

PENERAPAN DATA MINING CLUSTERING DALAM MENGELOMPOKAN BUKU DENGAN METODE K-MEANS

Januardi Nasir

Prodi Sistem Informasi

STMIK Indonesia Padang

Email: januardinasir@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*. 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.

Kata kunci: perpustakaan, *data mining*, *clustering*, *RapidMiner*

ABSTRACT

*Library as a means of information sources and knowledge to store library materials used by users to explore knowledge sources of information. The research was conducted at one of the libraries in the city of Batam. This library has a wide collection of books such as general books, scientific works, languages, history, and so on. The problem that often occurs is that the borrowed books are sometimes not available, besides that the library also experiences difficulties because they do not know how many books are borrowed so that the library looks again at the transaction records for borrowing books in the guest book. For this reason, a system with processing large amounts of data was created using the *k-means* data mining technique. From the results obtained, data on book borrowers who have been processed get 9 items of books that are borrowed in cluster 1, the least borrowed books are in cluster 2 as many as 15 items, books that are borrowed quite a lot are in cluster 0 as many as 12 items.*

Keywords: *libraries, data mining, clustering, RapidMiner*

1. PENDAHULUAN

Pada era 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. 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 di lemari arsip yang jumlahnya sangat banyak. 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

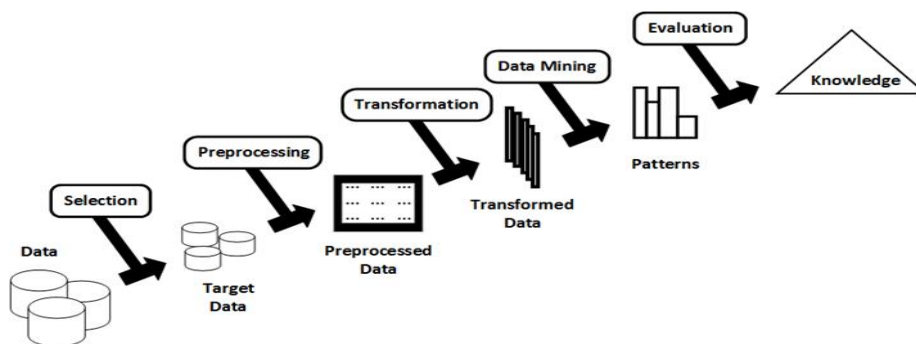
menyimpan bahan pustaka yang dipakai oleh pemakai untuk menggali ilmu sumber informasi. Penelitian dilakukan pada salah satu Perpustakaan yang ada di kota Batam. Perpustakaan ini memiliki luas area 50 m², dirancang agar pengunjung merasa nyaman. Perpustakaan ini memiliki beragam koleksi buku yang tersedia, 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. 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 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*.

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 [1]. 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. Teknik clustering meminimalisir suatu objek data yang mirip dan mengklaster data secara ilmiah [2].

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.

1.1. *Knowledge Discovery in Database (Kdd)*

Knowledge Discovery in Database (KDD) suatu teknik pembentukan pola atau rule dalam informasi. Informasi yang dihasilkan didapatkan dari suatu data yang besar atau dikenal dengan tambang data yang disimpan dalam basis data yang awalnya belum diketahui dan menghasilkan suatu data yang potensial bermanfaat. Iterasi dalam data mining disebut proses KDD. Berikut langkah-langkah dari KDD:



Gambar 1. Proses KDD

Seleksi data yaitu sekumpulan data harus melalui tahap KDD. Data yang dipakai harus melalui proses data mining dengan penyeleksian data dari kumpulan data, kemudian data di simpan dalam database secara terpisah dan disimpan dalam suatu file/berkas. Pembersihan data; tahap ini dilakukan pembersihan data atau membuang data yang tidak konsisten dan memperbaiki data yang salah serta melakukan pemeriksaan data

yang tidak konsisten, data redundansi, duplikasi data, sehingga menghasilkan informasi yang relevan dan berguna. Transformasi data yaitu data diubah menjadi bentuk yang sesuai dengan yang di mining. Pengkodean merupakan proses data mining yang akan dicari dalam database. Data mining merupakan teknik atau metode dalam pembentukan pola sehingga menghasilkan informasi yang dibutuhkan. pola yang ditemukan; dari pola yang dihasilkan maka didapatkanlah suatu informasi yang mudah dipahami oleh pengguna dan menjadi sumber pengetahuan dalam pengambilan keputusan.

1.2. *Data Mining*

Suatu konsep yang dipakai dalam menghasilkan suatu aturan dalam penemuan pengetahuan disebut data mining. *Data mining* atau tambang data merupakan metode, teknik, *artificial intelegent* dan mesin pembelajaran yang diekstraksi sehingga menghasilkan suatu pengetahuan dan informasi yang berguna yang tersimpan dalam suatu database besar. Pada prinsipnya data mining mewarisi banyak aspek dan teknik bidang-bidang ilmu. *Data mining* bukanlah sesuatu hal yang baru, karena *data mining* merupakan akar dari berbagai bidang ilmu tersebut.

1.3. *Teknik Clustering*

Clustering disebut sebagai *segmention*. Metode ini mengidentifikasi kelompok dalam sebuah kasus yang didasarkan pada kelompok atribut yang memiliki kemiripan. Cara kerja *clustering* memisahkan sejumlah kelompok data berdasarkan ciri masing-masing, dimana objeknya dapat berupa orang, peristiwa dan lainnya yang didistribusikan ke dalam kelompok sehingga terdapat beberapa tingkatan yang saling berhubungan antar *cluster*, kuat dan lemahnya antar anggota dari *cluster* yang berbeda terlihat pada anggota *cluster* yang sama [5].

1.4. *Algoritma K-Means*

K-Means merupakan algoritma clustering yang berulang-ulang. 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 lebih efisien. Yang menjadi kelemahan dalam algoritma *k-means* saat menentukan *cluster* awal, karena bergantung pada inisial data yang diberikan [6]

1.5. *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*. 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. *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.

2. METODOLOGI PENELITIAN



Gambar 2. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1. Seleksi Data

Tahapa awal yang dilakukan adalah seleksi data. Data peminjaman buku yang akan diolah tahun 2016-2018, dengan jumlah data sebanyak 36 diambil dari jumlah peminjam tiap bulanya.

3.2. 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.

3.3. Perhitungan Algoritma

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

(1) Menentukan Jumlah *Cluster*

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

(2) Menentukan titik pusat *Cluster*

c0	Ferbruari 2016	14	20	35	15
c1	Mei 2018	35	24	21	9
c2	Maret 2018	19	31	12	10

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

(3) 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* [9].

$$d(xi, ui) = \sqrt{(xi - ui)^2} d(xi, ui) = \sqrt{(xi - ui)^2} \quad (2)$$

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

$$(C0,k1) = \sqrt{(14 - 22)^2 + (20 - 15)^2 + (35 - 30)^2 + (15 - 14)^2}$$

$$\sqrt{(14 - 22)^2 + (20 - 15)^2 + (35 - 30)^2 + (15 - 14)^2} = 11$$

$$(C1,k1) = \sqrt{(35 - 22)^2 + (24 - 15)^2 + (21 - 30)^2 + (9 - 14)^2}$$

$$\sqrt{(35 - 22)^2 + (24 - 15)^2 + (21 - 30)^2 + (9 - 14)^2} = 19$$

$$(C2,k1) = \sqrt{(35 - 22)^2 + (31 - 15)^2 + (12 - 30)^2 + (10 - 14)^2}$$

$$\sqrt{(35 - 22)^2 + (31 - 15)^2 + (12 - 30)^2 + (10 - 14)^2} = 25$$

$$(C0,k2) = \sqrt{(14 - 14)^2 + (20 - 20)^2 + (35 - 35)^2 + (15 - 15)^2}$$

$$\sqrt{(14 - 14)^2 + (20 - 20)^2 + (35 - 35)^2 + (15 - 15)^2} = 0$$

$$(C1,k2) = \sqrt{(35 - 14)^2 + (24 - 20)^2 + (21 - 35)^2 + (9 - 15)^2}$$

$$\sqrt{(35 - 14)^2 + (24 - 20)^2 + (21 - 35)^2 + (9 - 15)^2} = 26$$

$$(C2,k2) = \sqrt{(35 - 14)^2 + (31 - 20)^2 + (12 - 35)^2 + (10 - 15)^2}$$

$$\sqrt{(35 - 14)^2 + (31 - 20)^2 + (12 - 35)^2 + (10 - 15)^2} = 26$$

$$(C0,k3) = \sqrt{(14 - 10)^2 + (20 - 14)^2 + (35 - 15)^2 + (15 - 7)^2}$$

$$\sqrt{(14 - 10)^2 + (20 - 14)^2 + (35 - 15)^2 + (15 - 7)^2} = 23$$

$$(C1,k3) = \sqrt{(35 - 10)^2 + (24 - 14)^2 + (21 - 15)^2 + (9 - 7)^2}$$

$$\sqrt{(35 - 10)^2 + (24 - 14)^2 + (21 - 15)^2 + (9 - 7)^2} = 28$$

$$(C2,k3) = \sqrt{(35 - 10)^2 + (31 - 14)^2 + (12 - 15)^2 + (10 - 7)^2}$$

$$\sqrt{(35 - 10)^2 + (31 - 14)^2 + (12 - 15)^2 + (10 - 7)^2} = 20$$

$$(C0,k4) = \sqrt{(14 - 19)^2 + (20 - 33)^2 + (35 - 37)^2 + (15 - 23)^2}$$

$$\sqrt{(14 - 19)^2 + (20 - 33)^2 + (35 - 37)^2 + (15 - 23)^2} = 16$$

$$(C1,k4) = \sqrt{(35 - 19)^2 + (24 - 33)^2 + (21 - 37)^2 + (9 - 23)^2}$$

$$\sqrt{(35 - 19)^2 + (24 - 33)^2 + (21 - 37)^2 + (9 - 23)^2} = 28$$

$$(C2,k4) = \sqrt{(35 - 19)^2 + (31 - 33)^2 + (12 - 37)^2 + (10 - 23)^2}$$

$$\sqrt{(35 - 19)^2 + (31 - 33)^2 + (12 - 37)^2 + (10 - 23)^2} = 28$$

$$(C0,k5) = \sqrt{(14-8)^2 + (20-30)^2 + (35-31)^2 + (15-12)^2}$$
$$\sqrt{(14-8)^2 + (20-30)^2 + (35-31)^2 + (15-12)^2} = 13$$

$$(C1,k5) = \sqrt{(35-8)^2 + (24-30)^2 + (21-31)^2 + (9-12)^2}$$
$$\sqrt{(35-8)^2 + (24-30)^2 + (21-31)^2 + (9-12)^2} = 30$$

$$(C2,k5) = \sqrt{(35-8)^2 + (31-30)^2 + (12-31)^2 + (10-12)^2}$$
$$\sqrt{(35-8)^2 + (31-30)^2 + (12-31)^2 + (10-12)^2} = 22$$

$$(C0,k6) = \sqrt{(14-10)^2 + (20-23)^2 + (35-29)^2 + (15-17)^2}$$
$$\sqrt{(14-10)^2 + (20-23)^2 + (35-29)^2 + (15-17)^2} = 8$$

$$(C1,k6) = \sqrt{(35-10)^2 + (24-23)^2 + (21-29)^2 + (9-17)^2}$$
$$\sqrt{(35-10)^2 + (24-23)^2 + (21-29)^2 + (9-17)^2} = 27$$

$$(C2,k6) = \sqrt{(35-10)^2 + (31-23)^2 + (12-29)^2 + (10-17)^2}$$
$$\sqrt{(35-10)^2 + (31-23)^2 + (12-29)^2 + (10-17)^2} = 22$$

$$(C0,k7) = \sqrt{(14-12)^2 + (20-23)^2 + (35-18)^2 + (15-12)^2}$$
$$\sqrt{(14-12)^2 + (20-23)^2 + (35-18)^2 + (15-12)^2} = 18$$

$$(C1,k7) = \sqrt{(35-12)^2 + (24-23)^2 + (21-18)^2 + (9-12)^2}$$
$$\sqrt{(35-12)^2 + (24-23)^2 + (21-18)^2 + (9-12)^2} = 23$$

$$(C2,k7) = \sqrt{(35-12)^2 + (31-23)^2 + (12-18)^2 + (10-12)^2}$$
$$\sqrt{(35-12)^2 + (31-23)^2 + (12-18)^2 + (10-12)^2} = 12$$

$$(C0,k8) = \sqrt{(14-9)^2 + (20-21)^2 + (35-17)^2 + (15-10)^2}$$
$$\sqrt{(14-9)^2 + (20-21)^2 + (35-17)^2 + (15-10)^2} = 18$$

$$(C1,k8) = \sqrt{(35-9)^2 + (24-21)^2 + (21-17)^2 + (9-10)^2}$$
$$\sqrt{(35-9)^2 + (24-21)^2 + (21-17)^2 + (9-10)^2} = 26$$

$$(C2,k8) = \sqrt{(35-9)^2 + (31-21)^2 + (12-17)^2 + (10-10)^2}$$
$$\sqrt{(35-9)^2 + (31-21)^2 + (12-17)^2 + (10-10)^2} = 15$$

$$(C0,k9) = \sqrt{(14-15)^2 + (20-27)^2 + (35-21)^2 + (15-14)^2}$$
$$\sqrt{(14-15)^2 + (20-27)^2 + (35-21)^2 + (15-14)^2} = 16$$

$$(C1,k9) = \sqrt{(35-15)^2 + (24-27)^2 + (21-21)^2 + (9-14)^2}$$
$$\sqrt{(35-15)^2 + (24-27)^2 + (21-21)^2 + (9-14)^2} = 21$$

$$(C2,k9) = \sqrt{(35-15)^2 + (31-27)^2 + (12-21)^2 + (10-14)^2}$$
$$\sqrt{(35-15)^2 + (31-27)^2 + (12-21)^2 + (10-14)^2} = 11$$

$$(C0,k10) = \sqrt{(14 - 12)^2 + (20 - 15)^2 + (35 - 23)^2 + (15 - 21)^2}$$

$$\sqrt{(14 - 12)^2 + (20 - 15)^2 + (35 - 23)^2 + (15 - 21)^2} = 14$$

$$(C1,k10) = \sqrt{(35 - 12)^2 + (24 - 15)^2 + (21 - 23)^2 + (9 - 21)^2}$$

$$\sqrt{(35 - 12)^2 + (24 - 15)^2 + (21 - 23)^2 + (9 - 21)^2} = 28$$

$$(C2,k10) = \sqrt{(35 - 12)^2 + (31 - 15)^2 + (12 - 23)^2 + (10 - 21)^2}$$

$$\sqrt{(35 - 12)^2 + (31 - 15)^2 + (12 - 23)^2 + (10 - 21)^2} = 23$$

- (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 sebagai berikut:

Tabel 1. 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	

- (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*.

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 2. Cluster Baru Iterasi ke-1

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

- (6) 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.

Tabel 3. 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
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-2, 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 :

$$\begin{aligned} \text{Cluster0} &= 1,2,4,5,6,10,8,23,28,31,35,36 \\ \text{Cluster 1} &= 13,16,17,18,21,26,29,30,32 \\ \text{Cluster2} &= 3,7,8,9,11,12,14,15,20,22,24,25,27,33,34 \end{aligned}$$

Berikut adalah perhitungan C0 baru sebagai berikut:

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

$$\begin{aligned} C1,1 &= (34+25+26+30+24+35+23+33+21)/9 = 28 \\ C1,2 &= (21+21+23+28+24+16+18+27+12)/9 = 21 \\ C1,3 &= (12+15+27+25+21+19+21+18+14)/9 = 19 \\ C1,4 &= (16+21+12+10+9+24+23+10+20)/9 = 16 \end{aligned}$$

$$\begin{aligned} C2,1 &= (10+12+9+15+19+17+21+13+18+5+14+10+21+19+13)/15 = 14 \\ C2,2 &= (14+23+21+27+14+26+31+24+22+29+30+38+12+31+21)/15 = 24 \\ C2,3 &= (15+18+17+21+12+23+20+21+17+12+23+23+14+12+13)/15 = 18 \\ C2,4 &= (7+12+10+14+10+12+12+13+9+8+12+12+20+10+8)/15 = 11 \end{aligned}$$

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

Tabel 4. 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. Berikut adalah tabel hasil perhitungan jarak pengelompokan data dengan iterasi-3 adalah:

Tabel 5. Hasil Perhitungan Jarak Dan Pengelompokan Data Iterasi Ke-3

	c0	c1	c2		
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
k30	15	11	18		1
k31	9	18	12	1	
k32	21	9	24		1
k33	19	20	13		1
k34	9	15	7		1
k35	10	22	17	1	
k36	9	19	15	1	

Perulangan berhenti pada iterasi ke-3 karena nilai iterasi-2 dan iterasi-3 tetap dan tidak ada perubahan sehingga perulangan dihentikan pada iterasi ke-3, seperti terlihat pada tabel 6.

Tabel 6. Hasil pengelompokan Data Dari Iterasi Ke-1 Sampai Iterasi Ke-3

Iterasi1			Iterasi 2			Iterasi3		
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
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
		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
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		
1		

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

Cluster0 = 12 Peminjaman

Cluster1 = 9 Peminjaman

Cluster2 = 15 Peminjaman

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

4. KESIMPULAN

Hasil yang diperoleh antara hitungan manual dan *RapidMiner* dari data peminjam buku adalah sama dengan hasil yang telah diproses maka didapatkan jumlah buku yang banyak dipinjam terdapat pada *cluster 1* sebanyak 9 item, jumlah buku yang paling sedikit dipinjam terdapat pada *cluster 2* sebanyak 15 item, jumlah buku yang cukup banyak dipinjam terdapat pada *cluster 0* sebanyak 12 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.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada STMIK Indonesia Padang dalam Program Penelitian Peningkatan Kapasitas serta dukungannya yang telah diberikan kepada peneliti.

DAFTAR PUSTAKA

- [1] A. Ikhwan, D. Nofriansyah, and Sriani, "Penerapan Data Mining dengan Algoritma Fp-Growth untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma)," *Saintikom*, vol. 14, no. 3, pp. 211–226, 2015.
- [2] I. Budiman, "Data Clustering Menggunakan Metodologi Crisp-Dm Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma," *Semin. Nasioanal Apl. Teknol. Inf. 2012 (SNATI 2012)*, ISSN 1907-5022, vol. 2012, no. Snati, pp. 15–16, 2012.
- [3] Kusriani and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: CV Andi Offset, 2009.
- [4] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: ANDI, 2014.
- [5] N. Jannah and T. Yulianto, "Mengelompokkan Siswa Berprestasi Akademik Dengan," vol. 2, no. 2, 2016.
- [6] K. Handoko, "Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di Program Studi Tkj Akademi Komunitas Solok Selatan)," *J. Teknol. dan Sist. Inf.*, vol. 02, no. 03, p. 31, 2016.
- [7] Sugiyono, "Metode Penelitian Manajemen," no. ALFABETA, Bandung., 2014.
- [8] K. & E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta, 2009.
- [9] C. Purnamaningsih, R. Saptono, and A. Aziz, "Pemanfaatan Metode K-Means Clustering dalam Penentuan Penjurusan Siswa SMA," *J. Teknol. Inf. ITSmart*, vol. 3, no. 1, p. 27, 2016.