

**TEKNIK DATA MINING MENGGUNAKAN METODE K-MEANS UNTUK  
MENGCLUSTER DAN PENCARIAN BUKU DI PERPUSTAKAAN  
DAERAH KABUPATEN ASAHAN**

**Sela Gustin<sup>1</sup>, William Ramdhan<sup>2\*</sup>, Wan Mariatul Kifti<sup>3</sup>**

<sup>1,2,3</sup>Sistem Informasi, STMIK Royal Kisaran

*\*email*: : [william\\_ramdhan052@gmail.com](mailto:william_ramdhan052@gmail.com)

**Abstract:** The Asahan Regency Regional Library and Archives is one of the government agencies that has used a computer-based information system but only on the visitor data system. In the book search process, most visitors find it difficult to find the right reference book according to their needs. Therefore we need a system that can simplify and speed up the search for the book. One of them is by grouping books so that the search process will be faster, easier, and more accurate. By designing an information system that can assist the library in managing library materials in the form of grouping books using the K-Means method to facilitate the arrangement of books into shelves, making it easier for visitors to find library materials quickly and easily and further improve the quality of library services. This application was built using the PHP programming language, MySQL database, and using the K-Means method. By implementing the K-Means method, a clustering and search application is produced in the Regional Library and Archives of Asahan Regency.

**Keywords:** Libraries, clusters, K-means, PHP and MySQL

**Abstrak:** sudah memakai sistem informasi berbasis komputer tetapi hanya pada bagian sistem data pengunjung. proses pencarian buku, kebanyakan pengunjung sulit mendapatkan buku referensi yang tepat sesuai dengan kebutuhan mereka. Maka dari itu diperlukan suatu sistem yang dapat mempermudah dan mempercepat pencarian buku tersebut. Salah satunya dengan pengelompokkan buku sehingga proses pencariannya akan lebih cepat, mudah, dan akurat. Dengan membuat perancangan sistem informasi yang dapat membantu pihak perpustakaan dalam mengelola bahan pustaka dalam bentuk pengelompokkan buku menggunakan metode K-Means untuk memudahkan penataan buku ke dalam rak sehingga memudahkan pengunjung dalam pencarian bahan pustaka dengan cepat dan mudah serta lebih meningkatkan mutu pelayanan perpustakaan. Aplikasi ini dibangun menggunakan bahasa pemrograman PHP, database MySQL, dan menggunakan metode K-Means. Dengan mengimplementasikan metode K-Means, maka dihasilkan sebuah aplikasi pengclusteran dan pencarian di Perpustakaan Dan Kearsipan Daerah Kabupaten Asahan.

**Kata Kunci:** Perpustakaan; cluster; K-means; PHP dan MySQL.

## PENDAHULUAN

Perpustakaan Dan Kearsipan Daerah Kabupaten Asahan adalah salah satu instansi pemerintah yang sudah memakai sistem informasi berbasis komputer tetapi hanya pada bagian sistem data pengunjung. Dalam proses pencarian buku, kebanyakan pengunjung tidak mengetahui letak buku yang sesuai dengan keinginan pengunjung tersebut. Oleh sebab itu diperlukan suatu sistem yang dapat mempermudah dan mempercepat pencarian buku tersebut. Salah satunya dengan mengelompokkan buku sehingga proses pencariannya akan lebih cepat, mudah, dan akurat.

Perpustakaan yang menggunakan sistem pelayanan dan pengelolaan bahan pustaka secara manual dapat menimbulkan masalah tidak efisien dalam pencatatan dan pencarian bahan pustaka yang menjadi koleksi perpustakaan. Akibatnya banyak penyimpanan data yang tidak teratur dan terdapat beberapa data yang sama (redudansi). Sistem pencarian buku yang masih dilakukan secara manual dianggap tidak efisien lagi mengingat pesatnya peningkatan jumlah buku. Oleh karena itu, untuk mengatasi permasalahan tersebut diperlukan suatu pendukung keputusan untuk membantu pihak perpustakaan dalam memberikan pelayanan dan pengelolaan bahan pustaka yaitu dengan cara mengelompokkan bahan pustaka yang ada dengan menggunakan metode *K-Means Cluster* yang akan diterapkan dalam membangun perangkat lunak pencarian pada Aplikasi Perpustakaan Daerah Kabupaten Asahan.

*K-Means* merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada kedalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*.

Tujuan penelitian ini adalah membuat perancangan sistem informasi yang dapat membantu pihak perpustakaan dalam mengelola bahan pustaka dalam bentuk pengelompokkan buku menggunakan metode *K-Means* untuk memudahkan penataan buku ke dalam rak sehingga memudahkan pengunjung dalam pencarian bahan pustaka dengan cepat dan mudah serta lebih meningkatkan mutu pelayanan perpustakaan. Berdasarkan latar belakang masalah, penulis berniat untuk merancang dan membuat suatu sistem komputerisasi yang dapat meng*cluster* buku serta membuat suatu sistem pencarian buku dari hasil *cluster* buku tersebut.

### Definisi Data Mining

*Data mining* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *database sistem*, *data warehousing*, *statistik*, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing*. *Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti

dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar [1].

### **Tahapan Data Mining**

Untuk menganalisis data dalam penerapan *data mining* ini menggunakan proses tahapan *knowledge discovery in databases* (KDD) yang terdiri dari *data cleaning*, *data integration*, *Data Selection*, *Data Transformation*, Proses *mining*, *pattern evaluation*, *knowledge presentation* [2]. Tahap-tahap *data mining* ada 7 yaitu :

#### **Pembersihan data (*data cleaning*)**

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya [3].

#### **Integrasi data (*data integration*)**

Integrasi data merupakan penggabungan data dari berbagai *database* kedalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau *file* teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya [4].

#### **Seleksi Data (*data selection*)**

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja [5].

#### **Transformasi data (*data transformation*)**

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data [6]

**Proses *mining***

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data [7].

**Evaluasi pola (*pattern evaluation*)**

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat [8].

**Presentasi pengetahuan (*knowledge presentation*)**

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

***Clustering***

Klasterisasi atau *clustering* merupakan salah satu teknik dari salah satu fungsionalitas *data mining*, algoritma *clustering* merupakan algoritma pengelompokan sejumlah data menjadi kelompok-kelompok data tertentu

***K-Means***

Metode *K-Means* adalah Metode *clustering* berbasis jarak yang membagi data kedalam *cluster* dan algoritma ini bekerja pada atribut numerik. Metode *K-Means* termasuk dalam *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah. Metode *K-Means* berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain [9].

**METODE**

Berikut ini adalah langkah-langkah dalam algoritma *K-Means* : (1) menentukan jumlah cluster (k) pada data set; (2) tentukan nilai pusat (*centroid*); (3) penentuan nilai *centroid* pada tahap awal dilakukan secara random dengan rumus menentukan target awal *k-means*, rumus tersebut digunakan untuk mendapatkan target data atau jarak antara kelompok; (4) pada tahap iterasi digunakan rumus rata-rata, pada rumus tersebut dilakukan perhitungan untuk mencari nilai rata-rata seperti pada persamaan berikut ini :

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad \text{Persamaan (1)}$$

Keterangan :

$V_{ij}$  = *centroid* rata-rata *cluster* ke-i untuk variabel ke-j

- $N_i$  = jumlah anggota *cluster* ke- $i$   
 $i, k$  = indeks dari *cluster*  
 $j$  = indeks dari variabel  
 $X_{kj}$  = nilai data ke- $k$  variabel ke- $j$  dalam *cluster* tersebut

Pada masing-masing record, hitung jarak terdekat dengan *centroid* Jarak *centroid* yang digunakan adalah *Euclidean Distance* [10], dengan rumus seperti pada persamaan berikut :

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad \text{Persamaan (2)}$$

Keterangan :

- $De$  = *Euclidean Distance*  
 $i$  = banyaknya objek  
 $(x, y)$  = koordinat objek  
 $(s, t)$  = koordinat *centroid*

Kelompokkan objek berdasarkan jarak ke *centroid* terdekat Ulangi langkah ke-2, lakukan iterasi hingga *centroid* bernilai optimal

## METODE PENELITIAN

Dalam melakukan penelitian, peneliti harus memperhatikan dan mengetahui metode apa yang akan digunakan dalam penelitiannya. Dalam penelitian ini metode yang digunakan penulis adalah metode kualitatif. Penelitian kualitatif biasanya mengandalkan teknik observasi terlibat dan tak terkontrol, juga wawancara bebas dan mendalam. Jenis data yang dikumpulkan bersifat monografis atau berwujud kasus-kasus. Data-data tersebut dianalisis secara kualitatif [11]. Didalam proses penelitian ini, yang dimaksud dari metode kualitatif yaitu pada masalah yang sesuai dengan fenomena dan menggunakan data. Data tersebut diambil langsung dari Dinas Perpustakaan dan Kearsipan Daerah Kab. Asahan.

## HASIL DAN PEMBAHASAN

### Tampilan *Form* Menu Utama

*Form* menu utama berisi tampilan awal masuk ke dalam sistem. Berikut ini adalah tampilan *form* menu utama:



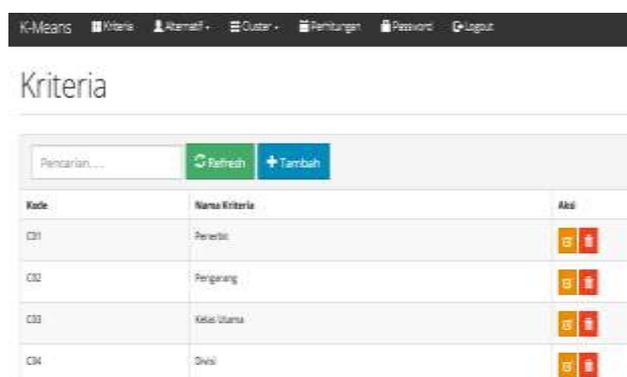
Gambar 1. Tampilan Form Login

### Tampilan Input dan Tampil Data Kriteria

Input data kriteria berfungsi sebagai salah satu implementasi metode *K-Means Clustering* untuk pengelompokan buku di Perpustakaan Daerah Kabupaten Asahan seperti tambah, simpan, edit, hapus, cari, refresh, dan keluar. Berikut adalah tampilan input kriteria:



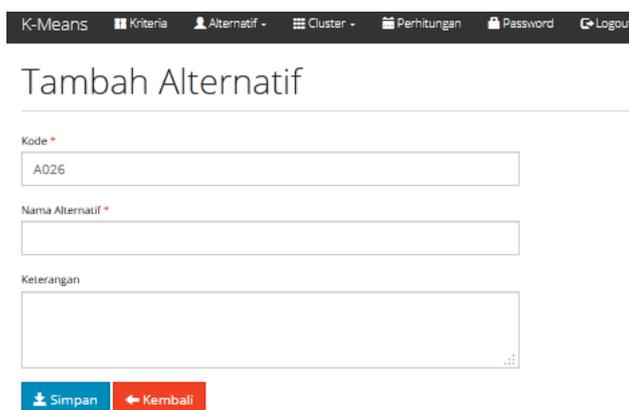
Gambar 2. Tampilan Input Data Kriteria



Gambar 3. Tampilan Tampil Data Kriteria

### Tampilan *Input* dan Tampil Data alternatif

*Input* data tes fisik berfungsi sebagai salah satu jenis implementasi metode *K-Means Clustering* untuk pengelompokkan buku di Perpustakaan Daerah Kabupaten Asahan seperti tambah, simpan, edit, hapus, cari, refresh, dan keluar. Berikut adalah tampilan *input* data alternatif:



Gambar 4. Tampilan *Input* Data alternative

### Tampilan Ubah dan Tampil Data Nilai Bobot Alternatif

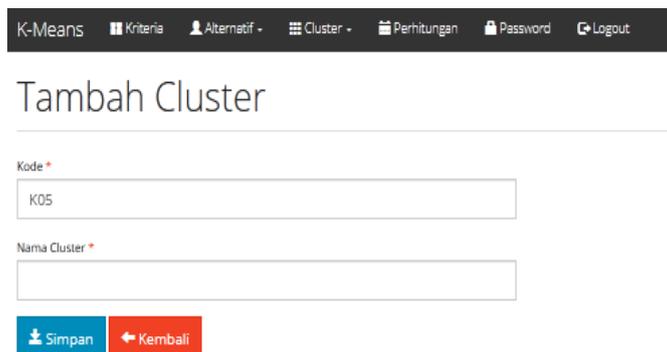
*Input* data tes fisik berfungsi sebagai salah satu jenis implementasi metode *K-Means Clustering* untuk pengelompokkan buku di Perpustakaan Daerah Kabupaten Asahan seperti simpan, edit, cari, refresh, dan keluar. Berikut adalah tampilan ubah data alternatif:



Gambar 5. Tampilan Ubah Data Nilai Bobot Alternatif

### Tampilan *Input* dan Tampil Data Cluster

*Input* data cluster berfungsi sebagai salah satu jenis implementasi metode *K-Means Clustering* untuk pengelompokkan buku di Perpustakaan Daerah Kabupaten Asahan seperti simpanproses, lihat dan keluar. Berikut adalah tampilan *input* data cluster.



Gambar 6. Tampilan *Input Data Cluster*

### Tampilan Ubah dan Tampil Data Nilai Pusat *Centroid*

Ubah data nilai pusat *centroid* berfungsi untuk mengubah nilai pusat centroid pada implementasi metode *K-Means Clustering* untuk pengelompokan buku di Perpustakaan Daerah Kabupaten Asahan seperti simpan proses, lihat dan keluar. Berikut adalah tampilan ubah data nilai pusat *centroid*



Gambar 7. Tampilan Ubah Data Nilai Pusat Centroid

### Tampilan *Input* dan Tampil Data Perhitungan

*Input* data perhitungan pada implementasi metode *K-Means Clustering* untuk pengelompokan buku di Perpustakaan Daerah Kabupaten Asahan seperti proses, lihat dan keluar. Berikut adalah tampilan *input* data itersi 2:



Gambar 8. Tampilan *Input* Data Perhitungan

## KESIMPULAN

Berdasarkan dari pembahasan, penulis dapat menarik beberapa kesimpulan yaitu dapat mengimplementasikan metode *K-Means Clustering* di Perpustakaan Daerah Kabupaten Asahan yang menghasilkan pengelompokan buku berdasarkan kepada *clusternya* masing-masing. Dapat mengelompokan buku berdasarkan pada *cluster* yang telah ditentukan sesuai dalam penerapan metode *K-Means Clustering* sehingga dapat mempermudah dalam pencarian buku.

## DAFTAR PUSTAKA

- [1] G. Abdillah *et al.*, “Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means,” *Sentika 2016*, vol. 2016, no. 2016, pp. 18–19, 2016.
- [2] F. Yunita, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru,” *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [3] L. Iryani, “Penerapan Datamining Menentukan Minat Baca Mahasiswa Di Perpustakaan Universitas Bina Darma Palembang Menggunakan Metode Clustering Application of Data Mining Determines Interest Reading in Students At Universitas Bina Darma Palembang Using Clustering Met,” *J. Inf. Technol. Comput. Sci.*, vol. 3, no. 1, 2020.
- [4] . F., F. T. Kesuma, and S. P. Tamba, “Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering,” *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.
- [5] A. Setiawan, I. F. Astuti, and A. H. Kridalaksana, “Klasifikasi Dan Pencarian Buku Referensi Akademik Menggunakan Metode Naïve Bayes Classifier (NBC) (Studi Kasus: Perpustakaan Daerah Provinsi Kalimantan Timur),” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 10, no. 1, p. 1, 2016, doi: 10.30872/jim.v10i1.17.
- [6] Y. Heriyanto, “Perancangan Sistem Informasi Rental Mobil Berbasis Web Pada PT.APM Rent Car,” *J. Intra-Tech*, vol. 2, no. 2, pp. 64–77, 2018.
- [7] J. Bri, R. Dalam, K. Baru, and J. Selatan, “Eriek Orlando Kata Kunci Tinjauan Pustaka Pengertian Aplikasi Konsep Dasar Web Jenis Aplikasi Web Web Berita dan Informasi,” vol. 16, 2017.
- [8] Fitri Ayu and Nia Permatasari, “perancangan sistem informasi pengolahan data PKL pada divisi humas PT pegadaian,” *J. Infra tech*, vol. 2, no. 2, pp. 12–26, 2018, [Online]. Available: <http://journal.amikmahaputra.ac.id/index.php/JIT/article/download/33/25>.
- [9] M. P. H. Setiawan and F. Masya, “Analisa Perancangan Sistem Informasi Pelaporan Fasilitas Umum dan Informasi Pembuatan E-KTP Untuk Masyarakat,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 5, no. 1, pp. 1–8, 2020, doi: 10.36341/rabit.v5i1.810.

- [10] P. Penggunaan *et al.*, “p 2337-4721,” vol. 5, no. 1, pp. 145–151, 2017.  
DAN JARINGAN PADA SISWA SMAN 1 AIR JOMAN,” *Jurdimas R.*, vol. 1, no. 2, pp. 81–86, 2018.
- [11] N. Manurung, “Sistem Pendukung Keputusan Pemberian Bonus Karyawan Menggunakan Metode AHP,” *JurTI (Jurnal Teknol. Informasi)*, vol. 1, no. 1, pp. 48–53, 2017.