

IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKAN BUKU MENGUNAKAN ALGORITMA *K-MEANS CLUSTERING* (STUDI KASUS : PERPUSTAKAAN POLITEKNIK LPP YOGYAKARTA)

Nisriina Nuur Hasanah¹⁾, Agus Sidiq Purnomo²⁾

¹⁾Teknologi Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta
email: 18111069@student.mercubuana-yogya.ac.id

²⁾Teknologi Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta
email: sidiq@mercubuana-yogya.ac.id

Abstract

LPP Yogyakarta Polytechnic Library is a facility provided as a learning resource. The number of collections available in the library makes it easier for students to access various knowledge. On the other hand, the available books come from donations, so there is still a lack of control over the availability of certain books. For example, books that are given by contributors are related to engineering sciences, while those that are sought after are books on plantation science. Based on these problems, in order to optimize the availability of books, it is necessary to analyze so that the ratio of availability and the number of enthusiasts is appropriate.

One alternative is to implement data mining using the K-Means Clustering method to help library managers analyze by grouping books from the number of borrowed and the number of available copies so that they can be used to support the decision whether to reproduce books with related titles or not.

The data used in this study is the history of borrowing books from the Yogyakarta LPP Polytechnic Library for approximately 2 years. The final result of the research in the form of book data is the most in demand, quite in demand, and the least in demand.

Keywords: Data Mining, K-Means, Clustering, Library

Abstrak

Perpustakaan Politeknik LPP Yogyakarta merupakan fasilitas yang disediakan sebagai sumber belajar. Banyaknya koleksi yang tersedia di Perpustakaan mempermudah mahasiswa mengakses berbagai ilmu pengetahuan. Disisi lain buku – buku yang tersedia berasal dari sumbangan, sehingga untuk kontrol ketersediaan buku tertentu masih kurang. Misalnya buku yang diberikan penyumbang buku terkait ilmu teknik, sedangkan yang banyak dicari buku ilmu perkebunan. Berdasarkan masalah tersebut, agar dapat mengoptimalkan ketersediaan buku, perlu analisis agar rasio ketersediaan dan banyaknya peminat sesuai.

Salah satu alternatif adalah dengan implementasi data mining menggunakan metode K-Means Clustering untuk membantu pengelola Perpustakaan menganalisis dengan mengelompokkan buku dari banyaknya peminjaman dan banyaknya ketersediaan eksemplar buku sehingga bisa digunakan menunjang keputusan apakah perlu memperbanyak buku dengan judul terkait atau tidak.

Data yang digunakan dalam penelitian ini adalah data riwayat peminjaman buku Perpustakaan Politeknik LPP Yogyakarta kurang lebih selama 2 tahun. Hasil akhir dari penelitian berupa data buku paling diminati, cukup diminati, dan sedikit diminati.

Keywords: Data Mining, K-Means, Clustering, Perpustakaan

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



1. PENDAHULUAN

Berkembangnya teknologi dipengaruhi perkembangan ilmu pengetahuan. Dengan perkembangan teknologi ini membuka banyak peluang berinovasi dalam berbagai bidang keilmuan. Salah satunya *data mining*. Sudah banyak penelitian terkait dengan *data mining* dengan berbagai metode sesuai dengan keperluan penelitian. Mulai dari *Clustering*, analisis, hingga untuk keperluan prediksi berdasarkan parameter yang digunakan.

Perpustakaan Politeknik LPP Yogyakarta merupakan salah satu fasilitas yang disediakan sebagai tempat bagi civitas akademika Politeknik LPP Yogyakarta untuk menggali ilmu pengetahuan. Banyak judul buku yang tersedia di Perpustakaan Politeknik LPP Yogyakarta namun dalam pengelolaannya dianggap masih belum optimal. Karena untuk beberapa buku rasio banyaknya ketersediaan buku dengan jumlah peminjam tidak seimbang sehingga dimungkinkan akan terdapat kekosongan apabila jumlah eksemplar tidak diperbanyak. Memang sejauh ini pengelola Perpustakaan Politeknik LPP Yogyakarta belum pernah mengajukan pengadaan atau memperbanyak eksemplar karena ketersediaan tempat yang terbatas. Disisi lain buku – buku baru atau penambahan eksemplar yang ada berasal dari sumbangan sehingga pengelola Perpustakaan tidak bisa mengelola atau mengontrol judul buku apa saja yang seharusnya diperbanyak. Sehingga beberapa judul buku yang seharusnya tidak diperbanyak karena sedikit peminatnya justru terlalu banyak eksemplarnya. Terkait dengan memperbanyak eksemplar, kedepannya pengelola Perpustakaan berencana untuk memperbanyak sendiri sehingga pengelola Perpustakaan harus tahu buku apa saja yang paling diminati,

cukup diminati, dan sedikit diminati agar dapat menentukan buku apa saja yang harus diperbanyak eksemplarnya.

K-Means merupakan salah satu metode *clustering* yang biasanya banyak digunakan pada implementasi *data mining* untuk mengklaster atau mengelompokkan data menjadi beberapa kelompok data. Metode ini mengelompokkan jumlah data yang besar berdasarkan titik pusatnya (*Centroid*). *Clustering* data akan mengelompokkan objek yang paling dekat dimana terdapat kesamaan dengan objek lain, serta data yang akan di-*cluster* diambil secara acak atau random. Data yang dikelompokkan dengan memiliki karakteristik *clustering*. (Merliana, Ernawati, & Santoso)

Dengan implementasi *data mining* dengan metode *K-Means Clustering* diharapkan dapat membantu pengelompokan buku yang paling diminati, cukup diminati, dan sedikit diminati yang ada di Perpustakaan Politeknik LPP Yogyakarta sehingga pengelola perpustakaan dapat dengan mudah melakukan analisis dan mengambil keputusan berdasarkan hasil pengelompokan tersebut.

2. TINJAUAN PUSTAKA

Penelitian mengenai implementasi *data mining* untuk menilai kemampuan nasabah koperasi dalam pemberian kredit yang dilihat dari identifikasi dengan memperhatikan data historis pinjaman. Dalam penelitian ini algoritma yang digunakan adalah *naive bayes*. Sumber data yang digunakan sendiri berasal dari koperasi simpan pinjam sumber rejeki. Dalam penelitian ini terdapat beberapa atribut yang digunakan seperti total telat, nilai jaminan, status kepemilikan, jenis bunga, status perkawinan, status pernikahan, jumlah plafon diusulkan, dan jangka waktu kredit. Algoritma *Naive Bayes* dapat digunakan untuk klasifikasi

kategori kredit nasabah yang hendak mengajukan pinjaman koperasi karena memiliki tingkat kebenaran klasifikasi sebesar 94,77% (Pandie, 2018).

Penelitian mengenai pengelompokan buku di Perpustakaan Yayasan Nurul Islam Indonesia Baru. Dalam penelitian ini terdapat beberapa parameter yang digunakan untuk penelitian, seperti no. rak buku, kategori buku, peminjaman buku, dan stok buku. Metode yang digunakan dalam penelitian ini adalah metode *K-Means Clustering*. Berdasarkan hasil penelitian, dalam merancang *data mining* berbasis desktop yang mengadopsi metode *K-Means Clustering* dapat digunakan dalam penyelesaian masalah Perpustakaan Yayasan Nurul Islam Indonesia Baru (Haryani, Nofriansyah, & Mariami, 2021).

Penelitian mengenai analisis sentimen terhadap acara televisi berdasarkan opini publik. Penelitian diimplementasikan dengan *data mining* menggunakan algoritma *Naïve Bayes*. Beberapa parameter yang digunakan seperti komentar atau tweet pengguna twitter untuk memberikan opini terhadap acara televisi. Dari hasil penelitian yang telah dilakukan terlihat bahwa algoritma *naïve bayes* dapat mengklasifikasikan suatu opini yang berupa tweet ke dalam dua kelas yaitu positif dan negatif dengan akurat. Tingkat keakuratan dari pengklasifikasian tersebut sangat dipengaruhi oleh proses training. Sehingga dapat disimpulkan dari hasil pengklasifikasian yang disajikan dalam bentuk grafik dapat terlihat dengan jelas informasi sentimen publik terhadap suatu acara televisi dan dapat dijadikan sebagai referensi untuk menonton acara televisi (Sentiaji & Bachtiar, 2014).

Penelitian mengenai implementasi klasifikasi kelulusan mahasiswa. Algoritma yang digunakan dalam penelitian ini adalah algoritma dengan

model *k-nearest neighbor*, *decision tree* serta *naive bayes*. Untuk parameter yang digunakan dalam penelitian ini seperti nim, nama, jenis kelamin, usia, IPS1, IPS2, IPS3, IPS4, IPS5, IPS6, IPS7, IPS8, status. Hasil penelitian ini menunjukkan bahwa setelah menggunakan model *K-Nearest Neighbor*, *Decision Tree* serta *Naive Bayes* untuk mengklasifikasi status kelulusan mahasiswa Teknik Informatika Universitas Islam Madura diperoleh hasil bahwa kinerja *Naive Bayes* lebih unggul dari *K-Nearest Neighbor* serta *Decision Tree*. Terbukti bahwa dari 35 data uji yang digunakan *Naive Bayes* memiliki nilai akurasi 89%, presisi 88% sedangkan *K-Nearest Neighbor* memiliki nilai akurasi 77% , presisi 76% dan *Decision Tree* memiliki nilai akurasi 74% dan presisi 84% (Hozairi, Anwari, & Alim, 2021).

Penelitian mengenai implementasi aplikasi *data mining* untuk menentukan *clustering* daerah penyebaran penyakit diare di DKI Jakarta. Algoritma yang digunakan pada penelitian ini adalah algoritma *K-Means*. Di DKI Jakarta sendiri sudah tersebar di 6 Kota yang terdiri dari 44 Kecamatan, dan 267 kelurahan. Untuk mengelompokkan penyebaran diare dibuat beberapa atribut yang digunakan seperti JSP, JSSP, SHARING dan OD untuk menentukan titik penyebaran diare. Pada penelitian ini berhasil menentukan kluster yang terdiri dari C1, C2 dan C3, dimana C1 berjumlah 11 Kelurahan, C2 berjumlah 34 dan C3 berjumlah 43 kelurahan dengan data awal yang digunakan berjumlah 88 data kelurahan. Dimana berdasarkan klustering bahwa C1 dikategorikan sebagai kelurahan yang memiliki titik rawan diare, sedangkan c2 dikategorikan sebagai kelurahan yang daerah kemungkinan rawan penyebaran diare dan sedang c3 dikategorikan sebagai kelurahan yang dikategorikan aman dari penyebaran diare (Santoso & Hadi, 2019).

Penelitian mengenai implementasi *data mining* untuk menentukan produk laris dan tidak laris dari toko Mitra 10 Maguwoharjo. Metode yang digunakan adalah *K-Means Clustering* untuk menentukan kelompok laris maupun tidak dan metode *Weighted Moving Average* untuk memperkirakan stok periode selanjutnya. Pada penelitian ini data yang digunakan adalah data penjualan selama 1 tahun pada tahun 2019 . Dari 29 barang yang tersedia, didapatkan terdapat 25 barang yang masuk dalam kategori laris sedangkan 4 barang lainnya masuk dalam kategori tidak laris yang akan digunakan untuk penentuan penambahan stok periode selanjutnya. Berdasarkan komparasi pengolahan data secara manual dan sistem didapatkan hasil nilai validitas yang tinggi (Samsudin & Purnomo, 2021).

3. METODE PENELITIAN

3.1 Pengumpulan Data

Tahap pertama pada penelitian ini adalah tahap pengumpulan data. Pada tahap ini dilakukan pengumpulan data dimana data yang dikumpulkan adalah data peminjaman buku Perpustakaan Politeknik LPP Yogyakarta selama 26 bulan, terhitung mulai tanggal 1 Januari 2020 sampai dengan 28 Februari 2022.

3.2 Preprocessing Data

Setelah data dikumpulkan dilanjutkan dengan *preprocessing* data. Proses ini merupakan agregasi pada data dengan mengakumulasi atau menjumlahkan peminjaman pada judul buku yang sama. Kemudian dilanjutkan dengan proses reduksi untuk mengambil data yang dibutuhkan agar bisa dengan mudah diolah.

3.3 Data Integration

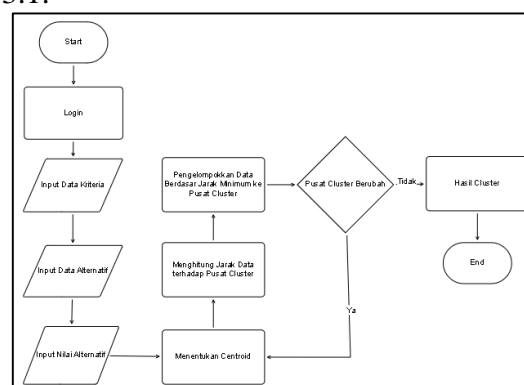
Setelah data direduksi, selanjutnya dilakukan analisis dan penggabungan data apabila data berasal dari beberapa sumber,

misal data dari beberapa excel digabungkan sehingga data terintegrasi.

3.4 Tahap Implementasi

3.4.1 Flowchart Sistem

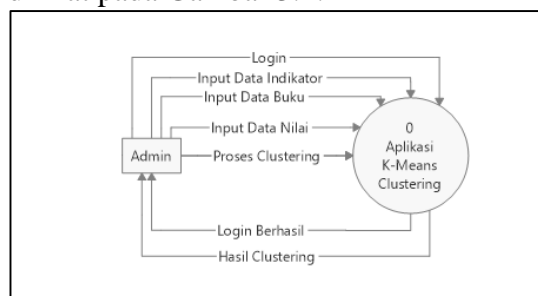
Flowchart (diagram alir) adalah salah satu jenis diagram yang merepresentasikan algoritma atau langkah – langkah instruksi yang berurutan dalam suatu sistem. *Flowchart* untuk penerapan *K-Means Clustering* seperti pada Gambar 3.1.



Gambar 3. 1 Flowchart Sistem

3.4.2 Perancangan Context Diagram

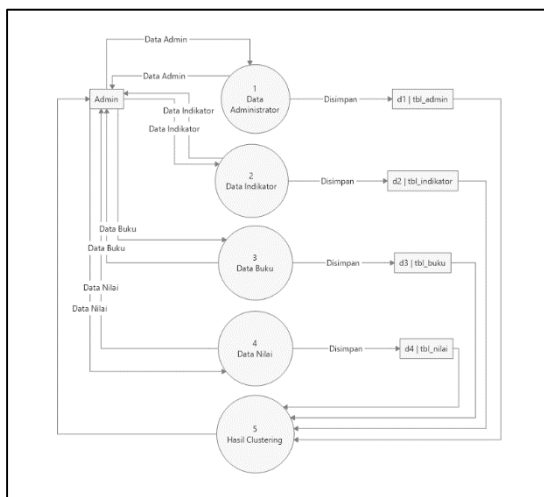
Context Diagram adalah diagram sederhana yang menggambarkan hubungan antara *entity* luar, masukan dan keluaran dari sistem. *Context Diagram* direpresentasikan dengan lingkaran tunggal yang mewakili keseluruhan sistem. Untuk *Context Diagram* dapat dilihat pada Gambar 3.2.



Gambar 3. 2 Context Diagram

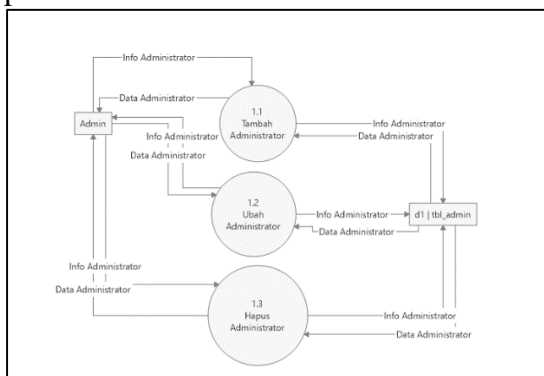
3.4.3 Perancangan Data Flow Diagram (DFD)

Data Flow Diagram (DFD) adalah diagram yang digunakan untuk memodelkan sebuah sistem secara logik. Di dalamnya terdapat gambaran asal data, tujuan data yang keluar dari sistem, dimana data disimpan, proses apa yang menghasilkan data tersebut dan interaksi antara data yang tersimpan serta proses yang dikenakan pada data tersebut. Untuk DFD level 1 dapat dilihat pada Gambar 3.3.



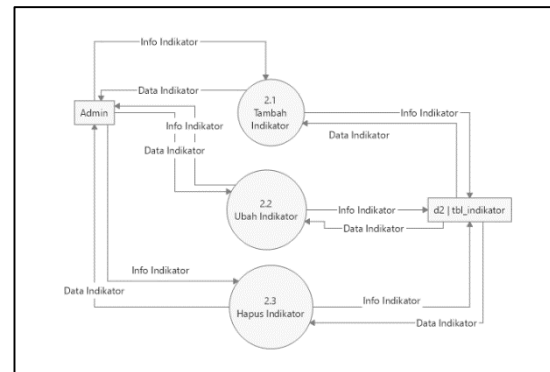
Gambar 3. 3 DFD level 1

Dari DFD level 1 terdapat rincian tiap proses yang digambarkan dengan DFD level 2. Pada DFD level 2 sendiri terdapat 4 proses yaitu proses pada administrator, indikator, buku, dan nilai. Untuk DFD level 2 proses 1 dapat dilihat pada Gambar 3.4 berikut.



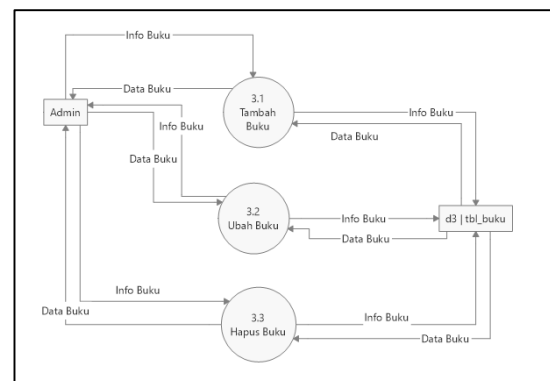
Gambar 3. 4 DFD Level 2 Proses 1

Selanjutnya adalah DFD Level 2 Proses 2 yang dapat dilihat pada Gambar 3.5 berikut ini.



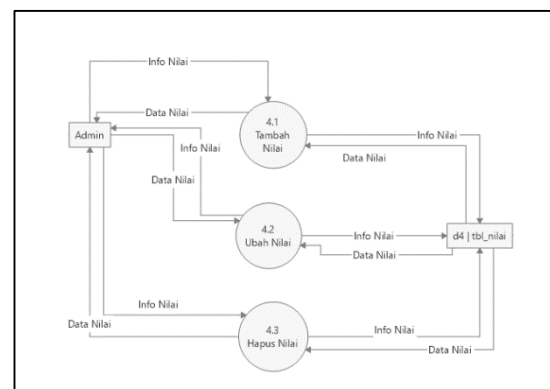
Gambar 3. 5 DFD Level 2 Proses 2

Selanjutnya adalah DFD level 2 Proses 3 yang dapat dilihat pada Gambar 3.6 berikut ini.



Gambar 3. 6 DFD Level 2 Proses 3

DFD yang terakhir adalah DFD level 2 Proses 4 yang dapat dilihat pada Gambar 3.7 berikut ini.



Gambar 3. 7 DFD Level 2 Proses 4

4. HASIL DAN PEMBAHASAN

4.1 Dataset

Dataset adalah himpunan data yang berasal dari informasi atau riwayat data masa lampau dan dikelola menjadi informasi untuk melakukan teknik dari ilmu *data mining*. Tabel dataset didapatkan dari rekap transaksi peminjaman buku di Perpustakaan Politeknik LPP Yogyakarta yang ditambahkan dengan jumlah eksemplar yang tersedia. Tabel dataset dapat dilihat pada Tabel 4.1 sebagai berikut.

Tabel 4. 1 Dataset Peminjaman Buku

No	Judul Buku	Jumlah Transaksi Peminjaman	Jumlah Anggota yang Meminjam	Jumlah Eksemplar
1	Buku Pintar Mandor (BPM) Seni Budidaya Tanaman Tebu	35	23	26
2	Buku Pintar Mandor (BPM) Tanaman Kelapa Sawit	41	20	19
3	Kupas Tuntas AGRIBISNIS KELAPA SAWIT, Mengelola Kebun dan Fabrik Kelapa Sawit Secara Efektif & Efisien	29	14	12
4	12 Budi Daya Tanaman Perkebunan Unggulan	19	12	6
5	Panduan lengkap KAKAO Manajemen Agribisnis dari Hulu hingga Hilir	18	12	8
6	Panduan lengkap Kelapa Sawit, manajemen agribisnis dari hulu hingga hilir	23	12	11
7	Panduan Budidaya KAKAO Rah Sukses Dengan Bertanam Kakao	18	11	7
8	TOP 15 tanaman Perkebunan	17	10	6
9	AKUNTANSI BIAYA EDISI 5	11	9	8
10	Analisis LAPORAN KEUANGAN Edisi Kelima	16	9	6
11	Cekelat, Pembudidayaan, Pengolahan, Pemasaran	10	6	3
12	Dasar-dasar Akuntansi jilid 1 Edisi 7	15	6	6
13	Gulma dan Pengendalannya	11	6	9
14	Wademecum KAKAO	10	6	3
15	AUDITING Petunjuk Praktis Pemeriksaan Akuntan oleh Akuntan Publik Edisi 4 Buku 1	6	5	4
16	Dasar-Dasar Teknologi Gula Rafinasi	6	5	5
17	Masa Depan perkebunan kelapa sawit Indonesia	8	5	2
18	Menghasilkan buah & bibit kakao unggul	10	5	6
19	Pengantar ILMU Penyakit Tumbuhan	5	5	4
20	Pengendalian Hayati, Hama-hama serangga tropis dan gulma seperti Hujan yang jatuh ke bumi	2	2	1
21	Success Pengelolaan Perkebunan Kelapa Sawit Produktivitas Tinggi	7	2	2
22	Teknologi Pengolahan Kelapa Sawit	3	2	2
23	TERMODINAMIKA TEKNIK Jilid 2 EDISI 4	2	2	2
24	TERMODINAMIKA TEKNIK Jilid 2 EDISI 4	2	2	2
25	UNTUNG SELANGHIT DARI AGRIBISNIS KOPI	2	2	2
26	Akuntansi	1	1	2
27	Akuntansi Biaya edisi keempat jilid 1	1	1	1
28	Akuntansi Biaya, perencanaan dan pengendalian	1	1	1
29	Akuntansi Intermedie edisi 9 jilid 1	2	1	1
30	Akuntansi Intermedie edisi ketiga jilid 2	1	1	1

4.2 Iterasi 1

Pada iterasi atau perulangan ke 1, dataset dialokasikan ke dalam cluster secara acak yang kemudian digunakan untuk menghitung *centroid* terdekat dari data yang telah ditentukan dari masing – masing *cluster*. Untuk *centroid* awal ini dipilih angka secara acak yang dapat mewakili tiap *cluster* yang dapat dilihat pada Tabel 4.2.

Tabel 4. 2 Penentuan *Centroid* Awal

<i>Cluster1</i> (paling diminati)	2 9	1 4	1 2
<i>Cluster2</i> (cukup diminati)	1 5	6	6
<i>Cluster3</i> (sedikit diminati)	6	5	4

Perhitungan jarak *centroid* menggunakan Persamaan berikut ini.

$$D_{(ij)} = \sqrt{(X_{1i} - X_{1j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

- a. Perhitungan jarak dari data terhadap pusat *Cluster1*

$$D_{11} =$$

$$\sqrt{(35 - 29)^2 + (23 - 14)^2 + (26 - 12)^2} = 17,69$$

$$D_{12} =$$

$$\sqrt{(41 - 29)^2 + (20 - 14)^2 + (19 - 12)^2} = 15,13$$

$$D_{13} =$$

$$\sqrt{(29 - 29)^2 + (14 - 14)^2 + (12 - 12)^2} = 0,00$$

$$D_{14} =$$

$$\sqrt{(19 - 29)^2 + (12 - 14)^2 + (6 - 12)^2} = 11,83$$

$$D_{15} =$$

$$\sqrt{(18 - 29)^2 + (12 - 14)^2 + (8 - 12)^2} = 11,87$$

dan seterusnya

- b. Perhitungan jarak dari data terhadap pusat *Cluster2*

$$D_{21} =$$

$$\sqrt{(35 - 15)^2 + (23 - 6)^2 + (26 - 6)^2} = 33,00$$

$$D_{22} =$$

$$\sqrt{(41 - 15)^2 + (20 - 6)^2 + (19 - 6)^2} = 32,26$$

$$D_{23} =$$

$$\sqrt{(29 - 15)^2 + (14 - 6)^2 + (12 - 6)^2} = 17,20$$

$$D_{24} =$$

$$\sqrt{(19 - 15)^2 + (12 - 6)^2 + (6 - 6)^2} = 7,21$$

$$D_{25} =$$

$$\sqrt{(18 - 15)^2 + (12 - 6)^2 + (8 - 6)^2} = 7,00$$

dan seterusnya.

c. Perhitungan jarak dari data terhadap pusat Cluster3

$$D_{31} = \sqrt{(35 - 6)^2 + (23 - 5)^2 + (26 - 4)^2} = 40,61$$

$$D_{32} = \sqrt{(41 - 6)^2 + (20 - 5)^2 + (19 - 4)^2} = 40,93$$

$$D_{33} = \sqrt{(29 - 6)^2 + (14 - 5)^2 + (12 - 4)^2} = 25,96$$

$$D_{34} = \sqrt{(19 - 6)^2 + (12 - 5)^2 + (6 - 4)^2} = 14,90$$

$$D_{35} = \sqrt{(18 - 6)^2 + (12 - 5)^2 + (8 - 4)^2} = 14,46$$

dan seterusnya.

Berdasarkan perhitungan di atas didapatkan hasil seperti dapat dilihat pada Tabel 4.3 berikut.

Tabel 4. 3 Perhitungan Jarak Pusat Cluster

No	Judul Buku	Jumlah Transaksi Peminjaman	Jumlah Anggota yang Meminjam	Jumlah Eksemplar	c1	c2	c3	jarak terdekat
1	Buku 1	35	23	26	17,69	33	40,61	17,69
2	Buku 2	41	20	19	15,13	32,26	40,93	15,13
3	Buku 3	29	14	12	0	17,2	25,96	0
4	Buku 4	19	12	6	11,83	7,21	14,9	7,21
5	Buku 5	18	12	8	11,87	7	14,46	7
6	Buku 6	23	12	11	6,4	11,18	19,67	6,4
7	Buku 7	18	11	7	12,45	5,92	13,75	5,92
8	Buku 8	17	10	6	14	4,47	12,25	4,47
9	Buku 9	11	9	8	19,1	5,39	7,55	5,39
10	Buku 10	16	9	6	15,17	3,16	10,95	3,16
11	Buku 11	10	6	3	22,49	5,83	4,24	4,24
12	Buku 12	15	6	6	17,2	0	9,27	0
13	Buku 13	11	6	9	19,92	5	7,14	5
14	Buku 14	10	6	3	22,49	5,83	4,24	4,24
15	Buku 15	6	5	4	25,96	9,27	0	0
16	Buku 16	6	5	5	25,67	9,11	1	1
17	Buku 17	8	5	2	24,94	8,12	2,83	2,83
18	Buku 18	10	5	6	21,86	5,1	4,47	4,47
19	Buku 19	5	5	4	26,85	10,25	1	1
20	Buku 20	5	5	4	26,85	10,25	1	1
21	Buku 21	2	2	1	31,53	14,49	5,83	5,83
22	Buku 22	7	2	2	26,98	9,8	3,74	3,74
23	Buku 23	3	2	2	30,33	13,27	4,69	4,69
24	Buku 24	2	2	2	31,19	14,18	5,39	5,39
25	Buku 25	2	2	2	31,19	14,18	5,39	5,39
26	Buku 26	1	1	2	32,45	15,39	6,71	6,71
27	Buku 27	1	1	1	32,77	15,68	7,07	7,07
28	Buku 28	1	1	1	32,77	15,68	7,07	7,07
29	Buku 29	2	1	1	31,92	14,8	6,4	6,4
30	Buku 30	1	1	1	32,77	15,68	7,07	7,07

Setelah dihitung dengan centroid dan telah ditemukan jarak terdekatnya, selanjutnya menentukan pengelompokan data ke-1. Untuk pengelompokan datanya dapat dilihat pada Tabel 4.4 berikut.

Tabel 4. 4 Pengelompokan Data ke-1

No	Judul Buku	c1	c2	c3
1	Buku 1	1		
2	Buku 2	1		
3	Buku 3	1		
4	Buku 4		1	
5	Buku 5		1	
6	Buku 6	1		
7	Buku 7		1	
8	Buku 8		1	
9	Buku 9		1	
10	Buku 10		1	
11	Buku 11			1
12	Buku 12		1	
13	Buku 13		1	
14	Buku 14			1
15	Buku 15			1
16	Buku 16			1
17	Buku 17			1
18	Buku 18			1
19	Buku 19			1
20	Buku 20			1
21	Buku 21			1
22	Buku 22			1
23	Buku 23			1
24	Buku 24			1
25	Buku 25			1
26	Buku 26			1
27	Buku 27			1
28	Buku 28			1
29	Buku 29			1
30	Buku 30			1

4.3 Iterasi 2

Pada iterasi atau perulangan ke 2, dilakukan proses penghitungan kembali karena masih ada data yang berpindah cluster. Sebelum proses penghitungan perlu menentukan centroid kembali dengan mengambil rata – rata dari setiap data yang ada pada cluster. Untuk menentukan centroid baru penghitungan menggunakan Persamaan berikut.

$$T_{(i)} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

a. Perhitungan *centroid* untuk *cluster1*

$$T_{11} = \frac{35+41+29+23}{4} = 32$$

$$T_{12} = \frac{23+20+14+12}{4} = 17,25$$

$$T_{13} = \frac{26+19+12+11}{4} = 17$$

b. Perhitungan *centroid* untuk *cluster2*

$$T_{21} = \frac{19+18+18+17+11+16+15+11}{8} = 15,63$$

$$T_{22} = \frac{12+12+11+10+9+9+6+6}{8} = 9,38$$

$$T_{23} = \frac{6+8+7+6+8+6+6+9}{8} = 7$$

c. Perhitungan *centroid* untuk *cluster3*

$$T_{31} = \frac{10+10+6+6+8+10+5+5+2+7+3+2+2+1+1+1+2+1}{18} = 4,56$$

$$T_{32} = \frac{6+6+5+5+5+5+5+5+2+2+2+2+2+1+1+1+1+1}{18} = 3,17$$

$$T_{33} = \frac{3+3+4+5+2+6+4+4+1+2+2+2+2+2+1+1+1+1}{18} = 2,56$$

Dari hasil pernghitungan di atas didapat *centroid* baru yang bisa dilihat pada Tabel 4.5 berikut.

Tabel 4. 5 *Centroid* Baru

<i>Cluster1</i> (paling diminati)	32	17,25	17
<i>Cluster2</i> (cukup diminati)	15,63	9,38	7
<i>Cluster3</i> (sedikit diminati)	4,56	3,17	2,56

Setelah didapatkan *centroid* baru, selanjutnya menghitung jarak data ke *centroid* baru menggunakan Persamaan seperti sebelumnya. Untuk hasil perhitungannya dapat dilihat pada Tabel 4.6.

Tabel 4. 6 Perhitungan Jarak Pusat *Cluster*

No	Judul Buku	Jumlah Transaksi Peminjaman	Jumlah Anggota yang Meminjam	Jumlah Eksemplar	c1	c2	c3	jarak terdekat
1	Buku 1	35	23	26	11,09	30,36	43,24	11,09
2	Buku 2	41	20	19	9,62	30,01	43,38	9,62
3	Buku 3	29	14	12	6,68	15,01	28,36	6,68
4	Buku 4	19	12	6	17,82	4,39	17,28	4,39
5	Buku 5	18	12	8	17,45	3,68	16,98	3,68
6	Buku 6	23	12	11	12,02	8,79	22,13	8,79
7	Buku 7	18	11	7	18,3	2,88	16,18	2,88
8	Buku 8	17	10	6	19,96	1,81	14,61	1,81
9	Buku 9	11	9	8	24,29	4,75	10,26	4,75
10	Buku 10	16	9	6	21,1	1,13	13,3	1,13
11	Buku 11	10	6	3	28,4	7,68	6,15	6,15
12	Buku 12	15	6	6	23,16	3,38	11,36	3,38
13	Buku 13	11	6	9	25,13	6,06	9,54	6,06
14	Buku 14	10	6	3	28,4	7,68	6,15	6,15
15	Buku 15	6	5	4	31,54	10,99	2,74	2,74
16	Buku 16	6	5	5	31,15	10,76	3,38	3,38
17	Buku 17	8	5	2	30,84	10,11	3,94	3,94
18	Buku 18	10	5	6	27,48	7,2	6,7	6,7
19	Buku 19	5	5	4	32,37	11,88	2,38	2,38
20	Buku 20	5	5	4	32,37	11,88	2,38	2,38
21	Buku 21	2	2	1	37,26	16,61	3,21	3,21
22	Buku 22	7	2	2	32,9	12,4	2,76	2,76
23	Buku 23	3	2	2	36,04	15,45	2,02	2,02
24	Buku 24	2	2	2	36,85	16,28	2,86	2,86
25	Buku 25	2	2	2	36,85	16,28	2,86	2,86
26	Buku 26	1	1	2	38,08	17,38	4,2	4,2
27	Buku 27	1	1	1	38,48	17,89	4,44	4,44
28	Buku 28	1	1	1	38,48	17,89	4,44	4,44
29	Buku 29	2	1	1	37,68	17,08	3,69	3,69
30	Buku 30	1	1	1	38,48	17,89	4,44	4,44

Setelah dihitung dengan *centroid* baru dan telah ditemukan jarak terdekatnya, selanjutnya menentukan pengelompokan data ke-2 yang dapat dilihat pada Tabel 4.7 berikut.

Tabel 4. 7 Pengelompokan Data ke-2

No	Judul Buku	c1	c2	c3
1	Buku 1	1		
2	Buku 2	1		
3	Buku 3	1		
4	Buku 4		1	
5	Buku 5		1	
6	Buku 6		1	
7	Buku 7		1	
8	Buku 8		1	
9	Buku 9		1	
10	Buku 10		1	
11	Buku 11			1
12	Buku 12		1	
13	Buku 13		1	
14	Buku 14			1
15	Buku 15			1
16	Buku 16			1
17	Buku 17			1
18	Buku 18			1
19	Buku 19			1
20	Buku 20			1
21	Buku 21			1
22	Buku 22			1
23	Buku 23			1
24	Buku 24			1
25	Buku 25			1
26	Buku 26			1
27	Buku 27			1
28	Buku 28			1
29	Buku 29			1
30	Buku 30			1

4.4 Iterasi 3

Pada iterasi atau perulangan ke 3, dilakukan proses penghitungan kembali karena masih ada data yang berpindah *cluster*. Sebelum proses penghitungan perlu menentukan *centroid* baru kembali dengan mengambil rata – rata dari setiap data yang ada pada *cluster* menggunakan Persamaan seperti sebelumnya. Berikut hasil pernghitungan *centroid* baru dapat dilihat pada Tabel 4.8.

Tabel 4. 8 *Centroid* Baru

<i>Cluster1</i> (paling diminati)	35	19	19
<i>Cluster2</i> (cukup diminati)	16,4 4	9,6 7	7,4 4
<i>Cluster3</i> (sedikit diminati)	4,56	3,1 7	2,5 6

Setelah didapatkan *centroid* baru, selanjutnya menghitung jarak data ke *centroid* baru menggunakan Persamaan seperti sebelumnya. Untuk hasil perhitungannya dapat dilihat pada Tabel 4.9.

Tabel 4. 9 Perhitungan Jarak Pusat *Cluster*

No	Judul Buku	Jumlah Transaksi Peminjaman	Jumlah Anggota yang Meminjam	Jumlah Eksemplar	c1	c2	c3	jarak terdekat
1	Buku 1	35	23	26	8,06	29,43	43,24	8,06
2	Buku 2	41	20	19	6,08	29,04	43,38	6,08
3	Buku 3	29	14	12	10,49	14,04	28,36	10,49
4	Buku 4	19	12	6	21,77	3,75	17,28	3,75
5	Buku 5	18	12	8	21,42	2,86	16,98	2,86
6	Buku 6	23	12	11	16,03	7,81	22,13	7,81
7	Buku 7	18	11	7	22,29	2,1	16,18	2,1
8	Buku 8	17	10	6	23,96	1,58	14,61	1,58
9	Buku 9	11	9	8	28,23	5,51	10,26	5,51
10	Buku 10	16	9	6	25,1	1,65	13,3	1,65
11	Buku 11	10	6	3	32,4	8,64	6,15	6,15
12	Buku 12	15	6	6	27,17	4,2	11,36	4,2
13	Buku 13	11	6	9	29,07	6,75	9,54	6,75
14	Buku 14	10	6	3	32,4	8,64	6,15	6,15
15	Buku 15	6	5	4	35,52	11,95	2,74	2,74
16	Buku 16	6	5	5	35,11	11,7	3,38	3,38
17	Buku 17	8	5	2	34,84	11,08	3,94	3,94
18	Buku 18	10	5	6	31,46	8,09	6,7	6,7
19	Buku 19	5	5	4	36,35	12,83	2,38	2,38
20	Buku 20	5	5	4	36,35	12,83	2,38	2,38
21	Buku 21	2	2	1	41,26	17,38	3,21	3,21
22	Buku 22	7	2	2	36,91	13,33	2,76	2,76
23	Buku 23	3	2	2	40,02	16,41	2,02	2,02
24	Buku 24	2	2	2	40,83	17,24	2,86	2,86
25	Buku 25	2	2	2	40,83	17,24	2,86	2,86
26	Buku 26	1	1	2	42,06	18,53	4,2	4,2
27	Buku 27	1	1	1	42,47	18,85	4,44	4,44
28	Buku 28	1	1	1	42,47	18,85	4,44	4,44
29	Buku 29	2	1	1	41,68	18,04	3,69	3,69
30	Buku 30	1	1	1	42,47	18,85	4,44	4,44

Setelah dihitung dengan *centroid* baru dan telah ditemukan jarak terdekatnya, selanjutnya menentukan pengelompokan data ke-3. Untuk pengelompokan datanya dapat dilihat pada Tabel 4.10 berikut.

Tabel 4. 10 Pengelompokan Data ke-3

No	Judul Buku	c1	c2	c3
1	Buku 1	1		
2	Buku 2	1		
3	Buku 3	1		
4	Buku 4		1	
5	Buku 5		1	
6	Buku 6		1	
7	Buku 7		1	
8	Buku 8		1	
9	Buku 9		1	
10	Buku 10		1	
11	Buku 11			1
12	Buku 12		1	
13	Buku 13		1	
14	Buku 14			1
15	Buku 15			1
16	Buku 16			1
17	Buku 17			1
18	Buku 18			1
19	Buku 19			1
20	Buku 20			1
21	Buku 21			1
22	Buku 22			1
23	Buku 23			1
24	Buku 24			1
25	Buku 25			1
26	Buku 26			1
27	Buku 27			1
28	Buku 28			1
29	Buku 29			1
30	Buku 30			1

4.5 Hasil

Setelah melalui beberapa proses iterasi dan tidak ada perubahan pada *cluster* maka didapatkan hasil akhir seperti *cluster* 1 merupakan paling diminati, *cluster* 2 cukup diminati, dan *cluster* 3 sedikit diminati, berikut hasil akhir *clustering* yang dapat dilihat pada Tabel 4.11.

Tabel 4. 11 Hasil Akhir *Clustering*

No	Judul Buku	Cluster
1	Buku Pintar Mandor (BPM) Seri Budidaya Tanaman Tebu	Paling diminati
2	Buku Pintar Mandor (BPM) Tanaman Kelapa Sawit	Paling diminati
3	Kupas Tuntas AGRIBISNIS KELAPA SAWIT, Mengelola Kebun dan Pabrik Kelapa Sawit Secara Efektif & Efisien	Paling diminati
4	12 Budi Daya Tanaman Perkebunan Unggulan	Cukup Diminati
5	Panduan lengkap KAKAO Manajemen Agribisnis dari Hulu hingga Hilir	Cukup Diminati
6	Panduan lengkap Kelapa Sawit, manajemen agribisnis dari hulu hingga hilir	Cukup Diminati
7	Panduan Budidaya KAKAO Raih Sukses Dengan Bertanam Kakao	Cukup Diminati
8	TOP 15 tanaman Perkebunan	Cukup Diminati
9	AKUNTANSI BIAYA EDISI 5	Cukup Diminati
10	Analisis LAPORAN KEUANGAN Edisi Kelima	Cukup Diminati
11	Cokelat, Pembudidayaan, Pengolahan, Pemasaran	Sedikit Diminati
12	Dasar-dasar Akuntansi jilid 1 Edisi 7	Cukup Diminati
13	Gulma dan Pengelolaannya	Cukup Diminati
14	Vademecum KAKAO	Sedikit Diminati
15	AUDITING Petunjuk Praktis Pemeriksaan Akuntan oleh Akuntan Publik Edisi 4 Buku 1	Sedikit Diminati
16	Dasar-Dasar Teknologi Gula Rafinasi	Sedikit Diminati
17	Masa Depan perkebunan kelapa sawit Indonesia	Sedikit Diminati
18	Menghasilkan benih & bibit kakao unggul	Sedikit Diminati
19	Pengantar ILMU Penyakit Tumbuhan	Sedikit Diminati
20	Pengendalian Hayati, Hama-hama serangga tropis dan gulma	Sedikit Diminati
21	Seperti Hujan yang jatuh ke bumi	Sedikit Diminati

N o	Judul Buku	Cluster
22	Sukses Pengelolaan Perkebunan Kelapa Sawit Produktivitas Tinggi	Sedikit Diminati
23	Teknologi Pengolahan Kelapa Sawit	Sedikit Diminati
24	TERMODINAMIKA TEKNIK Jilid 2 EDISI 4	Sedikit Diminati
25	UNTUNG SELANGIT DARI AGRIBISNIS KOPI	Sedikit Diminati
26	Akuntansi	Sedikit Diminati
27	Akuntansi Biaya edisi keempat jilid 1	Sedikit Diminati
28	Akuntansi Biaya, perencanaan dan pengendalian	Sedikit Diminati
29	Akuntansi Intermediate edisi 9 jilid 1	Sedikit Diminati
30	Akuntansi Intermediate edisi ketiga jilid 2	Sedikit Diminati

SIMPULAN

Berdasarkan dari penelitian yang telah dilakukan dapat ditarik kesimpulan seperti:

1. Implementasi *K-Means Clustering* berdasarkan data uji yang telah ditentukan dapat mengelompokkan data judul buku menjadi kelompok paling diminati, cukup diminati, dan sedikit diminati. Dari 30 judul buku yang diuji, 3 diantaranya merupakan judul buku paling diminati seperti Buku Pintar Mandor (BPM) Seri Budidaya Tanaman Tebu, Buku Pintar Mandor (BPM) Tanaman Kelapa Sawit, Kupas Tuntas AGRIBISNIS KELAPA SAWIT, Mengelola Kebun dan Pabrik Kelapa Sawit Secara Efektif & Efisien.
2. Hasil akhir *clustering* diharapkan dapat digunakan sebagai bahan pertimbangan atau penunjang keputusan untuk memperbanyak judul buku yang paling diminati sehingga rasio antara ketersediaan dan banyaknya peminat seimbang.

DAFTAR PUSTAKA

- Haryani, Nofriansyah, D., & Mariami, I. (2021). Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering. *Jurnal Cyber Tech Vol. 1*, 1-12.
- Hozairi, Anwari, & Alim, S. (2021). IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES. *Jurnal Ilmiah NERO Vol. 6 No. 2*, 133-144.
- Merliana, N. P., Ernawati, & Santoso, A. J. (n.d.). ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-MEANS CLUSTERING. *PROSIDING SEMINAR NASIONALMULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI_U)*.
- Pandie, E. S. (2018). Implementasi Algoritma Data Mining Naive Bayes Pada Koperasi. *J-ICON, Vol. 6 No. 1*, 15-20.
- Samsudin, A. R., & Purnomo, A. S. (2021). Implementasi Data Mining Untuk Menentukan Produk Laris Dengan K-Means Clustering dan Weighted Moving Average. *SMAI*, 29-36.
- Santoso, T. B., & Hadi, A. S. (2019). Aplikasi Data Mining untuk Clustering Daerah Penyebaran Penyakit Diare di DKI Jakarta Menggunakan Algoritma K-

MEANS. *JURNAL ILMIAH
FIFO*, 131-143.

Sentiaji, A. R., & Bachtiar, A. M. (2014).
ANALISIS SENTIMEN
TERHADAP ACARA TELEVISI

BERDASARKAN OPINI
PUBLIK. *Jurnal Ilmiah Komputer
dan Informatika (KOMPUTA)*, 1-
6.