 &= 66 \\
 (c0, k5) &= \sqrt{(156 - 80)^2 + (97 - 78)^2 + (86 - 83)^2 + (95 - 76)^2} \\
 &= \sqrt{5.776 + 361 + 9 + 361} \\
 &= 81 \\
 (c1, k5) &= \sqrt{(90 - 80)^2 + (60 - 78)^2 + (59 - 83)^2 + (65 - 76)^2} \\
 &= \sqrt{100 + 324 + 576 + 121} \\
 &= 33 \\
 (c2, k5) &= \sqrt{(60 - 80)^2 + (57 - 78)^2 + (78 - 83)^2 + (100 - 76)^2} \\
 &= \sqrt{400 + 441 + 25 + 576} \\
 &= 38
 \end{aligned}$$

- (4) 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*, yaitu seperti pada *Table III Centroid Pertama* :

TABLE III  
 CENTROID PERTAMA

	c0	c1	c2		
k1	28	55	84	1	
k2	0	86	104	1	
k3	68	55	63		1
k4	81	39	66		1
k5	81	33	38		1
k6	52	52	11		1
k7	84	27	33		1
k8	99	28	36		1
k9	57	52	53		1
k10	27	47	37	1	
k11	79	23	47		1
k12	90	77	27		1
k13	19	48	91	1	
k14	50	41	62		1
k15	95	43	33		1
k16	89	0	50		1
k17	72	37	48		1
k18	104	50	0		1
k19	71	45	39		1
k20	64	31	42		1
k21	84	53	39		1
k22	86	36	46		1
k23	68	43	45		1
k24	91	40	54		1
k25	62	73	71		1
k26	69	50	59		1
k27	68	53	48		1
k28	71	34	37		1
k29	84	24	44		1
k30	105	42	51		1
k31	72	31	64		1
k32	110	42	49		1
k33	90	50	35		1
k34	82	52	42		1
k35	70	50	54		1
k36	60	61	63	1	

- (5) Lakukan iterasi dan kemudian tentukan posisi C0 baru dengan cara menghitung rata-rata dari data-data yang berada pada *centroid* yang sama. 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*.

$$\text{Cluster 0} = 1, 2, 10, 13, 36$$

$$\text{Cluster 1} = 3, 4, 5, 7, 8, 9, 11, 14, 16, 17, 20, 22, 23, 24, 26, 28, 29, 30, 31, 32$$

$$\text{Cluster 2} = 6, 12, 15, 18, 19, 21, 25, 27, 33, 34$$

- (6) Hitung pusat *cluster* baru. Untuk *cluster* pertama ada 5 data setelah pengelompokan pertama :

$$C0, 1 = \frac{135 + 156 + 78 + 145 + 97}{5} = 122 ;$$

$$C0, 2 = \frac{89 + 97 + 86 + 88 + 100}{5} = 92 ;$$

$$C0, 3 = \frac{80 + 86 + 87 + 73 + 98}{5} = 85 ;$$

$$C0, 4 = \frac{95 + 79 + 90 + 95 + 87}{5} = 89 ;$$

Untuk *cluster* kedua ada 21 data setelah pengelompokan pertama sehingga :

$$\begin{aligned}
 C1,1 &= \frac{90 + 89 + 80 + 78 + 67 + 100 + 87 + 110 + 90 + 87 + 98 + 78 + 89 + 76 + 90 + 89 + 65 + 99 + 57 + 87 + 89}{21} \\
 &= 85; \\
 C1,2 &= \frac{108 + 89 + 78 + 73 + 72 + 89 + 78 + 87 + 60 + 89 + 72 + 79 + 89 + 80 + 87 + 76 + 78 + 89 + 78 + 98 + 89}{21} \\
 &= 83; \\
 C1,3 &= \frac{78 + 80 + 83 + 74 + 57 + 90 + 56 + 78 + 59 + 78 + 79 + 87 + 80 + 89 + 98 + 80 + 46 + 54 + 77 + 87 + 80}{21} \\
 &= 76; \\
 C1,4 &= \frac{85 + 56 + 76 + 78 + 76 + 93 + 79 + 78 + 65 + 77 + 87 + 64 + 89 + 55 + 89 + 87 + 67 + 67 + 55 + 80 + 56}{21} \\
 &= 74;
 \end{aligned}$$

Untuk *cluster* ketiga, ada 10 data setelah pengelompokan sehingga :

$$\begin{aligned}
 C2,1 &= \frac{56 + 76 + 78 + 60 + 88 + 78 + 98 + 90 + 67 + 78}{10} = 77; \\
 C2,2 &= \frac{56 + 60 + 45 + 57 + 78 + 89 + 112 + 87 + 87 + 86}{10} = 76; \\
 C2,3 &= \frac{79 + 69 + 98 + 78 + 87 + 90 + 100 + 98 + 87 + 97}{10} = 88; \\
 C2,4 &= \frac{98 + 87 + 87 + 100 + 86 + 95 + 95 + 89 + 87 + 67}{10} = 89;
 \end{aligned}$$

Dari perhitungan *cluster* yang telah diperoleh di atas maka akan mendapatkan *centroid* baru untuk iterasi-2 yaitu seperti pada *Table IV Titik Pusat Cluster 2*:

TABLE IV  
 TITIK PUSAT CLUSTER 2

C0	122	92	85	89
C1	85	83	76	74
C2	77	76	88	89

- (7) Ulangi langkah ke-3 hingga *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, seperti data pada *Table V Centroid Kedua*.



TABLE V  
 CENTROID KEDUA

	c0	c1	c2	
k1	15	50	60	1
k2	36	73	82	1
k3	37	28	36	1
k4	47	20	39	1
k5	46	10	23	1
k6	72	46	32	1
k7	50	13	18	1
k8	66	29	35	1
k9	23	29	30	1
k10	44	21	13	1
k11	19	21	15	1
k12	58	29	25	1
k13	27	64	71	1
k14	18	26	38	1
k15	66	46	33	1
k16	57	30	43	1
k17	38	7	19	1
k18	72	45	30	1
k19	37	17	12	1
k20	34	22	23	1
k21	45	27	17	1
k22	52	17	25	1
k23	34	17	15	1
k24	59	19	36	1
k25	62	73	71	1
k26	39	19	29	1
k27	35	28	20	1
k28	14	16	15	1
k29	54	21	35	1
k30	74	37	40	1
k31	45	28	48	1
k32	70	34	41	1
k33	55	61	15	1
k34	53	23	27	1
k35	37	20	26	1
k36	29	33	34	1

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

$$Cluster\ 0 = 1, 2, 9, 11, 13, 14, 28, 36.$$

$$Cluster\ 1 = 3, 4, 5, 7, 8, 16, 17, 20, 22, 24, 26, 29, 30, 31, 32, 33, 34, 35.$$

$$Cluster\ 2 = 6, 10, 12, 15, 18, 19, 21, 23, 25, 27, 33$$

(8) Hitung pusat *cluster* baru. Untuk *cluster* pertama ada 8 data setelah pengelompokan pertama :

$$C0, 1 = \frac{135 + 156 + 100 + 87 + 145 + 110 + 89 + 98}{8} = 115 ;$$

$$C0, 2 = \frac{89 + 97 + 89 + 78 + 88 + 87 + 76 + 100}{8} = 88 ;$$

$$C0, 3 = \frac{80 + 86 + 90 + 56 + 73 + 78 + 80 + 98}{8} = 80 ;$$

$$C0, 4 = \frac{95 + 79 + 93 + 79 + 95 + 78 + 87 + 87}{8} = 87 ;$$

Untuk *cluster* kedua ada 17 data setelah pengelompokan pertama sehingga :

$$C1,1 = \frac{90 + 89 + 80 + 78 + 67 + 90 + 87 + 98 + 87 + 78 + 76 + 90 + 80 + 65 + 99 + 57 + 87 + 78}{17} = 85;$$

$$C1,2 = \frac{108 + 89 + 78 + 73 + 72 + 60 + 89 + 72 + 79 + 80 + 100 + 78 + 78 + 89 + 78 + 86 + 98}{17} = 83;$$

$$C1,3 = \frac{78 + 80 + 83 + 74 + 57 + 59 + 78 + 79 + 87 + 89 + 88 + 56 + 46 + 54 + 77 + 97 + 87}{17} = 76;$$

$$C1,4 = \frac{85 + 56 + 76 + 78 + 65 + 77 + 87 + 64 + 55 + 79 + 76 + 67 + 67 + 55 + 67 + 80}{17} = 74;$$

Untuk *cluster* ketiga, ada 11 data setelah pengelompokan sehingga :

$$C2, 1 = \frac{56 + 78 + 76 + 78 + 60 + 88 + 78 + 89 + 98 + 90 + 67}{11} = 78;$$

$$C2, 2 = \frac{56 + 86 + 60 + 45 + 57 + 78 + 89 + 89 + 112 + 87 + 87}{11} = 77;$$

$$C2,3 = \frac{79 + 87 + 69 + 98 + 78 + 87 + 90 + 80 + 100 + 98 + 87}{11} = 87;$$

$$C2,4 = \frac{98 + 90 + 87 + 87 + 100 + 86 + 95 + 89 + 95 + 87 + 89}{11} = 91;$$

Dari perhitungan pusat *cluster* yang telah diperoleh diatas maka akan mendapatkan *centroid* baru untuk iterasi-3 yaitu seperti *Table VI Titik Pusat Cluster 3*:

TABLE VI  
 TITIK PUSAT CLUSTER 3

C0	115	88	80	87
C1	82	83	75	71
C2	78	77	87	91

Karena proses yang dilakukan pada itersi ke-2 masih belum mendapat hasil yang sama dengan iterasi sebelumnya (iterasi-1), maka perlu melakukan beberapa it iterasi erasi 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 langkah kembali. Berikut adalah *tabel VII Centroid Kedua* hasil perhitungan jarak pengelompokan data dengan iterasi-3 adalah



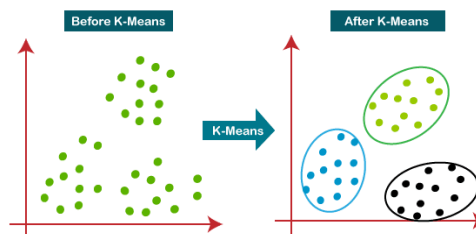
1	1	1	
1	1	1	
1	1	1	
1	1	1	
	1	1	1
	1	1	1
1	1	1	1
1	1	1	1

Perulangan dihentikan setelah hasil dari perhitungan pada iterasi ke-2 dan iterasi ke-3 tidak mengalami perubahan pada titik *cluster*, hasil dari perhitungan adalah sebagai berikut :

- Cluster0* = 8 Peminjaman
- Cluster1* = 17 peminjaman
- Cluster2* = 11 Peminjaman

Dari kesimpulan diatas dapat diketahui peminjaman buku menengah terdapat pada *cluster1*, sedangkan untuk peminjaman buku banyak terdapat pada *cluster0* dan *cluster2* adalah peminjaman buku sedikit.

Gambar 2 merupakan gambar hasil pengelompokan clustering menggunakan *RapidMinerK-Means*



Gambar 2 Rapidminer K-means

#### IV. KESIMPULAN

Hasil yang diperoleh hitungan manual dan *RapidMiner* dari data peminjaman buku adalah sama dengan hasil yang telah diproses maka didapatkan jumlah buku yang banyak dipinjam terdapat pada *cluster0* sebanyak 8 item, jumlah buku yang sedikit dipinjam terdapat pada *cluster1* sebanyak 17 item, dan jumlah buku yang cukup banyak dipinjam terdapat pada *cluster2* sebanyak 11 item. Penerapan data mining dengan metode K-Means clustering dapat diterapkan pada pengelompokan buku sehingga membantu pihak Perpustakaan sehingga dapat mengetahui buku mana yang sering dipinjam.

#### DAFTAR PUSTAKA

- [1] N. N. Hasanah and A. S. Purnomo, "Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik LPP Yogyakarta)," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 4, no. 2, pp. 300–311, 2022.
- [2] E. Bakker, "Implementasi Data Mining Clustering Data Perpustakaan Menggunakan Algoritma K-Means untuk Menentukan Penambahan Koleksi Buku di Perpustakaan UPY," in *Seri Prosiding Seminar Nasional Dinamika Informatika*, 2020, vol. 4, no. 1.
- [3] M. Silalahi, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," *Indonesian Journal of Computer Science*, vol. 10, no. 1, 2021.
- [4] N. I. Royanti and B. Ismanto, "PENGELOMPOKKAN KEAKTIFAN PEMINJAMAN BUKU DI PERPUSTAKAAN STMIK WIDYA PRATAMA DENGAN METODE K-MEANS," *IC-Tech*, vol. 15, no. 1, 2020.
- [5] E. Yuliana, M. Syafii, B. E. Damanik, and S. Suhelmi, "Penerapan K-Means dalam Pengelompokan Buku Perpustakaan Amik dan Stikom Tunas Bangsa Pematang Siantar," *AFoSJ-LAS (All Fields of Science Journal Liaison Academia and Society)*, vol. 1, no. 4, pp. 262–270, 2021.
- [6] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *Jurnal Sistem Komputer Dan Informatika (JSON)*, vol. 1, no. 2, pp. 151–157, 2020.
- [7] D. Sunia, K. Kurniabudi, and P. A. Jusia, "Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means," *Jurnal Ilmiah Mahasiswa Teknik Informatika*, vol. 1, no. 2, pp. 121–134, 2019.
- [8] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *Jurnal Sistem Komputer Dan Informatika (JSON)*, vol. 1, no. 2, pp. 151–157, 2020.
- [9] R. P. Primanda, A. Alwi, and D. Mustikasari, "Data Mining Seleksi Siswa Berprestasi Untuk Menentukan Kelas Unggulan Menggunakan Metode K-Means Clustering (Studi Kasus di MTS Darul Fikri)," *KOMPUTEK*, vol. 5, no. 1, pp. 88–100, 2021.
- [10] N. N. Hasanah and A. S. Purnomo, "Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik LPP Yogyakarta)," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 4, no. 2, pp. 300–311, 2022.

- [11] S. Gunawan, A. Aprilio, and R. Rhandy, "Implementasi k-means, suffix tree dan Dewey Decimal Classification untuk shelving buku perpustakaan," *Jurnal Algoritma, Logika dan Komputasi*, vol. 2, no. 1, 2019.
- [12] D. A. Fakhri and S. Defit, "Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering," *Jurnal Informasi dan Teknologi*, pp. 160–166, 2021.