KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 3, No 6, Juni 2023, Hal 925-933
DOI 10.30865/klik.v3i6.826
https://djournals.com/klik

Penerapan Data Mining Dalam Pengelompokkan Buku Yang Dipinjam Menggunakan Algoritma K-Means

Herlina Latipa Sari*, Ila Yati Beti

Fakultas Ilmu Komputer, Informatika, Universitas Dehasen Bengkulu, Bengkulu, Indonesia
Email: 1,*herlinalatipasari@unived.ac.id , 2ilayb@unived.ac.id
Email Penulis Korespondensi: herlinalatipasari@unived.ac.id

Abstrak−Tujuan dari penelitian ini adalah menerapkan data mining dalam mengelompokkan buku yang di pinjam berdasarkan dengan jumlah ketersediaan buku, sehingga dengan mengimplementasikan pengelompokkan data dengan algoritma k-means ini akan menentukan buku yang paling banyak diminati yang dilihat dari jumlah ketersediaan buku, buku yang dibaca, dan buku yang dipinjam di Perpustakaan Universitas Dehasen Bengkulu. Data mining adalah suatu proses atau kegiatan yang dilakukan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut sehingga menjadi sebuah informasi yang dapat digunakan untuk sesuatu yang bermanfaat. Dari hasil penerapan data mining untuk pengelompokkan judul buku yang dipinjam di perpustakaan Universitas Dehasen Bengkulu menggunakan algoritma k-means yang telah diterapkan judul buku telah terkelompok menjadi 2 cluster sesuai kriteria yang ditentukan yaitu jumlah buku yang tersedia, jumlah buku yang di baca dan jumlah buku yang dipinjam dengan hasil perhitungan yang ditunjukkan dari cluster I (tinggi) dari pengelompokkan data k-means terdapat 10 jumlah kode dan judul buku dengan jumlah buku yang tersedia sebanyak 295 dengan rata-rata jumlah buku yang dibaca adalah 9 buku dan rata-rata jumlah buku yang dipinjam sebanyak 11 buku dan pada cluster II (rendah) dari pengelompokkan data k-means terdapat 10 jumlah kode dan judul buku dengan jumlah buku yang tersedia sebanyak 115 dengan rata-rata jumlah buku yang dibaca adalah 5 buku dan rata-rata jumlah buku yang dipinjam sebanyak 4 buku.

Kata Kunci: Penerapan; Data Mining; Pengelompokkan; Algoritma; K-Means.

Abstract−The purpose of this study is to apply data mining in grouping books that are borrowed based on the number of available books, so that by implementing data grouping with the k-means algorithm it will determine the books that are most in demand as seen from the number of available books, books read, and books borrowed from the Dehasen Bengkulu University Library. Data mining is a process or activity carried out to collect large data and then extract the data so that it becomes information that can be used for something useful. From the results of applying data mining to grouping book titles borrowed at the Dehasen Bengkulu University library using the k-means algorithm which has been applied the book titles have been grouped into 2 clusters according to the specified criteria, namely the number of books available, the number of books read and the number of books read. borrowed with the calculation results shown from cluster I (high) from the k-means data grouping there are 10 the number of codes and book titles with the number of books available being 295 with the average number of books read being 9 books and the average number of books read 11 books were borrowed and in cluster II (low) of the k-means data grouping there were 10 numbers of codes and book titles with 115 available books with an average number of books read 5 books and an average number of books borrowed as many as 4 books.

Keywords: Data Mining; Grouping; Algorithm; K-Means

1. PENDAHULUAN

Perpustakaan Universitas Dehasen (UNIVED) Bengkulu merupakan unsur penunjang perguruan tinggi dalam kegiatan pendidikan, penelitian dan pengabdian kepada masyarakat. Dalam rangka menunjang kegiatan tri darma tersebut, perpustakaan Universitas Dehasen Bengkulu menyediakan berbagai bahan pustaka yang sangat berguna bagi pelaksanaan dan peningkatan proses belajar mengajar, karena perpustakaan merupakan pusat rujukan yang melayani informasi dan pembelajaran terhadap civitas akademika diantaranya adalah dosen dan mahasiswa Universitas Dehasen Bengkulu.

Pada penelitian ini akan dilakukan kajian data mining dalam menerapkan algoritma K-Means pada pengelompokkan judul buku yang di pinjam di perpustakaan Universitas Dehasen Bengkulu berdasarkan dengan jumlah ketersediaan buku, sehingga dengan mengimplementasikan Pengelompokkan Data dengan Algoritma K-Means ini akan menentukan buku yang paling banyak diminati yang dilihat dari jumlah ketersediaan buku, buku yang dibaca, dan buku yang dipinjam. Dan juga dapat memberikan informasi berupa kategori buku yang paling banyak diminati oleh pengunjung yaitu seluruh civitas akademika Universitas Dehasen Bengkulu atau anggota perpustakaan Universitas Dehasen Bengkulu.

Dalam proses peminjaman buku di perpustakaan UNIVED terkadang terjadi buku yang akan dipinjam tidak tersedia dikarenakan sudah dipinjam. Untuk mengatasi hal tersebut terdapat teknik yang dapat digunakan salah satunya adalah data mining. Data mining merupakan tools memperkirakan perilaku dan tren masa depan, memungkinkan untuk membuat keputusan yang proaktif dan berdasarkan pengetahuan [1]. Definisi lain mengatakan bahwa data mining adalah sebuah kegiatan yang meliputi pengumpulan data, pemakaian data historis, pola atau hubungan dalam data yang berukuran besar [2], dari kedua definisi diatas dapat disimpulkan bahwa data mining adalah suatu proses atau kegiatan yang dilakukan untuk mengumpulkan data yang berukuran besar lalu kemudian mengekstraksi data tersebut sehingga menjadi sebuah informasi yang dapat digunakan untuk sesuatu yang bermanfaat [3].

Clustering adalah proses pengelompokkan data ke dalam kluster yang sesuai terhadap objek-objek yang berbeda kluster, pada metode clustering terdapat beberapa algoritma salah satunya adalah algoritma k-means [4]. Clustering adalah proses pengelompokkan data kedalam cluster berdasarkan parameter tertentu sehingga objek-objek dalam sebuah cluster

memiliki tingkat kemiripan yang tinggi satu sama lain dan sangat tidak mirip dengan objek yang lain pada cluster yang berbeda [5].

K-means adalah salah satu metode clustering non-hirarki yang mempartisi data kedalam bentuk satu atau lebih kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok lain[6]. Pada algoritma ini akan ditentukan terlebih dahulu jumlah cluster untuk kategori data dan pemulihan *centroid point* tidak harus ditentukan dan dibatasi, hal ini berarti bahwa boleh memilih sembarang objek untuk penentuan titik tengah cluster [7]. Penentuan centroid dilakukan dengan cara mengambil data pertama sebagai centroid pertama, data kedua sebagai centroid kedua, dan seterusnya hingga jumlah centroid yang diperlukan. Langkah berikutnya adalah dengan menghitung jarak dari titik yang akan di cluster ke setiap centroid yang ada dan dikelompokkan sesuai dengan jarak terdekat kepada centroid-nya. Bila semua titik sudah masuk kedalam pengelompokkan maka Langkah pertama selesai. Kemudian Langkah berikutnya, kita perlu menghitung kembali k-centroid baru sebagai barycenters dari kelompok yang dihasilkan. Setelah memiliki k-centroid yang baru, pengelompokkan di uji kembali terhadap k-centroid [8]. Luaran dari penerapan data mining menggunakan algoritma k-means dapat digunakan untuk membantu dalam pengambilan keputusan di masa mendatang, k-means merupakan metode clustering non hirarki yang berusaha mempartisi data yang ada kedalam bentuk satu atau lebih cluster atau dapat mempunyai tujuan untuk membagi data menjadi beberapa kelompok [9].

Berikut beberapa referensi dengan menggunakan algoritma k-means, menurut penelitian Yulia dan Mersi Silalahi yang telah menerapkan data mining clustering dalam mengelompokkan buku dengan metode k-means dari hasil penelitian yang terdiri dari 3 cluster yaitu cluster 0 sebanyak 12 item, cluster 1 sebanyak 9 item dan cluster 2 sebanyak 15 kali [10]. Menurut Daud Siburian, Sundari Retno Andani dan Eka Purnama Sari dimana penelitian ini menghasilkan nilai yang sama, yakni cluster tertinggi menghasilkan 6 jenis buku diantaranya matematika, geografi, kimia, PKN, penjaskesrek dan komputer. Sedangkan cluster terendah sebanyak 6 kali jenis buku yaitu Bahasa Indonesia,, Bahasa inggris, biologi, fisika, agama dan seni budaya. Sehingga dapat disimpulkan bahwa metode k-means pada penelitian ini dapat mengelompokkam peminjaman buku perpustakaan sekolah dengan baik, mengacu pada perhitungan manual [11]. Januardi Nasir telah melakukan penelitian dengan judul penerapan data mining clustering dalam mengelompokkan buku dengan metode k,means diperoleh data peminjaman buku yang telah diproses mendapatkan buku yang banyak dipinjam terdapat dalam cluster 3 sebanyak 6 item, buku yang paling sedikit dipinam terdapat pada cluster 2 sebanyak 15 item, buku yang cukup banyak dipinjam terdapat pada cluster 1 sebanyak 12 item [12].

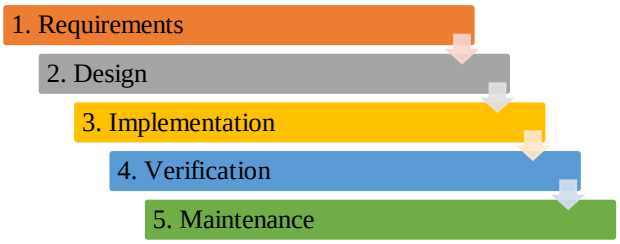
Implementasi data mining untuk pengelompokkan buku menggunakan algoritma k-means clustering (studi kasus : perpustakaan politeknik LPP Yogyakarta) merupakan penelitian yang dilakukan oleh Nisriina Nuur Hasanah dan Agus Sidiq Purnomo dengan tujuan penelitian ini untuk mengklasifikasi penyakit ISPA dan mendapatkan akurasi yang teoat dan cepat dalam mengklasifikasi gejala penyakit ISPA menggunakan metode k-means, dengan hasil Analisa persentase untuk tiap cluster adalah cluster pertama memiliki persentase 35% data, cluster kedua 45% data dan cluster ketiga 20% data, dan pengujian menggunakan validasi DBI (*davies bouldin index)* diperoleh nilai tiap-tiap cluster, pengujian cluster 1 menghasilkan nilai DBI -0,244 cluster 2 nilai DBI -0,250 dan cluster 3 nilai DBI -0,239 karena nilai DBI dari cluster 3 lebih kecil maka cluster tersebut bisa disebut optimal [13]. Begitupun juga penelitianCornelia Selvi Dinta Sembiring, Latifah Hanum dan Saut Parsaoran Tamba yang melakukan penerapan data mining menggunakan algoritma k-means untuk menentukan judul skripsi dan jurnal penelitian (studi kasus FTIK UNPRI) dengan tujuan untuk mengelompokkan data mahasiswa sesuai dengan skill dan basic yang didominasi pada matakuliah yang paling banyak diminati dengan perbandingan score dalam pembagian clustering yaitu pada 29% C1, 21% C2, 22%C3, 13%C4 dan 15%C5 [14].

Berdasarkan dengan uraian dari penelitian yang telah dilakukan tersebut, penelitian ini bertujuan menerapkan data mining Tujuan dari penelitian ini adalah menerapkan data mining dalam mengelompokkan buku yang di pinjam berdasarkan dengan jumlah ketersediaan buku, sehingga dengan mengimplementasikan pengelompokkan data dengan algoritma k-means ini akan menentukan buku yang paling banyak diminati yang dilihat dari jumlah ketersediaan buku, buku yang dibaca, dan buku yang dipinjam di Perpustakaan Universitas Dehasen Bengkulu. Sehingga k-means clustering meminimalisasi object function yang telah diatur pada proses clusterisasi, dengan cara minimalisasi variasi antar 1 cluster dengan maksimalisasi variasi dengan data di cluster lainnya [15].

2. METODOLOGI PENELITIA

2.1 Metode *Waterfall*

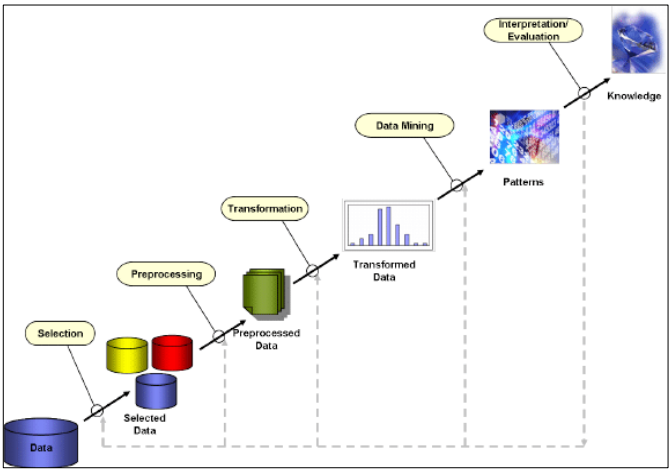
Metode penelitian yang digunakan penulis adalah metode *Waterfall*. Metode *waterfall* merupakan metode yang sering digunakan oleh penganalisa sistem pada umumnya [16]. Metode *waterfall* sering dinamakan siklus hidup klasik (*classic life cycle*) yang bersifat sistematis, berurutan dalam membangun software [17]. Model ini merupakan model satu arah yang dimulai dari tahap persiapan sampai perawatan, dimana hal ini menggambarkan pendekatan yang sistematis dan juga berurutan pada pengembangan perangkat lunak, dimulai dengan spesifikasi kebutuhan pengguna lalu berlanjut melalui tahapan-tahapan *Requirement*, perancangan (*design*), Implementasi (*implementation*), verifikasi (*verification)* dan pemeliharaan (*maintenance*) [18] serta penyerahan sistem ke para pelanggan/pengguna (*deployment*) yang diakhiri dengan dukungan pada perangkat lunak lengkap yang dihasilkan.



Gambar 1. Metode *Waterfall*

Dari gambar 1 merupakan tahapan dari metode Waterfall beberapa diantaranya adalah sebagai berikut [19] :

a. *Requirement* : tahap ini pengembang sistem diperlukan komunikasi yang bertujuan untuk memahami perangkat lunak yang diharapkan oleh pengguna dan Batasan perangkat lunak tersebut. Informasi dapat diperoleh melalui wawancara, diskusi atau survei langsung. Informasi dianalisis untuk mendapatkan data yang dibutuhakn oleh pengg*una.*
b. *Design* : pada tahap ini, pengembang membuat desain sistem yang dapat membantu menentukan perangkat keras (*hardware)* dan persyaratan dan persyaratan dan juga membantu dalam mendefinisikan arsitektur sistem secara keseluruhan.



Gambar 2. Proses *Knowledge Discovery in Database* (KDD)

Dari gambar 2 rangkaian proses iterative dari data mining dalam KDD [20] sebagai berikut:

1. Data cleaning, menghilangkan noise dan data yang inkonsisten.
2. Data integration, menggabungkan data dari berbagai sumber data yang berbeda
3. Data selection, mengambil data yang relevan dengan tugas analisis dari database
4. Data transformation, Mentransformasi atau menggabungkan data ke dalam bentuk yang sesuai untuk penggalian lewat operasi summary atau aggregation.
5. Data mining, proses esensial untuk mengekstrak pola dari data dengan metode cerdas.
6. Pattern evaluation, mengidentifikasikan pola yang menarik dan merepresentasikan pengetahuan berdasarkan interestingness measures.
7. Knowledge presentation, penyajian pengetahuan yang digali kepada pengguna dengan menggunakan visualisasi dan teknik representasi pengetahuan.

c. *Implementation* : pada tahap ini, sistem pertama kali dikembangkan di program kecil yang disebut unit, yang terintegrasi dalam tahap selanjutnya [21]. Setiap unit dikembangkan dan diuji untuk fungsionalitas yang disebut sebagai unit testing. Dalam penelitian ini penerapan data mining menggunakan algoritma *k-means* yang cara kerjanya membagi data numeric menjadi beberapa cluster. Algoritma *k-means* memiliki proses kerja yaitu :
   1. Menentukan k untuk kluster yang dianalisis
   2. Menentukan k sebagai titik pusat secara acak
   3. Menghitung jarak dari setiap k dari centroid dari cluster menggunakan rumus Euclidean distance.
   4. Mengalokasikan objek ke dalam centroid terdekat.
   5. Proses iterasi, lalu menentukan tempat centroid yang baru.
   6. Mengulangi Langkah no 3 jika tempat centroid tidak sama.
d. *Verification* : pada tahap ini, sistem dilakukan verifikasi dan pengujian apakah sistem sepenuhnya atau sebagian memenuhi persyaratan sistem, pengujian dapat dikategorikan kedalam unit testing.
e. *Maintenance* : ini adalah tahap akhir dari metode *waterfall*, perangkat lunak yang sudah jadi dijalanklan serta dilakukan pemeliharaan. Pemeliharaan termasuk dalam memperbaiki kesalahan yang tidak ditemukan pada Langkah sebelumnya.

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Algoritma K-Means dalam Pengelompokan Judul Buku yang dipinjam

Sampel data yang digunakan dalam penelitian ini adalah data peminjaman buku pada bulan juli 2022 di Perpustakaan Dehasen Bengkulu dapat dilihat pada table 1.

Tabel 1. Sampel Data Peminjaman Buku

Kode Buku	Judul Buku	Jml Buku
420 JUW h	1 Hari Tuntas Menguasai Grammar	10
420 Suk j	1 Jam Pintar Percakapan Bahasa Inggris	10
005 SUD a	10 Animasi Kartun Flash	10
005.1 UTA I	10 Langkah Belajar Logika dan Algoritma, menggunakan Bahasa C dan C++ di GNU dan Linux	10
001.422.0285 KOM m	10 Model Penelitian dan Pengolahan dengan SPSS 10.01	10
001.422.0285 KOM 1	10 Model Penelitian dan Pengolahannya dengan SPSS 14	15
005 BUD p	10 Proyek Robot Spektakuler	15
658.7 KOM c	100 Celah Keamanan Windows XP	15
371.9 KEW 1	100 Ide Membimbing Anak ADHD	10
371.9 BRO 1	100 Ide Membimbing Anak Autis	10
006.76 SIA 1	100 Kasus Pemrograman Visual C#	25
300 CHA w	100 Ways to Motivate Others	25
410 ARI s	1001 Kesalahan Berbahasa	25
650 SAT 1	101 Bisnis Laris Modal di Bawah 10 juta	40
006.69 MUL j	101 Jurus Cepat Menguassai Adobe Photoshop CS3 untuk Photographer	40
006.8 ISM 1	101 Kreasi Coreldraw X4	30
796.334 CHA s	101 Sesi Latihan Sepak Bola untuk Pemain Muda	30
005.3 ENT t	101 Tip & Trik Coreldraw X4	30
005.74 ENT t	101 Tip & Trik Microsoft Access 2007	25
005 ENT m	101 Tip dan Trik Dunia Kerja	25

Dalam penerapan algoritma K-Means dalam pengelompokkan judul buku yang dipinjam terdiri dari 5 (lima) tahapan adalah sebagai berikut [22] :

3.1.1 Menentukan k sebagai jumlah *cluster* yang ingin dibentuk

Pada penelitian penerapan data mining untuk pengelompokkan judul buku yang dipinjam di perpustakaan Universitas Dehasen Bengkulu menggunakan algoritma k-means, cluster yang ingin dibentuk berjumlah 2 cluster yaitu cluster 1 Tinggi dan cluster 2 rendah.

3.1.2 Membangkitkan nilai *random* untuk pusat *cluster* awal (*centroid*) sebanyak k.

Penentuan titik awal cluster diambil dari data yang telah didapatkan pada Tabel 1. dengan mengambil nilai tertinggi untuk cluster I dan nilai terendah untuk cluster II. Titik awal cluster diambil secara acak dengan mengacu pada nilai maksimum dan nilai minimum pada data yang didapatkan.

a. Titik awal Cluster I (Tinggi) : {40; 17; 20}
b. Titik awal Cluster II (Rendah) : {10; 2; 3}

3.1.3 Menghitung jarak setiap data input terhadap masing-masing *centroid*

Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidean* (*Euclidean Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Cluster I : $d(x,y) = \sqrt{(40-10)^2 + (17-4)^2 + (20-3)^2} = 36{,}85$

Cluster II : $d(x,y) = \sqrt{(10-10)^2 + (2-4)^2 + (3-3)^2} = 2$

Dan seterusnya sehingga menghasilkan nilai seperti Tabel 2.

Tabel 2. Nilai Euclidean

Kode Buku	Jml Buku dibaca	Jml Buku dipjm	C1	C2	Jarak Terdekat
420 JUW h	4	3	36,85	2	C2
420 Suk j	4	5	35,97	2,82	C2
005 SUD a	5	3	36,51	3	C2
005.1 UTA I	4	6	35,56	3,60	C2
001.422.0285 KOM m	4	4	36,40	2,24	C2
001.422.0285 KOM 1	7	4	31,32	7,14	C2
005 BUD p	7	5	30,82	7,34	C2
658.7 KOM c	8	4	31,01	7,87	C2
371.9 KEW 1	2	4	37,16	1	C2
371.9 BRO 1	5	5	35,62	3,61	C2
006.76 SIA 1	11	4	22,74	17,52	C2
300 CHA w	13	9	19,03	19,54	C1
410 ARI s	8	5	23,04	16,28	C2
650 SAT 1	7	19	10,05	34,37	C1
006.69 MUL j	7	15	11,18	32,69	C1
006.8 ISM 1	11	8	16,73	22,49	C1
796.334 CHA s	14	5	18,27	23,41	C1
005.3 ENT t	17	6	17,20	25,18	C1
005.74 ENT t	3	20	20,52	22,69	C1
005 ENT m	3	17	20,74	20,54	C2

3.1.4 Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil)

Dari hasil Tabel 2 tersebut setiap pertanyaan akan dipindahkan ke masing-masing cluster berdasarkan hasil *Euclidean*.

Tabel 3. *Cluster I* (tinggi*)*

Kode Buku	Judul Buku	Jml Buku	Jml Buku Dibaca	Jml Buku Dipinjam
300 CHA w	100 Ways to Motivate Others	25	13	9
420 Suk j	1 Jam Pintar Percakapan Bahasa Inggris	25	3	20
006.8 ISM 1	101 Kreasi Coreldraw X4	30	11	8
796.334 CHA s	101 Sesi Latihan Sepak Bola untuk Pemain Muda	30	14	5
005.74 ENT t	101 Tip & Trik Microsoft Access 2007	30	17	6
650 SAT 1	101 Bisnis Laris Modal di Bawah 10 juta	40	7	19
006.69 MUL j	101 Jurus Cepat Menguassai Adobe Photoshop CS3 untuk Photographer	40	7	15

Tabel 4. *Cluster II (*rendah*)*

Kode Buku	Judul Buku	Jml Buku	Jml Buku Dibaca	Jml Buku Dipinjam
371.9 KEW 1	100 Ide Membimbing Anak ADHD	10	2	4
371.9 BRO 1	100 Ide Membimbing Anak Autis	10	5	5
420 JUW h	1 Hari Tuntas Menguasai Grammar	10	4	3
420 Suk j	1 Jam Pintar Percakapan Bahasa Inggris	10	4	5
005 SUD a	10 Animasi Kartun Flash	10	5	3
005.1 UTA I	10 Langkah Belajar Logika dan Algoritma, menggunakan Bahasa C dan C++ di GNU dan Linux	10	4	6
001.422.0285 KOM m	10 Model Penelitian dan Pengolahan dengan SPSS 10.01	10	4	4
001.422.0285 KOM 1	10 Model Penelitian dan Pengolahannya dengan SPSS 14	15	7	4
005 BUD p	10 Proyek Robot Spektakuler	15	7	5
658.7 KOM c	100 Celah Keamanan Windows XP	15	8	4
006.76 SIA 1	100 Kasus Pemrograman Visual C#	25	11	4
410 ARI s	1001 Kesalahan Berbahasa	25	8	5
005 ENT m	101 Tip dan Trik Dunia Kerja	25	3	17

3.1.5 Memperbaharui Nilai *Centroid.*

Nilai *centroid* baru diperoleh dari rata-rata *cluster* .

$$v = \frac{\sum_{i=1}^{n} x_i}{n}$$

Titik Centroid Baru Cluster I : {31,43; 10,29; 11,71}
Titik Centroid Baru Cluster II : {14,62; 5,54; 5,31}

3.1.6 Menghitung jarak setiap data input terhadap masing-masing *centroid.*

Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidean* (*Euclidean Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Tabel 5. Nilai Euclidean

Kode Buku	Cluster I	Cluster II	Jarak Terdekat
420 JUW h	23,97	5,3846	C2
420 Suk j	23,32	4,87	C2
005 SUD a	23,73	5,19	C2
005.1 UTA I	23,05	4,91	C2
001.422.0285 KOM m	23,63	5,04	C2
001.422.0285 KOM 1	18,44	1,99	C2
005 BUD p	18,05	1,54	C2
658.7 KOM c	18,29	2,81	C2
371.9 KEW 1	24,24	5,96	C2
371.9 BRO 1	23,07	4,66	C2
006.76 SIA 1	10,07	11,81	C1
300 CHA w	7,49	13,31	C1
410 ARI s	9,57	10,68	C1
650 SAT 1	11,72	28,88	C1
006.69 MUL j	9,75	27,21	C1
006.8 ISM 1	4,04	16,55	C1
796.334 CHA s	7,81	17,56	C1
005.3 ENT t	8,93	19,19	C1
005.74 ENT t	12,77	18,17	C1
420 JUW h	11,06	15,84	C1

3.1.7 Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).

Dari hasil Tabel 5, tersebut setiap pertanyaan akan dipindahkan ke masing-masing cluster berdasarkan hasil euclidean.

Tabel 6. Cluster I (tinggi)

Kode Buku	Judul Buku	Jml Buku	Jml Buku Dibaca	Jml Buku Dipinjam
006.76 SIA 1	100 Kasus Pemrograman Visual C#	25	11	4
300 CHA w	100 Ways to Motivate Others	25	13	9
410 ARI s	1001 Kesalahan Berbahasa	25	8	5
005.74 ENT t	101 Tip & Trik Microsoft Access 2007	25	3	20
005 ENT m	101 Tip dan Trik Dunia Kerja	25	3	17
006.8 ISM 1	101 Kreasi Coreldraw X4	30	11	8
796.334 CHA s	101 Sesi Latihan Sepak Bola untuk Pemain Muda	30	14	5
005.3 ENT t	101 Tip & Trik Coreldraw X4	30	17	6
650 SAT 1	101 Bisnis Laris Modal di Bawah 10 juta	40	7	19
006.69 MUL j	101 Jurus Cepat Menguassai Adobe Photoshop CS3 untuk Photographer	40	7	15

Tabel 7. Cluster II (rendah)

Kode Buku	Judul Buku	Jml Buku	Jml Buku Dibaca	Jml Buku Dipinjam
420 JUW h	1 Hari Tuntas Menguasai Grammar	10	4	3
420 Suk j	1 Jam Pintar Percakapan Bahasa Inggris	10	4	5

005 SUD a	10 Animasi Kartun Flash	10	5	3
005.1 UTA I	10 Langkah Belajar Logika dan Algoritma, menggunakan Bahasa C dan C++ di GNU dan Linux	10	4	6
001.422.0285 KOM m	10 Model Penelitian dan Pengolahan dengan SPSS 10.01	10	4	4
371.9 KEW 1	100 Ide Membimbing Anak ADHD	10	2	4
371.9 BRO 1	100 Ide Membimbing Anak Autis	10	5	5
001.422.0285 KOM 1	10 Model Penelitian dan Pengolahannya dengan SPSS 14	15	7	4
005 BUD p	10 Proyek Robot Spektakuler	15	7	5
658.7 KOM c	100 Celah Keamanan Windows XP	15	8	4

3.1.8 Memperbaharui nilai *centroid*. Nilai *centroid* baru diperoleh dari rata-rata *cluster*

$$v = \frac{\sum_{i=1}^{n} x_i}{n}$$

Titik Centroid Baru Cluster I : {29,5; 9,4; 10,8}
Titik Centroid Baru Cluster II : {11,5; 5; 4,3}

Centroid baru ini dihitung menggunakan rumus seperti rumus iterasi per-1 dan karena tidak ada data yang berpindah cluster dan cluster ke-1 dan ke-2 hasilnya sama, maka proses centroid yang baru dihentikan dan berakhir pada iterasi ke-2.

3.1.9 Menghitung jarak setiap data input terhadap masing-masing *centroid.*

Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidean* (*Euclidean Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Tabel 8. Nilai Euclidean

Kode Buku	Cluster I	Cluster II	Jarak Terdekat
420 JUW h	21,69	2,22	C2
420 Suk j	21,05	1,93	C2
005 SUD a	21,46	1,98	C2
005.1 UTA I	20,79	2,48	C2
001.422.0285 KOM m	21,35	1,83	C2
001.422.0285 KOM 1	16,19	4,04	C2
005 BUD p	15,80	4,09	C2
658.7 KOM c	16,08	4,62	C2
371.9 KEW 1	21,94	3,37	C2
371.9 BRO 1	20,81	1,66	C2
006.76 SIA 1	8,31	14,78	C1
300 CHA w	6,04	16,38	C1
410 ARI s	7,47	13,85	C1
650 SAT 1	13,54	32,13	C1
006.69 MUL j	11,56	30,50	C1
006.8 ISM 1	3,26	19,79	C1
796.334 CHA s	7,42	20,58	C1
005.3 ENT t	9,00	22,11	C1
005.74 ENT t	12,08	20,80	C1
005 ENT m	9,98	18,64	C1

3.1.10 Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).

Dari hasil Tabel 8. tersebut setiap pertanyaan akan dipindahkan ke masing-masing cluster berdasarkan hasil euclidean.

Tabel 9. Cluster I (Tinggi)

Kode Buku	Judul Buku	Jml Buku	Jml Buku Dibaca	Jml Buku Dipinjam
006.76 SIA 1	100 Kasus Pemrograman Visual C#	25	11	4

300 CHA w	100 Ways to Motivate Others	25	13	9
410 ARI s	1001 Kesalahan Berbahasa	25	8	5
005.74 ENT t	101 Tip & Trik Microsoft Access 2007	25	3	20
005 ENT m	101 Tip dan Trik Dunia Kerja	25	3	17
006.8 ISM 1	101 Kreasi Coreldraw X4	30	11	8
796.334 CHA s	101 Sesi Latihan Sepak Bola untuk Pemain Muda	30	14	5
005.3 ENT t	101 Tip & Trik Coreldraw X4	30	17	6
650 SAT 1	101 Bisnis Laris Modal di Bawah 10 juta	40	7	19
006.69 MUL j	101 Jurus Cepat Menguassai Adobe Photoshop CS3 untuk Photographer	40	7	15

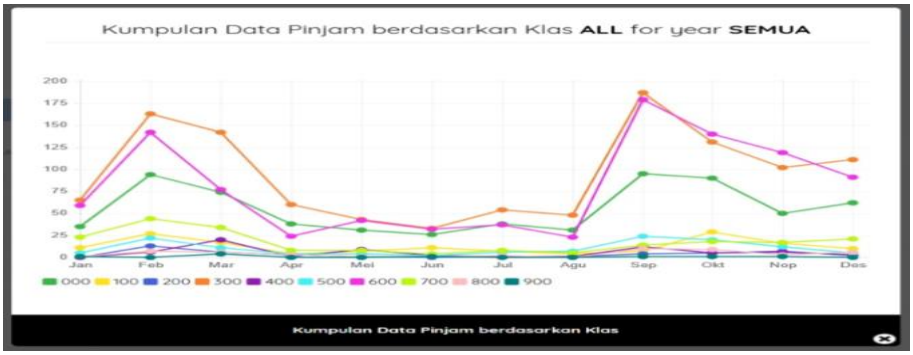
Tabel 10. Cluster II (Rendah)

Kode Buku	Judul Buku	Jml Buku	Jml Buku Dibaca	Jml Buku Dipinjam
420 JUW h	1 Hari Tuntas Menguasai Grammar	10	4	3
420 Suk j	1 Jam Pintar Percakapan Bahasa Inggris	10	4	5
005 SUD a	10 Animasi Kartun Flash	10	5	3
005.1 UTA I	10 Langkah Belajar Logika dan Algoritma, menggunakan Bahasa C dan C++ di GNU dan Linux	10	4	6
001.422.0285 KOM m	10 Model Penelitian dan Pengolahan dengan SPSS 10.01	10	4	4
371.9 KEW 1	100 Ide Membimbing Anak ADHD	10	2	4
371.9 BRO 1	100 Ide Membimbing Anak Autis	10	5	5
001.422.0285 KOM 1	10 Model Penelitian dan Pengolahannya dengan SPSS 14	15	7	4
005 BUD p	10 Proyek Robot Spektakuler	15	7	5
658.7 KOM c	100 Celah Keamanan Windows XP	15	8	4

Pada iterasi ke dua titik pusat pada setiap *cluster* sudah tidak berubah dan tidak ada lagi terdapat data yang berpindah dari satu *cluster* ke *cluster* yang lain. Maka dari itu, judul buku telah terkelompok menjadi 2 cluster sesuai kriteria yang ditentukan yaitu jumlah buku yang tersedia, jumlah buku yang di baca dan jumlah buku yang dipinjam menggunakan algoritma k-means, dari cluster I (tinggi) dari pengelompokkan data k-means pada Tabel 9 terdapat 10 jumlah kode dan judul buku dengan jumlah buku yang tersedia sebanyak 295 dengan rata-rata jumlah buku yang dibaca adalah 9 buku dan rata-rata jumlah buku yang dipinjam sebanyak 11 buku dan pada cluster II (rendah) dari pengelompokkan data k-means pada Tabel 10 terdapat 10 jumlah kode dan judul buku dengan jumlah buku yang tersedia sebanyak 115 dengan rata-rata jumlah buku yang dibaca adalah 5 buku dan rata-rata jumlah buku yang dipinjam sebanyak 4 buku.

3.2 Hasil Pengujian Penerapan Algoritma K-Means dalam Pengelompokkan Data Pinjam

Untuk mendapatkan hasil pengelompokkan pada tahap selanjutnya dilakukan tahapan dengan menampilkan hasil akhir dalam mengelompokkan jumlah peminjaman buku pada tahun 2022 pada perpustakaan UNIVED, sehingga dapat diketahui hasil pengelompokkan seperti yang terlihat pada gambar 3.



Gambar 3. Sebaran Cluster k-means

4. KESIMPULAN

Berdasarkan dengan penelitian yang dilakukan maka penelitian ini menghasilkan pengelompokkan judul buku yang di pinjam di berdasarkan dengan jumlah ketersediaan buku, sehingga dengan mengimplementasikan pengelompokkan data dengan algoritma k-

means ini akan menentukan buku yang paling banyak diminati yang dilihat dari jumlah ketersediaan buku, buku yang dibaca, dan buku yang dipinjam. Dari hasil penerapan data mining untuk pengelompokkan judul buku yang dipinjam di perpustakaan universitas dehasen bengkulu menggunakan algoritma k-means yang telah diterapkan judul buku telah terkelompok menjadi 2 cluster sesuai kriteria yang ditentukan yaitu jumlah buku yang tersedia, jumlah buku yang di baca dan jumlah buku yang dipinjam dengan hasil perhitungan yang ditunjukkan dari cluster I (tinggi) dari pengelompokkan data k-means terdapat 10 jumlah kode dan judul buku dengan jumlah buku yang tersedia sebanyak 295 dengan rata-rata jumlah buku yang dibaca adalah 9 buku dan rata-rata jumlah buku yang dipinjam sebanyak 11 buku dan pada cluster II (rendah) dari pengelompokkan data k-means terdapat 10 jumlah kode dan judul buku dengan jumlah buku yang tersedia sebanyak 115 dengan rata-rata jumlah buku yang dibaca adalah 5 buku dan rata-rata jumlah buku yang dipinjam sebanyak 4 buku.

REFERENCES

[1] T. Alfina T, B. Santosa, dan A.R. Barakbah, "Analisa Perbandingan Metode *Hierarchical Clustering,* K-means dan Gabungan Keduanya dalam Dluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)," Jurnal Teknik ITS. Vol.1, (Sept, 2012), ISSN :2301-9271,A-521-A-525.

[2] A. Riani, Y. Susianto, dan N. Rahman, "Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naïve Bayes," *Jurnal of Innovation Information Technology and Application (JINITA),* Vol.1, No.01, Desember 2019, pp 23-34, DOI : 10.35970/jinita.v1i01.64.

[3] D.S. Silitonga, S. Saefullah, dan R.Dewi, "Analisis Metode Naïve Bayes dalam Memprediksi Tingkat Pemahaman Mahasiswa terhadap Mata Kuliah berdasarkan Posisi Duduk," dalam *Prosiding Seminar Nasional Riset Information Science (SENARIS),* 2019, vol.1, hlm. 427-436.

[4] L. Iryani, "Penerapan Datamining Menentukan Minat Baca Mahasiswa Di Perpustakaan Universitas Bina Darma Palembang, Menggunakan Metode Clustering," *INTECOMS J. Inf. Technol. Comput. Sci.,* vol.3, no. 1, pp. 82-89, 2020, doi:10.31539/intercoms.v3i1.1251.

[5] Asminah, "Sistem Penentuan Penambahan Koleksi Buku di Perpustakaan Menggunakan Metode *K-Means Clustering*," *Journal of information system research (JOSH),* vol.4, No.1, pp 330-338, 2022, doi:10.47065/josh.v4i1.2383.

[6] D.A. Fakhri, S. Defit, and Sumijan, "Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-means Clustering," *j.Inf* dan *Teknol.,* vol.3, pp. 160-166, 2021, doi : 10.37034/jidt.v3i3.137.

[7] G.E.I. kambey, R. Sengkey, A. Jacobus, "Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia," *Jurnal Teknik Informatika,* vol.15, no. 2, hal. 75-82, 2020, p-ISSN : 2301-8402, e-ISSN : 2685-368X.

[8] Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia," *Paradigma-Jurnal Informatika dan Komputer,* vol. 22, No. 22, hal 183-188, 2018, DOI: http://doi.org/10.31294/p.v21i2.

[9] Nurmah, S. Nurajizah, A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan pada Toko Fashion Hijab Banten," *Jurnal Teknik Komputer AMIK BSI,* vol. 7, No. 2, hal 158-163, 2021, DOI : 10.31294/jtk.v4i2.

[10] Yulia, M.Silalahi, "Penerapan Data Mining Clustering dalam mengelompokkan buku dengan metode k-means," Indonesian Journal of Computer Science, Vol.10, No.1, Ed. 2021, Hal 77-89, ISSN 2302-4364 (print) dan 2549-7286 (online).

[11] D. Siburian, S. R. Andani, I.P. Sari, "Implementasi Algoritma k-means untuk pengelompokkan peminjaman buku pada perpustakaan sekolah," Journal of machine learning and artificial intelligence, Vol.1, No.2, Juni 2022, pp 115-124, DOI : 10-55123/jomlai.v1i2.725, ISSN : 2828-9102 (print), 2828-9099 (online)

[12] J. Nasir, "Penerapan Data Mining Clustering dalam Mengelompokkan Buku dengan Metode K-Means," Junal Simetris, Vol.2, No.2, November 2020, P-ISSN : 2252-4983, E-ISSN : 2549-3108.

[13] N. N. Hasanah, A.S. Purnomo, "Implementasi Data Mining untuk pengelompokkan Buku menggunakan Algoritma K-Means Clustering (Studi KAsus : Perpustakaan Politeknik LPP Yogyakarta)," Jurnal Teknologi dan Sistem Informasi Bisnis, Vol. 4, No. 2, Juli 2022, Hal. 300-311, DOI : http://doi.org/10.47233/jteksis.v4i2.499.

[14] C.S.D. Sembiring, L. Hanum, P. Tamba, "Penerapan data mining menggunakan algoritma k-means untuk menentukan judul skripsi dan jurnal penelitian (studi kasus FTIK UNPRI," Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA), Vol.5, No.2, Februari 2022, hal 80-85, E-ISSN : 2580-2879.

[15] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, "Covid-19 Case and Deaths in Southeast Asia Clustering using K-Means Algoritma ", Journal of physics : Conference Series, vol,1783, no. 1, p.012027, 2021.

[16] M. S. Rosa A.S, Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek., Cetakan Pe. Bandung : INFORMATIKA BANDUNG, 2018.

[17] A. B. Wahid, "Analisis Metode Waterfall Untuk Pengembangan Sistem Informaso," *Jurnal Ilmu-ilmu Informatika dan Manajemen STMIK,* oktober 2020, ISSN: 1978-3310, E-ISSN: 2615-3467.

[18] S. H. Bariah dan M. I. S. Putra, "Penerapan Metode Waterfall pada Perancangan Sistem Informasi Pengolahan Data Nilai Siswa," *Jurnal Petik,* vol. 6 no. 1, hal 1-6, 2020, p-ISSN: 2640-7363, e-ISSN: 2614-6606.

[19] A. Taufik, A Christian, T. Asra, "Perancangan Sistem Informasi Penjualan Peralatan Kesehatan Dengan Metode Waterfall," *Jurnal Teknik Komputer AMIK BSI, vol. V,no. 1, hal 59-64,* 2019, DOI : 10.31294/jtk.v4i2.

[20] Nurahman dan Prihandoko, "Perbandingan Hasil Analisis Teknik Data Mining 'Metode Decision, Tree Naive Bayes , Smo dan Part untuk Mendiagnosa Penyakit Diabetes Militus," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi,* vol. 4, no. 1, Jan 2019.

[21] N. A. Febriyati, A. D. GS, and A. Wanto. " GRDP Growth Rate Clustering in Surabaya City uses the K-Means Algorithm," International Journal of Information System & Technology, vol. 3, no. 2, pp. 276-283, 2020.

[22] M. A. Hanafiah, A. Wanto, and P.B. Indonesia, "Implementation of Data Mining Algoritms for Grouping Poverty Lines by Distric/City in North Sumatra," International Journal of Information System & Technology, vol. 3, no. 2, pp. 315-322, 2020.