

Penerapan Clustering K-Means Untuk Mendukung Pengelolaan Koleksi Pada Perpustakaan Fakultas Teknik Universitas Negeri Gorontalo

¹Indhitya R. Padiku, ²Agus Lahinta

¹² Program Studi Sistem Informasi, Universitas Negeri Gorontalo, jalan jenderal Sudirman no.6
Provinsi Gorontalo, Indonesia, 96128
e-mail: 1indypadiku@ung.ac.id , 2lahinta@gmail.com

Abstrak

Salah satu unit fasilitas yang berperan sebagai penunjang pembelajaran mahasiswa dalam suatu universitas adalah perpustakaan. Perpustakaan merupakan tempat dimana terdapat beranekaragaman koleksi pustaka yang dapat memperkaya pengetahuan pengunjungnya. Sehingga, pengelolaan penambahan koleksi menjadi hal yang penting dalam menunjang hal tersebut. Saat ini penambahan koleksi buku di Perpustakaan Fakultas Teknik UNG hanya berdasarkan acuan dari data statistik buku terlaris, rekomendasi dosen maupun buku terkini tanpa mempertimbangkan koleksi mana yang diminati oleh mahasiswa. Hal tersebut dapat berdampak negatif jika terdapat penambahan koleksi pada kategori atau jenis buku yang kurang diminati untuk dibaca yang dapat menimbulkan masalah dari sisi pelayanan dan kerugian finansial perpustakaan. Beranjak dari masalah tersebut maka akan diterapkan metode clustering K-Means untuk mengelompokan minat baca pengunjung perpustakaan dalam hal ini adalah mahasiswa. Adapun variabel penunjang yang akan digunakan dalam pengelompokan tersebut akan dibagi sesuai dengan asal jurusan mahasiswa dengan kategori buku, stok, jumlah buku yang dipinjam dan jumlah buku yang dibaca. Pengelompokan ini diharapkan dapat dijadikan rekomendasi untuk pengelola perpustakaan dalam pengelolaan penambahan koleksi buku di perpustakaan Fakultas Teknik UNG.

Kata kunci: *Clustering; Clustering K-Means; koleksi; Perpustakaan*

Abstract

One of the facility units that acts as a supporter of student learning in a university is the library. The library is a place where there are various collections of libraries that can enrich the knowledge of visitors. Thus, the management of additional collections is important in supporting this. Currently, the addition of book collections in the UNG Engineering Faculty Library is only based on references from statistical data on best-selling books, lecturer recommendations and the latest books without considering which collections are of interest to students. This can have a negative impact if there are additional collections in categories or types of books that are less desirable to read which can cause problems from the service side and financial losses of the library. Moving on from this problem, the K-Means clustering method will be applied to classify the reading interest of library visitors in this case are students. The supporting variables that will be used in the grouping will be divided according to the origin of the student majors by category of books, stock, number of books borrowed and number of books read. This grouping is expected to be a recommendation for library managers in managing the addition of book collections in the UNG Faculty of Engineering library.

Keywords: *Clustering; Clustering K-Means; collection; library*

Diterima: 30 Mei 2021
Disetujui: 31 Mei 2022
Dipublikasi: 29 Juni 2022

©2022 Indhitya R. Padiku, Agus Lahinta
Under the license CC BY-SA 4.0

Pendahuluan

Dalam rangka menunjang proses belajar mengajar pada suatu universitas maka diperlukan fasilitas – fasilitas penunjang dimana salah satu diantaranya adalah perpustakaan. Perpustakaan merupakan tempat dimana terdapat beranekaragaman koleksi pustaka yang dapat memperkaya pengetahuan pengunjungnya.

Menurut Larasati dalam Azis dkk, perpustakaan adalah suatu unit kerja yang dijadikan sebagai tempat menyimpan koleksi bahan Pustaka yang diatur dengan cara tertentu secara sistematis agar dapat dipergunakan sebagai sumber informasi secara berkesinambungan bagi pengunjungnya (Azis dkk, 2016). Sehingga, untuk memelihara kesinambungan dari pelayanan perpustakaan kepada pengunjung maka perlu adanya pengelolaan terhadap koleksi buku yang ada dimana koleksi buku ini menjadi hal yang utama dalam suatu perpustakaan. Salah satu pengelolaan yang menjadi hal utama adalah pengelolaan terkait jumlah koleksi buku.

Dalam Jurnal oleh Deka dkk yang berjudul Klasterisasi Judul Buku dengan menggunakan Metode K-Means Jurnal ini membahas tentang penggunaan metode data mining dengan Teknik clustering dalam membantu proses pengelolaan letak buku di perpustakaan pusat Universitas Islam Indonesia. Hasil dari penelitian tersebut adalah terbentuk cluster yang di dalamnya berisi buku yang berada dalam kategori dimana terdapat koleksi buku dengan judul yang serupa dan hal tersebut dapat membantu pustakawan dalam mengelola peletakan buku (Deka,2014). Maka penggunaan metode k-means ini diterapkan juga pada penelitian kali ini untuk pengeleloan terkait jumlah koleksi buku karena hal tersebut menjadi hal utama dalam peningkatan kualitas pelayanan perpustakaan.

Saat ini di Perpustakaan Fakultas Teknik UNG terdapat 1202 eksemplar dimana jumlah tersebut terdiri dari 7 kategori yaitu: Buku Umum, Teknik Industri, Teknik Arsitektur, Teknik Informatika, Teknik Kriya, Teknik Elektro, dan Teknik Sipil. Penambahan koleksi buku dilakukan berdasarkan acuan buku mana yang terlaris dipasaran, rekomendasi dosen maupun buku terkini terkait dengan tren yang ada. Penambahan koleksi belum dilakukan dari hasil pertimbangan koleksi mana yang diminati oleh mahasiswa untuk dibaca atau dipinjam. Hal tersebut dapat berdampak negatif jika terdapat penambahan koleksi pada kategori atau jenis buku yang kurang diminati untuk dibaca yang dapat menimbulkan masalah dari sisi pelayanan dan kerugian finansial perpustakaan.

Dalam penelusuran minat baca mahasiswa dapat dilakukan dengan berbagai macam metode, tetapi dalam penelitian ini akan diusulkan salah satu metode data mining yaitu *clustering* dengan model *K-Means*. *Clustering K-Means* merupakan metode

data mining yang cocok dalam pengelompokan objek data yang memiliki kesamaan identitas. Menurut Xie dkk, *Clustering K-Means* bekerja dengan cara membagi objek data ke dalam beberapa kelompok/*cluster* yang berbeda sesuai dengan ukuran kesamaan dari data - data tersebut, sehingga untuk objek data yang berada dalam cluster mempunyai tingkat kesamaan terbesar sedangkan untuk objek data antar cluster yang berbeda mempunyai tingkat kesamaan terkecil (Xie dkk, 2020). Hasil dari metode ini diharapkan dapat menjadi bahan pertimbangan yang lebih akurat dibandingkan dengan penggunaan metode yang bersifat deskriptif.

Berdasarkan hal tersebut maka akan diterapkan metode clustering K-Means untuk mendukung pengelolaan penambahan koleksi dengan cara mengelompokan minat baca pengunjung perpustakaan dalam hal ini pengunjungnya difokuskan pada mahasiswa. Menurut Pulukadang, dkk (2015), Clustering K-Means dianggap sebagai metode yang efektif dalam mengelompokan data. Penerapan metode clustering K-means ini akan diterapkan spesifik untuk mengetahui minat baca dari mahasiswa. Adapun variabel penunjang yang akan digunakan dalam pengelompokan tersebut akan dibagi sesuai dengan asal jurusan mahasiswa dengan kategori buku, stok, jumlah buku yang dipinjam dan jumlah buku yang dibaca. Tujuan dari penelitian ini diharapkan dapat dijadikan rekomendasi untuk pengelola perpustakaan dalam pengelolaan koleksi buku di perpustakaan Fakultas Teknik UNG.

Metode

Menurut kategori pengelompokan, *clustering* terbagi menjadi dua yaitu utuh dan parsial. Jika semua data bisa bergabung menjadi satu (dalam konsep *partitioning*), bisa dikatakan semua data kompak menjadi satu kelompok. Namun jika ada satu atau beberapa data yang tidak ikut bergabung dalam kelompok mayoritas, data tersebut dikatakan memiliki perilaku menyimpang atau dikenal dengan istilah *outlier/noise/"uninterested background"*. Beberapa metode yang dapat mendeteksi *outlier* ini diantaranya adalah DBSCAN dan *K-Means* (dengan sejumlah komputasi tambahan).

Pada penelitian akan menerapkan metode clustering k-means sebagai metode analisis data. K-Means termasuk dalam *partitioning clustering* yaitu setiap data harus masuk dalam cluster tertentu dan memungkinkan bagi setiap data yang termasuk dalam cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain. K-Means memisahkan data ke k daerah bagian yang terpisah, dimana k adalah bilangan integer positif. **K** yang dimaksudkan dalam K-Means adalah jumlah partisi, sehingga algoritma ini mengelompokkan setiap titik pada data X dalam salah satu

partisi K. Semakin besar nilai K berdampak pada banyaknya cluster untuk mengelompokkan data X. Nilai K ini juga menjadi parameter yang dibutuhkan oleh algoritma K-Means. Tidak ada ketetapan mutlak bagaimana menentukan nilai K yang optimal. Biasanya, penentuan nilai K didasarkan atas informasi yang diketahui sebelumnya tentang seberapa banyak cluster data yang muncul pada data X. Algoritma K-Means sangat terkenal karena kemudahan dan kemampuannya untuk mengklaster data besar dan outlier dengan sangat cepat. Secara sederhana algoritma K-Means adalah sebagai berikut :

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

Tahapan pertama dalam proses ini dilakukan melalui pengumpulan data bersama dengan Operator Perpustakaan Fakultas Teknik UNG. Hasil dari proses tersebut didapatkan data terkait buku apa saja yang dibaca dalam periode tahun 2018-2021. Adapun kriteria buku yang dibaca diperoleh dari data buku apa saja yang dipinjam dengan tidak memasukkan data terkait peminjaman laporan kerja praktek dan skripsi/tugas akhir.

Secara garis besar Desain Penelitian ini dapat dijelaskan sebagai berikut :

1. Mengidentifikasi masalah;

Pada tahap ini akan dilakukan identifikasi masalah apa saja yang berhubungan dengan penambahan koleksi pada perpustakaan fakultas Teknik UNG dan faktor serta sub-faktor yang mempengaruhinya. Identifikasi dilakukan dengan wawancara. Adapun indikator keberhasilan pada tahap ini adalah permasalahan telah teridentifikasi secara jelas dengan dilengkapi ketersediaan data dari masing-masing faktor/variable yang mempengaruhi masalah.

2. Implementasi metode *Clustering K-Means*;

Menganalisis data yang sudah ada dengan menerapkan metode *clustering k-means*, dan mengeksekusi metodenya menggunakan aplikasi *data mining*. Keluaran dari tahapan ini adalah beberapa *cluster* yang terbentuk dari masing-masing objek data yang mempunyai bobot kesamaan. Indikator keberhasilan tahapan ini adalah setiap objek data akan tergabung dalam *cluster*/kelompok dimana masing-masing kelompok

dapat menjadi acuan dalam mengatur penambahan koleksi di perpustakaan fakultas Teknik UNG.

Hasil dan Pembahasan

Hasil

Ada sebanyak 101 buku yang akan dianalisis tingkat peminatannya berdasarkan jumlah peminjaman pada masing-masing judul bukunya, Data tersebut akan menjadi sumber data yang selanjutnya akan diterapkan metode clustering K-Means untuk pengelompokan minat. Pengelompokan akan dibagi dalam 3 kategori, dimana dalam metode clustering K-Means berarti akan menghasilkan 3 cluster dengan nilai $k=3$. Adapun penerapan lebih jelas berdasarkan tahapan clustering K-Means adalah sebagai berikut :

1. Menentukan jumlah cluster (k)

Nilai K untuk penelitian ini akan ditentukan sebanyak 3 kelompok. Dimana ketiga kelompok tersebut akan dipetakan dengan nama Cluster A, Cluster B dan Cluster C.

2. Menentukan centroid m (pusat cluster awal) secara acak sebanyak jumlah nilai K yang sudah ditentukan pada tahap pertama. Penentuan centroid awal ditentukan secara acak dari nilai frekuensi pada data peminjaman, Adapun nilainya adalah sebagai berikut :

- $m_A = \{3, 4, 0, 0\}$
- $m_B = \{1, 0, 0, 0\}$
- $m_C = \{2, 1, 0, 0\}$

3. Menentukan jarak pada masing-masing data ke masing – masing centroid menggunakan jarak Euclidean dan menentukan jarak yang bernilai minimum/terdekat dari centroid untuk menjadi anggota cluster yang pertama.

Pada tahap ini akan dilakukan perhitungan jarak untuk keseluruhan nilai frekuensi dari masing-masing buku yang jumlahnya 101 buku. Adapun contoh penerapan rumus pada tahapan ini akan dilakukan pada data No. 1 dengan judul buku “Analisa desain & pemrograman berorientasi objek dengan UML dan VISUAL BASIC .NET”.

Tabel 1 Contoh Data untuk menentukan Jarak Euclidean

Judul Buku	2018	2019	2020	2021
Analisa desain & pemrograman berorientasi obyek dengan UML dan VISUAL BASIC. NET	0	1	0	0

Berikut perhitungan jarak *Euclidean* berdasarkan data Tabel 1 dengan *centroid* awal pada tahap pertama.

- Jarak ke m_A Cluster A (d_{mA}) = $\sqrt{(0 - 3)^2 + (1 - 4)^2 + (0 - 5)^2 + (0 - 0)^2} = 4.24$
- Jarak ke m_B Cluster B (d_{mB}) = $\sqrt{(0 - 1)^2 + (1 - 0)^2 + (0 - 0)^2 + (0 - 0)^2} = 1.41$
- Jarak ke m_C Cluster C (d_{mC}) = $\sqrt{(0 - 2)^2 + (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2} = 2$

Adapun hasil dari perhitungan jarak *Euclidean* dan penentuan anggota *cluster* berdasarkan jarak terdekat ke *centroid* terhadap keseluruhan data pada Tabel Master.

4. Menentukan nilai *centroid* baru berdasarkan perhitungan nilai rata-rata pada anggota masing-masing *cluster*. Berdasarkan dari informasi dari Tabel Master, maka akan dihitung rata-rata dari kelompok nilai frekuensi pada Tabel disesuaikan dengan *Clustemya*.

Berdasarkan data maka didapatkan nilai *centroid* yang baru yaitu sebagai berikut:

- $m_{A1} = \{3, 8, 0, 0\}$
- $m_{B1} = \{0.39, 0.82, 0.20, 0.04\}$
- $m_{C1} = \{2.86, 0.29, 0.14, 0\}$

Hasil *centroid* tersebut adalah hasil dari perhitungan rata-rata masing-masing *cluster* yang akan diterapkan pada data berikutnya.

5. Mengulangi tahap 3-5 sampai data *cluster* sudah tidak berubah.
Pada tahap ini akan mengulang tahap ketiga sampai kelima dan perhitungan akan berhenti jika anggota *cluster* sudah tidak berubah.
6. Menentukan nilai *centroid* baru berdasarkan perhitungan nilai rata-rata pada anggota masing-masing *cluster*.

Berdasarkan dari informasi dan data, maka akan dihitung rata-rata dari kelompok disesuaikan dengan *Clustemya*

Tabel 2 *Centroid* baru untuk masing – masing *Cluster*

<i>Cluster A</i>				
<i>Jumlah Anggota Cluster : 1</i>				
Rata – rata	3.00	8.00	0.00	0.00
<i>Cluster B</i>				
<i>Jumlah Anggota Cluster : 91</i>				
Rata – rata	0.34	0.86	0.22	0.04
<i>Cluster C</i>				
<i>Jumlah Anggota Cluster : 9</i>				
Rata – rata	2.78	0.00	0.00	0.00

Berdasarkan data pada Tabel 2 di atas maka didapatkan nilai centroid yang baru yaitu sebagai berikut :

- $m_{A2} = \{3, 8, 0, 0\}$
- $m_{B2} = \{0.34, 0.86, 0.22, 0.04\}$
- $m_{C2} = \{2.78, 0, 0, 0\}$

6. Menentukan nilai *centroid* baru berdasarkan perhitungan nilai rata – rata pada anggota masing-masing *cluster*

Tabel 3 *Centroid* baru untuk masing – masing *Cluster*

<i>Cluster A</i>				
<i>Jumlah Anggota Cluster : 1</i>				
Rata – rata	3.00	8.00	0.00	0.00
<i>Cluster B</i>				
<i>Jumlah Anggota Cluster : 91</i>				
Rata – rata	0.34	0.86	0.22	0.04
<i>Cluster C</i>				
<i>Jumlah Anggota Cluster : 9</i>				
Rata – rata	2.78	0.00	0.00	0.00

Berdasarkan data pada Tabel di atas maka didapatkan nilai centroid yang baru yaitu sebagai berikut :

- $m_{A3} = \{3, 8, 0, 0\}$
- $m_{B3} = \{0.34, 0.86, 0.22, 0.04\}$
- $m_{C3} = \{2.78, 0, 0, 0\}$

Pembahasan

Hasil analisis dari metode clustering k-means yang telah dilakukan akan dievaluasi menggunakan Davies Bouldin Index (DBI). Menurut Widiarina dan Wahono (Widiarina dan Wahono, 2015) semakin kecil nilai DBI (≥ 0) maka cluster yang terbentuk adalah yang lebih baik. Adapun hasil dari penerapan persamaan DBI untuk evaluasi hasil cluster adalah sebagai berikut :

a. *Sum of Square within Cluster (SSW) & Sum of Square between Cluster (SSB)*

$$SSW_1 = 0 ; SSB_{12} = 7.62$$

$$SSW_2 = 0.8 ; SSB_{13} = 8$$

$$SSW_3 = 0.5 ; SSB_{23} = 2.59$$

b. Rasio (R)

$$R_{12} = 0.10$$

$$R_{13} = 0.06$$

$$R_{23} = 0.52$$

c. DBI

Dari hasil rasio di atas maka dapat dipetakan dalam Tabel sebagai berikut:

Tabel 4 Pemetaan Rasio DBI

R	1	2	3	Nilai Max
1	0.00	0.11	0.06	0.11
2	0.11	0.00	0.52	0.52
3	0.06	0.52	0.00	0.52

Berdasarkan Tabel 4 maka Nilai DBI adalah rata-rata dari Nilai Max dari R_1 , R_2 , R_3 yaitu 0,38. Dimana nilai tersebut mendekati nilai 0, sehingga dapat dikatakan hasil evaluasi penggunaan metode *clustering kmeans* dari data frekuensi peminjaman buku dari tahun 2018-2021 pada perpustakaan Fakultas Teknik Universitas Negeri Gorontalo adalah cukup baik.

Berdasarkan analisis yang sudah dilakukan terhadap 101 jumlah buku yang dipinjam dari tahun 2018-2020 maka didapatkan kesimpulan sebagai berikut :

- a. Jumlah anggota pada cluster A adalah sebanyak 1 buku, dimana jika dilihat berdasarkan minat dari jumlah mahasiswa pada jumlah frekuensi peminjaman buku maka cluster A merupakan kelompok buku yang paling sedikit diminati oleh mahasiswa yang meminjam buku di perpustakaan fakultas Teknik Univeristas Negeri Gorontalo.
- b. Jumlah anggota pada cluster B adalah sebanyak 91 buku, dimana jika dilihat berdasarkan minat dari jumlah mahasiswa pada jumlah frekuensi peminjaman buku maka cluster B merupakan kelompok buku yang paling banyak diminati oleh mahasiswa yang meminjam buku di perpustakaan fakultas Teknik Univeristas Negeri Gorontalo.
- c. Jumlah anggota pada cluster C adalah sebanyak 9 buku, dimana jika dilihat berdasarkan minat dari jumlah mahasiswa pada jumlah frekuensi peminjaman buku maka cluster C merupakan kelompok buku yang rata - rata diminati oleh mahasiswa yang meminjam buku di perpustakaan fakultas Teknik Univeristas Negeri Gorontalo.

Adapun hasil analisis yang didapat dalam penelitian ini membuktikan bahwa penggunaan metode *clustering k-means* dapat digunakan dalam pengelolaan buku pada perpustakaan serta dapat mendukung keputusan terkait pelayanan pada perpustakaan.

Kesimpulan

Penelitian ini bertujuan untuk membantu dalam proses pengambilan keputusan untuk mendukung pengelolaan penambahan koleksi dengan cara mengelompokkan minat baca pengunjung perpustakaan. Metode clustering K-Means digunakan untuk mengelompokkan frekuensi peminjaman buku yang paling banyak diminati atau yang paling sedikit diminati oleh peminjam khususnya mahasiswa di perpustakaan Fakultas Teknik Universitas Negeri Gorontalo. Hasil penelitian menunjukkan bahwa metode yang diusulkan dapat digunakan untuk dijadikan rekomendasi untuk pengelola perpustakaan dalam pengelolaan koleksi buku di perpustakaan Fakultas Teknik UNG.

Daftar Pustaka

- Azis, W. S., dan Atmajaya, D., 2016, Pengelompokan Minat Baca Mahasiswa menggunakan Metode K-Means, *Jurnal Ilmiah ILKOM*, Vol. 8, No. 2, 89-94.
- Can, F., E. A., 1990. Concepts and effectiveness of the cover-coefficient-based clustering methodology for text databases. *ACM Transactions on Database Systems* 15 (4), 483 - 517.
- Kumar, V., Steinbach, M., Tan, P., 2004. *Introduction to Data Mining*. Addison-Wesley
- Larose, D.T., 2005. *Discovering Knowledge in Data : An Introduction to Data Mining*. WileyInterscience, Canada.
- Nugraha, D., D., C., Fahmi, M., Naimah, Z., dan Setiani, N., 2014, *Seminar Nasional Aplikasi Teknologi Informasi, G1-G4*.
- Pulukadang, D. R., Mustafid, dan Farikhin, 2015, Pendekatan Clustering untuk Ekstraksi Pengetahuan pada Pembangunan Sistem Manajemen Pengetahuan, *Jurnal Sistem Informasi Bisnis*, Vol. 5, No. 2, 79-83.
- Wibisono, Y., 2011. Perbandingan Partition Around Medoids (PAM) dan K-means Clustering untuk Tweets, *Prosiding Konferensi Nasional Sistem Informasi*. Medan, Februari 25-26, 483 – 486.
- Xie, T., Liu, R., dan Wei, Z., 2020, Improvement of the Fast Clustering Algorithm Improved by K-Means in the Big Data, *Applied Mathematics and Nonlinear Sciences*, Vol.5, Issue 1, 1-10.