# K-MEANS CLUSTERING MENGGUNAKAN RAPIDMINER DALAM SEGMENTASI PEMINJAMAN BUKU (STUDI KASUS: UNIVERISTAS NASIONAL)

**Abidah Dzalulliyah, Agus Iskandar**
Sistem Informasi, Universitas Nasional
Jalan Sawo Manila, Pejaten Barat, Pasar Minggu, RT.14/RW.3, RT.14/RW.3, Ps. Minggu,
Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12520, Indonesia
*abidahdzalulliyah2021@student.unas.ac.id*

**ABSTRAK**

*Cyber Library* telah berhasil menyediakan layanan digital, seperti akses ke jurnal internasional dan *e-book*, serta sistem peminjaman yang terintegrasi secara daring. Akan tetapi, layanan *Cyber Library* Universitas Nasional perlu di optimalkan untuk meningkatkan minat baca pengunjung. Oleh karena itu, penelitian ini melakukan analisis *K-Means Clustering* menggunakan *RapidMiner* dalam segmentasi peminjaman buku. Tujuan penelitian ini adalah menganalisis preferensi buku dengan menentukan kategori buku yang paling banyak dan paling sedikit dipinjam, mengelompokkan data peminjaman buku berdasarkan frekuensi durasi peminjaman, melakukan segmentasi peminjaman berdasarkan frekuensi peminjaman, mengidentifikasi pola kunjungan dengan menggunakan data tipe keanggotaan. Data dianalisis berdasarkan algoritma *K-Means Clustering* menggunakan metode *Knowledge Discovery in Database* (KDD). Terdapat 4 hasil pada penelitian ini yaitu: (1) Kategori buku pada *cluster* 1 (kategori buku yang sedikit di pinjam) yaitu kategori 000, 100, 200, 400, 500, 600, 700, 800, dan 900, serta pada *cluster* 2 (kategori buku yang banyak di pinjam) yaitu kategori 300, (2) Frekuensi durasi peminjaman pada *cluster* 1 (frekuensi durasi peminjaman yang jarang muncul) yaitu durasi peminjaman buku 0, 1, 3, 4, 8, 9, 12, 14, dan 17 hari, dan pada *cluster* 2 (frekuensi durasi peminjaman yang sering muncul) yaitu durasi peminjaman buku 7 hari, (3) Frekuensi peminjaman pada *cluster* 1 (kategori yang sedikit di pinjam) yaitu kategori komputer dan informasi umum, filsafat dan psikologi, agama, bahasa, sains dan matematika, teknologi dan ilmu terapan, seni dan hiburan, sastra, serta sejarah dan geografi, dan pada *cluster* 2 (kategori yang banyak dipinjam) yaitu kategori ilmu sosial, (4) Tipe keanggotaan pada *cluster* 1 (tipe anggota yang jarang berkunjung ke perpustakaan) yaitu karyawan, dosen UNAS, mahasiswa Universitas Siber Asia, siswa SMP YMIK, siswa SMK YMIK, umum, dosen UNSIA, dan pengunjung bukan anggota, serta *cluster* 2 (tipe anggota yang sering berkunjung) yaitu mahasiswa Universitas Nasional.

**Kata Kunci**: *Data Peminjaman, Data Pengunjung, Algoritma K-Means Clustering, Aplikasi RapidMiner, Cyber Library Universitas Nasional.*

## 1. PENDAHULUAN

Perpustakaan adalah lembaga yang berfungsi sebagai pusat penyimpanan, pengelolaan, dan penyebaran informasi. Secara tradisional, perpustakaan dikenal sebagai tempat menyimpan koleksi buku, manuskrip, dan dokumen lainnya. Namun, fungsi perpustakaan tidak hanya terbatas pada penyimpanan, melainkan juga memberikan layanan informasi kepada masyarakat, mendukung pendidikan, serta menjadi pusat kegiatan intelektual. Perpustakaan modern telah mengalami evolusi signifikan dibandingkan dengan perpustakaan zaman dulu.

Saat ini, perpustakaan tidak hanya menyediakan buku fisik tetapi juga koleksi digital seperti e-book, jurnal elektronik, dan database daring. Perpustakaan modern berfungsi sebagai pusat komunitas, menawarkan ruang untuk diskusi, pelatihan, dan kegiatan sosial. Perpustakaan di Indonesia memiliki karakteristik tersendiri, sering kali berfokus pada pengembangan literasi masyarakat dan pelestarian budaya lokal. Namun, dibandingkan dengan perpustakaan internasional, banyak perpustakaan di Indonesia yang masih menghadapi tantangan dalam hal pendanaan, teknologi, dan koleksi yang memadai. Meski demikian, ada perpustakaan di Indonesia yang telah menerapkan teknologi modern.

*Cyber Library* di Universitas Nasional adalah salah satu perpustakaan modern di Indonesia yang menawarkan layanan berbasis teknologi. Perpustakaan *Cyber Library* Universitas Nasional menyediakan ruang belajar, akses ke sumber daya elektronik, dan fasilitas diskusi. Menurut [1], layanan perpustakaan seperti menyediakan koleksi buku yang memadai, sistem perpustakaan yang jelas, dan lingkungan perpustakaan yang kondusif sangat penting untuk diperhatikan. Oleh karena itu, terdapat tantangan yang dihadapi oleh pengelola *Cyber Library* Universitas Nasional dalam mengoptimalkan layanan dan meningkatkan minat baca pengunjung.

Pada penelitian yang dilakukan oleh [2], terdapat beberapa atribut data yang digunakan oleh penulis. Atribut tersebut akan tetapi hanya terbatas pada variabel bulan kunjungan, fakultas, dan golongan koleksi yang dipinjam. Penelitian tersebut tidak mempertimbangkan durasi peminjaman sebagai faktor dalam clustering. Identifikasi pola peminjaman buku pada akhirya kurang dapat dianalisis. Dengan menggunakan atribut data seperti kategori buku, durasi peminjaman, frekuensi peminjaman, dan tipe

keanggotaan, penelitian ini dapat memberikan wawasan dan analisis yang lebih mendalam tentang pola peminjaman, preferensi peminjaman, dan keaktifan pengunjung

Penelitian ini bertujuan untuk melakukan segmentasi peminjaman buku dengan klasterisasi pada data kategori buku, frekuensi durasi peminjaman, dan pengunjung perpustakaan. Dalam penelitian ini, penulis menggunakan algoritma *K-Means Clustering* karena algoritma *K-Means* Clustering merupakan algoritma yang sederhana dan prosesnya cepat, sehingga algoritma ini menjadi populer dan banyak digunakan untuk penelitian [3]. Dengan mengintegrasikan beberapa metode, yaitu *Knowledge Discovery in Database* (KDD), *Data Mining*, *K-Means*, *Clustering*, *Euclidean Distance*, *Manhattan Distance*, *RapidMiner*, *Davis Bouldin Index* (DBI), dan *Dewey Decimal Classification* (DDC). Metode ini dipilih karena keunggulannya dalam memberikan hasil yang lebih relevan dan akurat dibandingkan metode lain yang tersedia. Dengan adanya penelitian ini diharapkan agar pengelola *Cyber Library* dapat mempertimbangkan kebijakan serta layanan untuk meningkatkan kepuasan pengunjung. Penulis juga berharap agar hal ini dapat menjadi sumber wawasan bagi penelitian-penelitian selanjutnya dengan pembahasan serupa.

## 2. TINJAUAN PUSTAKA

Membahas mengenai apa saja teori yang digunakan sebagai dasar penelitian ini:

### 2.1. *Knowledge Discovery in Database* (KDD)

*Knowledge Discovery in Database* (KDD) ialah metode sebagai mencari pengetahuan dan informasi yang belum ditemukan pada suatu database. Data mining ialah salah satu tahapan dari proses KDD secara keseluruhan dan merupakan inti dari proses KDD [4].

### 2.2. *Data Mining*

Data mining merupakan proses untuk analisis data yang dimana tujuannya sebagai menemukan hubungan yang signifikan dan mencapai kesimpulan yang mana pun belum pernah terjadi sebelumnya yaitu dengan menggunakan metode terkini, dapat menguntungkan pemilik data [5].

### 2.3. *K-Means*

*K-Means* ialah algoritma yang mengelompokkan data ke dalam berbagai kelompok berdasarkan karakteristik yang sama. Pada penelitian [6] algoritma *K-Means* merupakan mekanisme yang tidak memiliki hierarki yang menggunakan beberapa komponen populasi sebagai pusat klaster awal.

### 2.4. *Clustering*

*Clustering* merupakan teknik untuk melakukan pengelompokkan data ke dalam beberapa kelompok atau cluster berdasarkan kesamaan atau perbedaan di antara data terebut. Menurut [7] *clustering* ialah pengklasteran dari dataset atau suatu objek ke dalam sebuah kelompok atau yang disebut dengan cluster, jadi dalam setiap *cluster* yang telah terbentuk berisi suatu data yang serupa dan sangat berbeda dengan data dalam *cluster* yang lainnya.

### 2.5. *Euclidean Distance*

*Euclidean Distance* digunakan dalam menentukan kemiripan objek yang terdapat pada algoritma *K-Means Clustering*. Menurut [8] *Euclidean Distance* merupakan metode yang digunakan oleh matematikawan Yunani Euclid untuk melakukan perhitungan jarak yang ada antara dua titik di dalam ruang geometris.

### 2.6. *Manhattan Distance*

*Manhattan Distance* adalah metode untuk melakukan perhitungan jarak mutlak di antara dua koordinat objek. Jarak ini juga disebut sebagai metrik taxicab atau jarak antara blok kota. Menurut [9] *Manhattan Distance* dimanfaatkan untuk metode pengukuran jarak yang didasarkan pada jumlah selisih antara dua objek data.

### 2.7. RapidMiner

RapidMiner merupakan salah satu aplikasi *data science* dan *machine learning* yang dimanfaatkan dalam berbagai proses analisis data. Perangkat lunak ini dapat digunakan hampir di semua perangkat komputer, dan perangkat lunak ini memiliki versi komunitas sehingga pengguna dapat menggunakan perangkat lunak ini secara gratis [10].

### 2.8. *Davies-Bouldin Index*

*Davies Bouldin Index* (DBI) ialah metrik yang dipakai sebagai perhitungan evaluasi hasil algoritma klastering dengan menggunakan aplikasi RapidMiner versi 10.5.0. Tujuan dari Davies Bouldin Index ini ialah untuk mengetahui seberapa baik algoritma klastering memisahkan berbagai kelompok data yang berbeda dan mendekati pusat klaster [11].

### 2.9. *Dewey Decimal in Database*

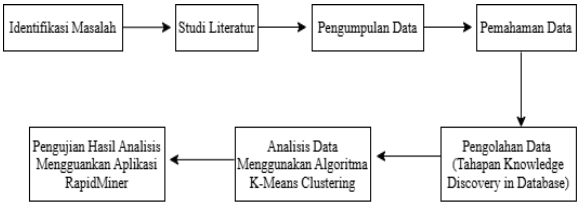
*Dewey Decimal Classification* (DDC) ialah sistem klasifikasi perpustakaan yang diciptakan oleh melvil dewey pada tahun 1876 dan menggunakan notasi decimal untuk mengorganisasi buku dan dokumen berdasarkan subjek atau bidang keilmuan.

## 3. METODE PENELITIAN

Membahas mengenai metode penelitian, tahapan penelitian, tahapan pengolahan data, tahapan *K-Means Clustering*, analisis menggunakan RapidMiner.

### 3.1. Tahapan Penelitian

Tahapannya sebagai berikut:

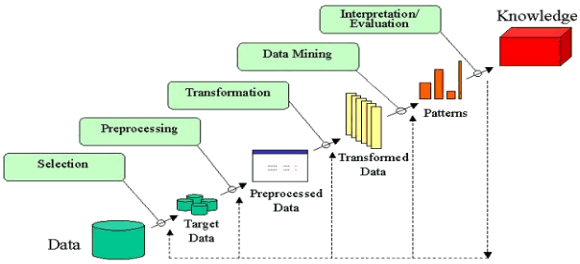



Gambar 1. Tahapan Penelitian

a. Identifikasi Masalah
   Hasil dari penelitian ini untuk melakukan pengelompokkan segmentasi peminjaman buku dan pengunjung dari data yang telah disediakan oleh *Cyber Library* Universitas Nasional.
b. *Studi Literatur*
   Melakukan referensi jurnal dari penelitian terdahulu dengan topik *Data Mining*, *Knowledge* Discovery *in Database* (KDD), *K-Means Clustering* dan yang berkaitan dengan penelitian ini.
c. Pengumpulan Data
   Melakukan pengumpulan data sekunder dan data primer, adapun data sekunder yaitu data yang sudah tersedia di *Cyber Library* data peminjaman buku dan data pengunjung dari tahun 2023-2024, sedangkan data primer yaitu membuat atribut baru dari data yang sudah tersedia. Atribut baru antara lain data kategori buku, durasi peminjaman, dan frekuensi peminjaman.
d. Pemahaman Data
   Mempelajari data yang telah diberikan oleh *Cyber Library* untuk dilakukan proses analisis data menggunakan algoritma *K-Means Clustering*.
e. Pengolahan Data
   Pengolahan data dilakukan melalui tahapan *Knowledge Discovery in Database* (KDD). Adapun tahapan KDD, sebagai berikut, *Data selection, Data preprocessing / cleaning, Data transformation, Data mining, Data interpretation*.
f. Analisis Data
   Pada tahap analisis data, algoritma pengklasteran *K-Means* digunakan untuk mengklasterkan data peminjaman buku dan data pengunjung perpustakaan.
g. Pengujian Hasil Analisis
   Pengujian hasil analisis digunakan untuk melakukan visualisasi dari perhitungan terhadap analisis data. Dengan menggunakan aplikasi RapidMiner versi 10.5.0.

### 3.2. Tahapan Pengolahan Data

Tahapan pengolahan data dilakukan dengan metode *Knowledge Discover in Database* tahapannya adalah sebagai berikut:




Gambar 2 Tahapan Pengolahan Data *[12]*

a. *Data Selection*
   Data yang di pilih untuk penelitian ini adalah data peminjaman buku dan data pengunjung *Cyber Library* dari tahun 2023-2024. Variabel data yang digunakan yaitu kategori buku, frekuensi durasi peminjaman, frekuensi peminjaman, dan tipe keanggotaan.
b. *Data Preprocessing*
   Pada penelitian ini dilakukan penghapusan data yang tidak digunakan pada tahap analisis.
c. *Data Transformation*
   Pada data peminjaman buku dengan membuat atribut baru yaitu kategori buku, durasi peminjaman, frekuensi peminjaman. Pada kategori buku melakukan pengelompokkan buku berdasarkan teori *dewey decimal classification* (DDC), durasi peminjaman di dapat dari hasil selisih antara tanggal pinjam dan tanggal kembali, frekuensi peminjaman di dapat dari seberapa sering judul buku muncul. Pada data pengunjung perpustakaan tipe keanggotaan di kelompokkan berdasarkan tahun yaitu 2023 dan 2024.
d. *Data Mining*
   Pengolahan data dengan memanfaatkan algoritma *K-Means* untuk melakukan klasterisasi.
e. *Data Evaluation*
   Evaluasi data pada proses mengubah pola untuk menghasilkan pengetahuan menggunakan aplikasi RapidMiner versi 10.5.0

### 3.3. Tahapan K-Means Clustering

Tahapan algoritma *K-Means Clustering* yang terdapat di flowchart:

a. Menentukan jarak antara setiap objek dengan koordinat titik tengah.
   - Menghitung jarak objek ke masing-masing centroid dengan persamaan *Euclidean Distance* dapat dihitung menggunakan rumus berikut ini [2]:

     $d(i,j)=\sqrt{(x1i-x1j)^2+(x2i-x2j)^2+\ldots+(xki-xkj)^2}$

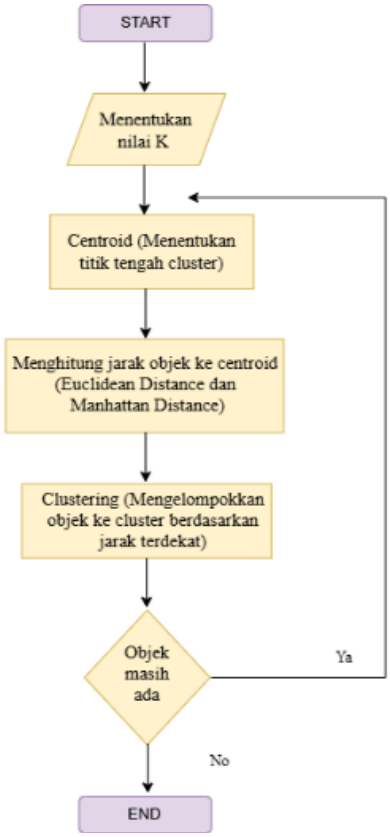
     Keterangan:
     d(i,j) = Jarak data ke i pusat cluster j
     xki = Data ke i pada atribut data ke k
     xkj = Titik pusat ke j pada atribut ke k
   - Menghitung jarak objek ke masing-masing centroid dengan persamaan *Manhattan Distance* dapat dihitung menggunakan rumus berikut ini [9]:

d(x,y) = $\sum_{i=1}^{n} |x_i - y_i|$
Keterangan:
d = Jarak antara x dan y
$x_i$ = Atribut dari data ke-i, (i =1,2,3,4,.....,n)
$y_i$ = Atribut pusat cluster ke-i, (i =1,2,3,4,.....,n)

b. Menghitung cluster baru
Untuk menentukkan posisi centroid baru menggunakan rumus berikut ini [2]:

$\mu j(t+1) = \frac{1}{Nsj} \sum_{jsj} xj$

Keterangan:
$\mu j(t+1)$ = Centroid baru di (t+1)
Nsj = Banyak data pada klaster Sj



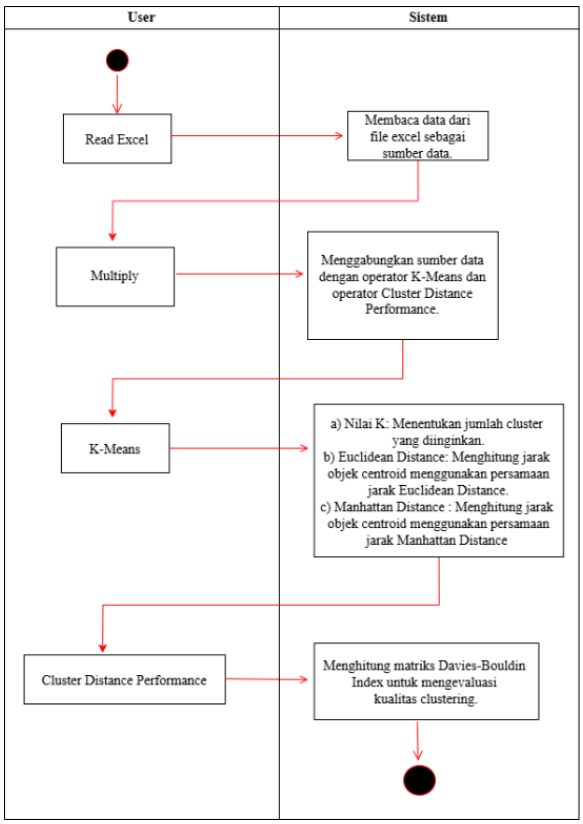

Gambar 3. Tahapan K-Means Clustering

### 3.4. Analisis Menggunakan RapidMiner

Penelitian ini memanfaatkan teknik analisis klasterisasi *K-Means Clustering* dengan penerapan pada aplikasi RapidMiner versi 10.5.0. Teknik ini merupakan metode klasterisasi yang berfungsi untuk menentukkan nilai K yang terbaik serta pada data penelitian ini membandingkan perhitungan jarak objek ke centroid menggunakan *Euclidean Distance* / *Manhattan Distance* yang terbaik. Untuk mengetahui perhitungan jarak objek ke centroid terbaik menggunakan *Euclidean Distance* / *Manhattan Distance* dengan melakukan perhitungan kualitas clustering pada metrik *Davies-Bouldin Index.*

Berikut ini langkah-langkah pada analisis klasterisasi yang terdapat pada *activity diagram* dengan algoritma K-Means menggunakan aplikasi RapidMiner:

a. *Start*: Memulai proses awal dari klasterisasi.
b. *Read* Excel: Membaca data dari file excel dan memasukkannya ke dalam proses klasterisasi.
c. *Multiply*: Menggabungkan sumber data dengan operator *K-Means* untuk menguji beberapa nilai K dan memilih nilai K yang menghasilkan nilai terbaik.
d. *K-Means*: Melakukan proses pengelompokkan data berdasarkan nilai K terbaik dengan menghitung jarak antara data dengan centroid mempergunakan *numerical measure Euclidean Distance dan Manhattan Distance*.
e. *Cluster distance performance*: Menghitung kualitas klastering dalam analisis data menggunakan *Davies Bouldin Index* untuk mengevaluasi hasil klasterisasi.
f. *End*: Menandai akhir dari proses klasterisasi.




Gambar 4. Activity Diagram Tahapan Analisis RapidMine

## 4. HASIL DAN PEMBAHASAN

Penelitian ini mengelompokkan data kategori buku, frekuensi durasi peminjaman, frekuensi peminjaman, dan tipe keanggotaan. Dengan melakukan perbanding nilai kelompok k=2 dengan k=3 untuk membandingkan nilai kelompok mana yang paling optimal, serta membandingkan perhitungan jarak *Euclidean Distance* dengan *Manhattan Distance*. Sehingga dihasilkan pengelompokkan sebagai berikut:

### 4.1. Kategori Buku

Tabel 1. Cluster Kategori Buku Dengan Nilai k=2

| Kategori Buku | Cluster |
|---|---|
| 000 | 1 |
| 100 | 1 |
| 200 | 1 |
| 300 | 2 |
| 400 | 1 |
| 500 | 1 |
| 600 | 1 |
| 700 | 1 |
| 800 | 1 |
| 900 | 1 |

Perhitungan jarak *Euclidean Distance* dan *Manhattan Distance* menghasilkan pengelompokkan yang sama. Pada K=2 berakhir di iterasi 2, sedangkan pada K=3 berakhir di iterasi 3.

Tabel 2. Cluster Kategori Buku Dengan Nilai k=3

| Kategori Buku | Cluster |
|---|---|
| 000 | 1 |
| 100 | 1 |
| 200 | 1 |
| 300 | 3 |
| 400 | 1 |
| 500 | 1 |
| 600 | 2 |
| 700 | 1 |
| 800 | 2 |
| 900 | 1 |

### 4.2. Frekuensi Durasi Peminjaman

Berdasarkan perhitungan jarak *Euclidean Distance* dan *Manhattan Distance* menghasilkan pengelompokkan yang sama. Pada K=2 berakhir di iterasi 2, sedangkan pada K=3 berakhir di iterasi 2.

Tabel 3. Cluster Frekuensi Durasi Peminjaman Dengan Nilai k=2

| Durasi Peminjaman | Cluster |
|---|---|
| 0 | 1 |
| 1 | 1 |
| 3 | 1 |
| 4 | 1 |
| 7 | 2 |
| 8 | 1 |
| 9 | 1 |
| 12 | 1 |
| 14 | 1 |
| 17 | 1 |

Tabel 4. Cluster Frekuensi Durasi Peminjaman Dengan Nilai k=3

| Durasi Peminjaman | Cluster |
|---|---|
| 0 | 3 |
| 1 | 3 |
| 3 | 3 |
| 4 | 3 |
| 7 | 2 |
| 8 | 1 |
| 9 | 1 |
| 12 | 3 |
| 14 | 3 |
| 17 | 3 |

### 4.3. Frekuensi Peminjaman

Berdasarkan perhitungan jarak *Euclidean Distance* dan *Manhattan Distance* menghasilkan pengelompokkan yang sama. Pada K=2 berakhir di iterasi 2, sedangkan pada K=3 berakhir di iterasi 3.

Tabel 1. Cluster Frekuensi Peminjaman Dengan Nilai k=2

| Frekuensi Peminjaman | Cluster |
|---|---|
| Komputer, Informasi Umum | 1 |
| Filsafat dan Psikologi | 1 |
| Agama | 1 |
| Ilmu Sosial | 2 |
| Bahasa | 1 |
| Sains dan Matematika | 1 |
| Teknologi dan Ilmu Terapan | 1 |
| Seni dan Hiburan | 1 |
| Sastra | 1 |
| Sejarah dan Geografi | 1 |

Tabel 2. Cluster Frekuensi Peminjaman Dengan Nilai k=3

| Frekuensi Peminjaman | Cluster |
|---|---|
| Komputer, Informasi Umum | 1 |
| Filsafat dan Psikologi | 1 |
| Agama | 1 |
| Ilmu Sosial | 3 |
| Bahasa | 1 |
| Sains dan Matematika | 1 |
| Teknologi dan Ilmu Terapan | 2 |
| Seni dan Hiburan | 1 |
| Sastra | 2 |
| Sejarah dan Geografi | 1 |

### 4.4. Tipe Keanggotaan

Berdasarkan perhitungan jarak *Euclidean Distance* dan *Manhattan Distance* menghasilkan pengelompokkan yang sama. Pada K=2 berakhir di iterasi 4, sedangkan pada K=3 berakhir di iterasi 2.

Tabel 3. Cluster Tipe Keanggotaan Dengan Nilai k=2

| Tipe Keanggotaan | Cluster |
|---|---|
| Mahasiswa Universitas Nasional | 2 |
| Karyawan | 1 |
| Dosen UNAS | 1 |
| Mahasiswa Universitas Siber Asia | 1 |
| Siswa SMP YMIK | 1 |
| Siswa SMK YMIK | 1 |
| Umum | 1 |
| Dosen UNSIA | 1 |
| Pengunjung Bukan Anggota | 1 |

Tabel 8. Cluster Tipe Keanggotaan Dengan Nilai k=3

| Tipe Keanggotaan | Cluster |
|---|---|
| Mahasiswa Universitas Nasional | 1 |
| Karyawan | 3 |
| Dosen UNAS | 3 |
| Mahasiswa Universitas Siber Asia | 3 |

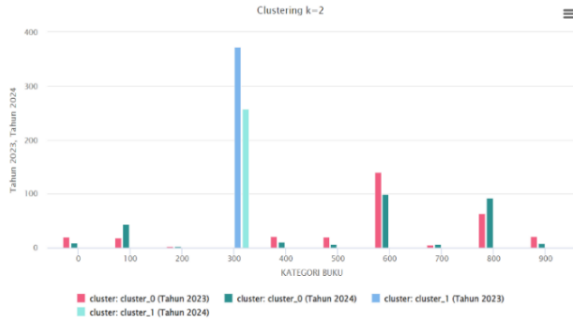
| Tipe Keanggotaan | Cluster |
|---|---|
| Siswa SMP YMIK | 3 |
| Siswa SMK YMIK | 3 |
| Umum | 3 |
| Dosen UNSIA | 3 |
| Pengunjung Bukan Anggota | 2 |

### 4.5. Penerapan Aplikasi RapidMiner

Pengelompokkan data peminjaman buku dan pengunjung perpustakaan menghasilkan hasil yang sama pada penerapan *K-Means Clustering*. Melakukan evaluasi metrik dengan *Davies Bouldin Index* untuk mengetahui nilai kelompok mana yang paling optimal.

### 4.6. Kategori Buku

Pada data kategori buku hasil pengelompokkan dengan nilai k=2 menggunakan aplikasi RapidMiner menghasilkan *cluster* 0 kategori buku yang sedikit di pinjam yaitu kategori buku 000, 100, 200, 400, 500, 600, 700, 800, dan 900, serta *cluster* 1 kategori buku yang banyak di pinjam yaitu kategori buku 300.

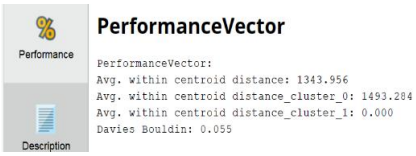



Gambar 5. Visualisasi Clustering Kategori Buku k=2

Hasil pengelompokkan dengan nilai k=3 menggunakan aplikasi RapidMiner menghasilkan *cluster* 0 sedikit di pinjam kategori buku 000, 100, 200, 400, 500, 700, 900. Pada *cluster* 1 cukup di pinjam kategori buku 600 dan 800. Dan untuk *cluster* 2 yang banyak di pinjam kategori buku 300.

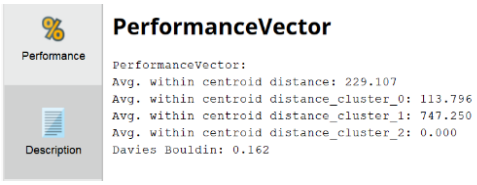

Gambar 6. Visualisasi Clustering Kategori Buku k=3

Hasil dari perhitungan jarak menggunakan *Euclidean Distance* dengan *Manhattan Distance* menghasilkan nilai yang sama ketika sudah di hitung dengan metrik evaluasi *Davies Bouldin Index*, sehingga menghasilkan nilai *Davies Bouldin Index* pada k=2 adalah 0,055.




Gambar 7. Metrik Evaluasi Dengan Nilai k=2

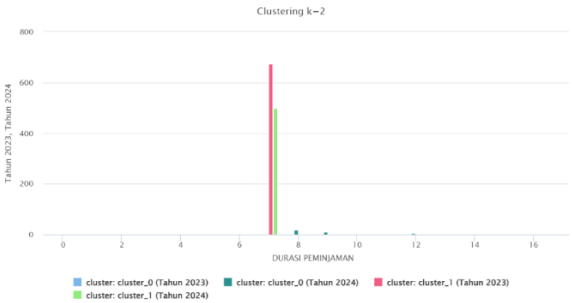
Nilai *Davies Bouldin Index* pada k=3 adalah 0,162. Nilai *Davies Bouldin Index* yang mendekati nilai 0 itu yang terbaik, oleh karena itu nilai kelompok k=2 yang menghasilkan nilai *Davies Bouldin Index* terbaik.




Gambar 8 Metrik Evaluasi Dengan Nilai k=3

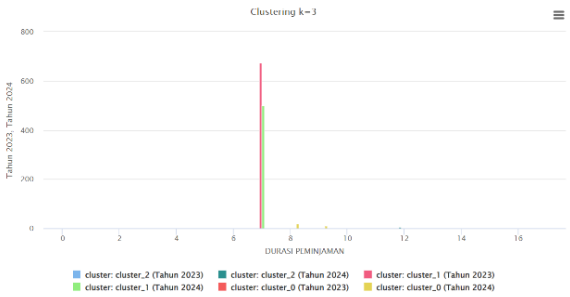
### 4.7. Frekuensi Durasi Peminjaman

Hasil pengelompokkan dengan nilai k=2 menggunakan aplikasi RapidMiner menghasilkan *cluster* 0 ialah kelompok data dengan frekuensi durasi peminjaman yang jarang muncul yaitu peminjaman buku dengan durasi 0, 1, 3, 4, 8, 9, 12, 14, dan 17 hari. *Cluster* 1 ialah kelompok data dengan frekuensi durasi peminjaman yang sering muncul yaitu peminjaman buku dengan durasi 7 hari.



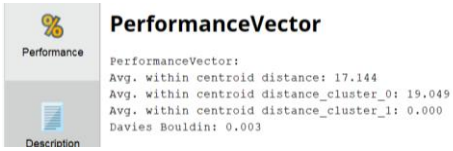

Gambar 5. Visualisasi Clustering Frekuensi Peminjaman k=2

Hasil pengelompokkan dengan nilai k=3 menggunakan aplikasi RapidMiner menghasilkan *cluster* 0 ialah kelompok data dengan frekuensi durasi peminjaman dengan kemunculan sedang yaitu peminjaman buku dengan durasi 8 dan 9 hari. *Cluster* 1 ialah kelompok data dengan frekuensi durasi peminjaman yang sering muncul yaitu peminjaman buku dengan durasi 7 hari. *Cluster* 2 ialah kelompok data dengan frekuensi durasi peminjaman yang jarang muncul yaitu peminjaman buku dengan durasi 0, 1, 3, 4, 12, 14, dan 17 hari.



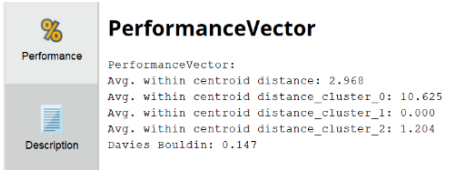

Gambar 10. Visualisasi Clustering Frekuensi Peminjaman k=3

Pada data durasi peminjaman hasil dari perhitungan jarak menggunakan *Euclidean Distance* dengan *Manhattan Distance* menghasilkan nilai yang sama ketika sudah di hitung dengan metrik evaluasi yaitu *Davies Bouldin Index*, sehingga menghasilkan nilai *Davies Bouldin Index* pada k=2 adalah 0,003.




Gambar 11. Metrik Evaluasi Dengan Nilai k=2

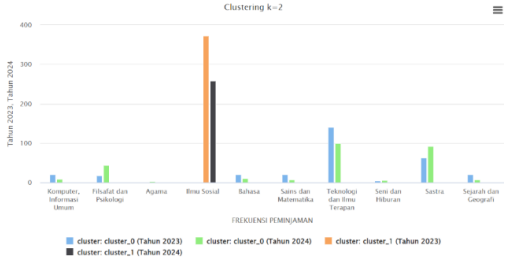
Nilai *Davies Bouldin Index* pada k=3 adalah 0,147. Nilai *Davies Bouldin Index* yang mendekati nilai 0 itu yang terbaik, oleh karena itu nilai kelompok k=2 yang menghasilkan nilai *Davies Bouldin Index* terbaik.




Gambar 12. Metrik Evaluasi Dengan Nilai k=3

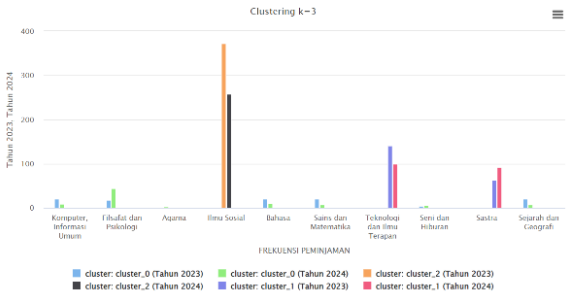
### 4.8. Frekuensi Peminjaman

Hasil pengelompokkan dengan nilai k=2 menggunakan aplikasi RapidMiner menghasilkan pengelompokkan pada *cluster* 0 dengan frekuensi pemijaman judul buku yang sedikit di pinjam yaitu kategori komputer dan informasi umum, filsafat dan psikologi, agama, bahasa, sains dan matematika, teknologi dan ilmu terapan, seni dan hiburan, sastra, sejarah dan geografi. Sementara itu, *cluster* 1 dengan frekuensi peminjaman judul buku yang banyak di pinjam yaitu kategori ilmu sosial.



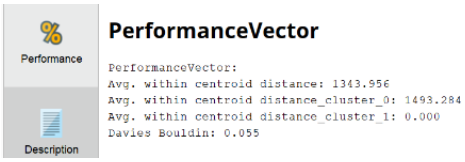

Gambar 13. Visualisasi Frekuensi Peminjaman k=2

Hasil pengelompokkan dengan nilai k=3 menggunakan aplikasi RapidMiner menghasilkan pengelompokkan pada *cluster* 0 dengan frekuensi pemijaman judul buku yang sedikit di pinjam yaitu kategori komputer dan informasi umum, filsafat dan psikologi, agama, bahasa, sains dan matematika, seni dan hiburan, sejarah dan geografi. Sementara itu, *cluster* 1 dengan frekuensi peminjaman judul buku yang cukup di pinjam yaitu teknologi dan ilmu terapan, dan sastra, serta *cluster* 2 dengan frekuensi peminjaman judul buku yang banyak di pinjam yaitu kategori ilmu sosial.



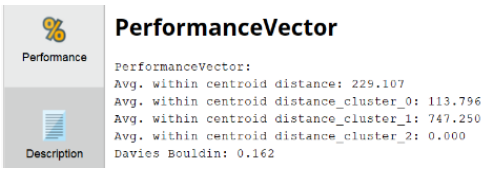

Gambar 14. Visualisasi Frekuensi Peminjaman k=3

Hasil dari perhitungan jarak menggunakan *Euclidean Distance* dengan *Manhattan Distance* menghasilkan nilai yang sama ketika sudah di hitung dengan metrik evaluasi yaitu *Davies Bouldin Index*, sehingga menghasilkan nilai *Davies Bouldin Index* pada k=2 adalah 0,055.




Gambar 15. Metrik Evaluasi Dengan Nilai k=3

Nilai *Davies Bouldin Index* pada k=3 adalah 0,162. Nilai *Davies Bouldin Index* yang mendekati nilai 0 itu yang terbaik, oleh karena itu nilai kelompok k=2 yang menghasilkan nilai *Davies Bouldin Index* terbaik.

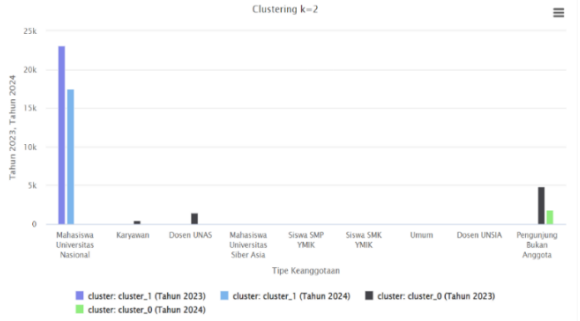



Gambar 16. Metrik Evaluasi Dengan Nilai k=3

### 4.9. Tipe Keanggotaan

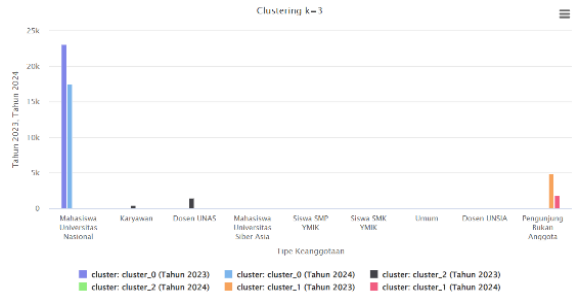
Hasil pengelompokkan dengan nilai k=2 menggunakan aplikasi RapidMiner menghasilkan pengelompokkan *cluster* 0 tipe keanggotaan yang jarang berkunjung yaitu karyawan, dosen UNAS, mahasiswa siber asia, siswa SMP YMIK, siswa SMK YMIK, umum, dosen UNSIA, pengunjung bukan anggota, sementara *cluster* 1 tipe keanggotaan yang

sering berkunjung adalah mahasiswa Universitas Nasional.



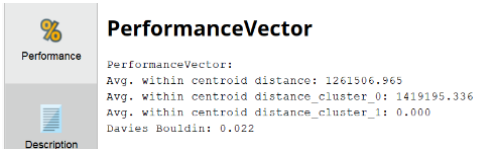

Gambar 17. Visualisasi Clustering Tipe Keanggotaan k=2

Hasil pengelompokkan dengan nilai k=3 menggunakan aplikasi RapidMiner menghasilkan pengelompokkan *cluster* 0 tipe keanggotaan yang sering berkunjung ke perpustakaan yaitu Mahasiswa Universitas Nasional, *cluster* 1 tipe keanggotaan yang sesekali berkunjung ke perpustakaan yaitu Pengunjung Bukan Anggota, *cluster* 2 tipe keanggotaan yang jarang berkunjung ke perpustakaan yaitu Karyawan, Dosen UNAS, Mahasiswa Universitas Siber Asia, Siswa SMP YMIK, Siswa SMK YMIK, Umum, Dosen UNSIA.



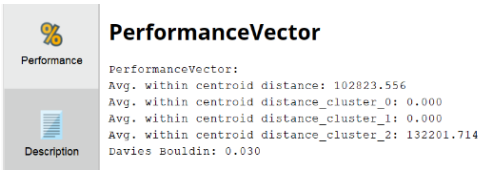

Gambar 18. Visualisasi Clustering Tipe Keanggotaan k=3

Hasil dari perhitungan jarak menggunakan *Euclidean Distance* dengan *Manhattan Distance* menghasilkan nilai yang sama, sehingga menghasilkan nilai *Davies Bouldin Index* pada k=2 adalah 0,022.




Gambar 19. Metrik Evaluasi Dengan Nilai k=2

Nilai *Davies Bouldin Index* pada k=3 adalah 0,030. Nilai *Davies Bouldin Index* yang mendekati nilai 0 itu yang terbaik, oleh karena itu nilai kelompok k=2 yang menghasilkan nilai *Davies Bouldin Index* terbaik.




Gambar 20. Metrik Evaluasi Dengan Nilai k=3

Berdasarkan hasil penelitian nilai k yang paling optimal yaitu k=2 di setiap variabel data, hasil nilai metrik evaluasi menggunakan *Davies Bouldin Index* pada perhitungan jarak *Euclidean Distance* dan *Manhattan Distance* menghasilkan nilai yang sama.

## 5. KESIMPULAN DAN SARAN

Kesimpulan pada penelitian ini adalah: (1) kategori buku yang tergolong dalam *cluster* 1 (sedikit di pinjam) yaitu 000, 100, 200, 400, 500, 600, 700, 800, dan 900, dan *cluster* 2 (banyak di pinjam) yaitu 300; (2) frekuensi durasi peminjaman pada *cluster* 1 (jarang muncul) yaitu 0, 1, 3, 4, 8, 9, 12, 14, dan 17 hari, dan pada *cluster* 2 (sering muncul) yaitu 7 hari; (3) frekuensi peminjaman yang tergolong dalam *cluster* 1 (sedikit di pinjam) yaitu komputer dan informasi umum, filsafat dan psikologi, agama, bahasa, sains dan matematika, teknologi dan ilmu terapan, seni dan hiburan, sastra, serta sejarah dan geografi, sementara *cluster* 2 (sering di pinjam) yaitu ilmu sosial; (4) Tipe keanggotaan yang tergolong dalam *cluster* 1 (jarang berkunjung) yaitu karyawan, dosen UNAS, mahasiswa Siber Asia, siswa SMP YMIK, siswa SMK YMIK, umum, dosen UNSIA, dan pengunjung bukan anggota, sementara *cluster* 2 (sering berkunjung) yaitu mahasiswa Universitas Nasional. Peneliti yang akan datang diharapkan dapat melakukan analisis tambahan terkait dengan frekuensi peminjaman dengan mengelompokkan judul buku yang banyak digemari. Selain hal tersebut, peneliti selanjutnya diharapkan dapat menggunakan pengukuran jarak ke centroid yang lain, seperti metode jarak *Minkowski* atau jarak *Mahalanobis* agar hal tersebut dapat memberikan perspektif yang baru bagi bidang penelitian terkait.

## DAFTAR PUSTAKA

[1] N. Q. A. Muhammad Ichsan Dedi Aprivian, "Prediksi Persediaan Buku Berdasarkan Pola Peminjaman Mahasiswa di Perpustakaan Universitas Ivet Menggunakan Metode K-Means Clustering," *J. Syst. Inf. Technol. Electron. Eng.*, vol. 3, no. 1, pp. 6–11, 2023.

[2] N. Karmila Sari and Y. Hendriyani, "Clustering Data Pengunjung UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang Menggunakan Algoritma K-Means," *J. Pendidik. Tambusai*, vol. 7, no. 3, pp. 29913–29923, 2023, doi: 10.31004/jptam.v7i3.11826.

[3] Mohamad Fahmi Hafidz and Sri Lestari, "Solution to Scalability and Sparsity Problems in Collaborative Filtering using K-Means Clustering and Weight Point Rank (WP-Rank),"

*J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 4, pp. 743–750, 2023, doi: 10.29207/resti.v7i4.4543.

[4] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, "Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 615–620, 2024, doi: 10.36040/jati.v8i1.8360.

[5] T. Maulana, R. Astuti, and Y. Arie Wijaya, "Implementasi Algoritma K-Means Dengan Optimize Parameter Grid Pada Data Kecelakaan Lalu Lintas Di Kota Cirebon," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 310–317, 2024, doi: 10.36040/jati.v8i1.8430.

[6] F. Handayani, "Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.6733.

[7] I. Wahyudi, L. Sarifah, and M. Sukron, "K-Means Clustering dengan Optimasi Algoritma Genetika untuk mengelompokkan daerah budidaya Cabai Jawa," vol. 6, no. 2, pp. 549–556, 2024, doi: 10.33650/jeecom.v4i2.

[8] R. Alya Shafira, Yahfizham, and A. Muliani Harahap, "Menentukan Jarak Terpendek Dalam Pengiriman BarangDengan Perbandingan Euclidean Distance DanManhattan Distance," *J. Sci. Soc. Res.*, vol. VI, no. 3, pp. 678–685, 2023.

[9] M. S. Pangestu and M. A. Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means," *Sainteks*, vol. 19, no. 2, p. 141, 2022, doi: 10.30595/sainteks.v19i2.14495.

[10] W. Sudrajat, I. Cholid, M. Rachmadi, and ..., "Penerapan Algoritma K-Means Dalam Segmentasi Pelanggan Pada Toko Sembako Menggunakan Rapidminer," *J. Ilm. Bin. ...*, vol. 3, no. 2, pp. 54–60, 2021, [Online]. Available: http://e-journal.stmik-bnj.ac.id/index.php/jb/article/view/57%0Ahttp://e-journal.stmik-bnj.ac.id/index.php/jb/article/download/57/60

[11] Y. Imam T. Umagapi, Basirung Umaternate, Hazriani, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa," *Semin. Nas. Sist. Inf. dan Teknol.*, vol. 7, no. 1, pp. 303–308, 2023.

[12] S. Fayyad, Piatetsky-Shapiro, "Overview of the KDD Process." Accessed: Nov. 19, 2024. [Online]. Available: https://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html