

Pemanfaatan K Means Clustering dalam Pengelompokan Judul Skripsi

Nisar¹⁾, Wasilah²⁾Haris Kusumajaya³⁾

^{1,3}Departemen Informatika, IIB Darmajaya, Jalan Zainal Abidin Pagar Alam No.93, Gedung Meneng, Bandar Lampung

²Departemen Sistem Informasi, IIB Darmajaya, Jalan Zainal Abidin Pagar Alam No.93, Gedung Meneng, Bandar Lampung

e-mail: * nisar@darmajaya.ac.id , wasilah@darmajaya.ac.id.

Abstrak

Abstrak— Skripsi merupakan istilah yang digunakan di Indonesia untuk menggambarkan karya tulis ilmiah. Skripsi berupa penjelasan tertulis hasil penelitian sarjana yang membahas fenomena dalam bidang ilmu tertentu dengan menggunakan kaidah-kaidah yang berlaku. Dalam studi kasus ini, analisis data mining dilakukan dengan menggunakan metode clustering K Means. Kriteria yang digunakan untuk mengelompokkan judul skripsi dengan K-Means adalah rpm, nama, judul, dosen pembimbing. Semakin bertambahnya jumlah mahasiswa dan variasi judul skripsi menyebabkan kesulitan dalam pengelompokan data skripsi. Penelitian ini bertujuan untuk mengelompokkan data skripsi mahasiswa program studi teknik informatika IIB Darmajaya. Pengelompokan dilakukan dengan menggunakan algoritma K-Means Clustering. Proses perhitungan algoritma clustering K-Means menggunakan aplikasi sederhana yang dibangun dengan menggunakan bahasa pemrograman PHP dan database MySQL.

Kata kunci—pengelompokan algoritma K-Means, pengelompokan skripsi.

Abstract

Abstract— Thesis is a term used in Indonesia to describe scientific writing. Thesis is in the form of a written explanation of the results of undergraduate research that discusses phenomena in certain fields of science by using applicable rules. In this case study, data mining analysis was carried out using the K-Means clustering method. The criteria used to classify thesis titles with K-Means are rpm, name, title, supervisor. The increasing number of students and variations in thesis titles cause difficulties in grouping thesis data. This study is intended to classify the thesis data of students of the informatics engineering study program IIB Darmajaya. Grouping needs to be done using an algorithm K-Means Clustering. The process of calculating the K-Means clustering algorithm uses a simple application built using the PHP programming language and MySQL database.

Keywords—K-Means algorithm, clustering data, thesis grouping.

1. PENDAHULUAN

Skripsi adalah karya ilmiah yang ditulis oleh mahasiswa program sarjana yang membahas topik atau bidang tertentu berdasarkan hasil tinjauan pustaka yang ditulis oleh para ahli, hasil penelitian lapangan, atau hasil pengembangan. Peningkatan jumlah

mahasiswa dan variasi judul skripsi menyebabkan banyaknya data skripsi yang harus diolah. Hal ini menyebabkan kesulitan dalam pengelompokan data skripsi.

Penerapan data mining dapat membantu menganalisa data yang diperoleh dari kondisi mahasiswa. Teknik data mining yang digunakan adalah dengan menggunakan teknik clustering. Algoritma K-Means merupakan algoritma non hirarki yang berasal dari metode data clustering, Metode K-Means ini mempartisi data kedalam kelompok sehingga data yang berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain [1]. Adapun tujuan dari pengelompokan data ini adalah untuk meminimumkan fungsi objektif yang diset dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi dalam antar kelompok [2].

Penerapan Algoritma K-Means telah dilakukan di berbagai bidang. Algoritma Analisis clustering K-Means pada Penyakit Infeksi Manusia [3]. Penerapan Algoritma K-Means Untuk Pemetaan Garis Kemiskinan Provinsi di Indonesia berdasarkan kesamaan tingkat pendapatan/kapita/bulan [4] [5] [6]. Pada penelitian lain, Algoritma K-Means digunakan untuk menentukan kategori peminatan penjualan bahan bangunan [7]. Kajian lainnya menggunakan Metode K-Means Clustering untuk mengkategorikan beras impor beras [8] dan penerapan Algoritma K-Means untuk penentuan jurusan di SMA [9]

Penelitian ini bertujuan untuk mengelompokkan data skripsi mahasiswa program studi teknik informatika IIB Darmajaya menggunakan algoritma K-Means Clustering. Kriteria yang digunakan untuk mengelompokkan judul skripsi dengan K-Means adalah npm, nama, judul, dan dosen pembimbing. Dalam penelitian ini judul skripsi mahasiswa akan dikelompokkan menurut tahun dan program studi dengan menggunakan konsep analisis Data Mining yaitu metode K Means Clustering.

2. METODE PENELITIAN

Penerapan data mining dapat membantu menganalisis data yang diperoleh. Teknik data mining yang digunakan dalam penelitian ini adalah teknik clustering, yaitu algoritma K-Means. Algoritma K-Means merupakan algoritma non-hierarchical yang berasal dari metode data clustering, metode K-Means ini mempartisi data kedalam kelompok-kelompok sehingga data dengan karakteristik yang sama dimasukkan kedalam kelompok yang sama dan data dengan karakteristik yang berbeda dikelompokkan kedalam kelompok yang berbeda[2]. Tujuan dari pengelompokan data ini adalah untuk meminimalkan fungsi tujuan yang ditetapkan dalam proses pengelompokan. Pada pengelompokan dilakukan minimalisasi variasi pada suatu kelompok dan memaksimalkan variasi antar kelompok.

2.1. Tahapan Penelitian

Metode yang digunakan dalam proses pengelompokan skripsi ini memiliki beberapa tahapan. Tahapan tersebut dimulai dari data *preprocessing*, *cleaning*, dan *clustering*. Tahap *preprocessing* bertujuan untuk mendapatkan bagian data yang diinginkan. Sedangkan *cleaning* digunakan untuk memeriksa data yang konsisten dan kesalahan ketik dan mengisi nilai yang hilang. Tahap selanjutnya adalah *clustering*, pada tahap grouping ini dimulai dengan menentukan jumlah *cluster*, setelah itu ditentukan titik pusat *cluster* (*centroid*) dan menghitung jarak antara data dengan titik pusat *cluster* dengan persamaan jarak *Euclidean*. Kemudian data tersebut di alokasikan ke *cluster* terdekat sehingga diketahui data tersebut berada di kelompok mana. Tahap selanjutnya adalah pengambilan keputusan, dimana data yang telah dikelompokkan akan diranking. Dari data tersebut dibuat sebuah matriks keputusan berdasarkan kriteria dan alternatif yang dimiliki yang kemudian dinormalisasi. Selanjutnya menentukan solusi ideal positif dan negatif diikuti dengan menghitung ukuran utilitas. Berikut ini adalah langkah-langkah dari algoritma *k-means*:

a. Penentuan *cluster* awal

Dalam menentukan n pusat *cluster* awal, dibangkitkan bilangan acak yang mewakili urutan data masukan. Pusat awal *cluster* diperoleh dari data tersebut, bukan dengan menentukan titik baru.

b. Perhitungan jarak ke pusat *cluster*

Untuk mengukur jarak antara data dan pusat *cluster* digunakan jarak *Euclidian*. Adapun algoritma untuk menghitung jarak antara data dan pusat cluster adalah:

$$(x, j) = \sqrt{(x_i - c_j)^2} \dots (1)$$

Penjelasan :

xi : Data kriteria

i : Centroid di cluster j

c. Pengelompokan Data

Jarak dari hasil perhitungan akan sebanding. Selanjutnya dipilih jarak terdekat antara data dan pusat *cluster*. Jarak ini menunjukkan bahwa data berada dalam satu kelompok dengan pusat *cluster* terdekat. Cara mengelompokkan data tersebut adalah:

- Pilih nilai jarak setiap pusat cluster dengan data.
- Cari nilai jarak terkecil.
- Kelompokkan data pada pusat cluster yang memiliki jarak terkecil.

d. Penentuan pusat *cluster* baru

Pusat *cluster* baru dapat dihitung berdasarkan nilai rata-rata anggota *cluster* dan pusat *cluster*. Pusat *cluster* baru digunakan untuk melakukan iterasi berikutnya, jika hasil yang diperoleh belum konvergen. Proses iterasi akan berhenti jika telah memenuhi maksimal iterasi yang dimasukkan oleh pengguna atau hasil yang dicapai telah konvergen (pusat *cluster* baru sama dengan pusat *cluster* sebelumnya).

Algoritma penentuan pusat *cluster* adalah sebagai berikut:

- a. Tentukan jumlah anggota setiap cluster
- b. Hitung pusat baru dengan menggunakan rumus berikut ini:

$$c_j^{(t+1)} = \frac{1}{N_j} \sum_{i \in C_j} x_i \dots (2)$$

Penjelasan :

$c_j^{(t+1)}$: Centroid baru pada iterasi pertama

N_j : Banyak data di cluster saja

Hasil dari operasi *clustering* yang terbentuk selanjutnya akan dievaluasi menggunakan indeks *Davies Bouldin* yang dihitung dengan persamaan:

$$C_i = \text{Cluster } i \text{ dan } c_i \text{ adalah centroid dari cluster } i \dots (3)$$

2.2. Pengelompokan

Clustering adalah suatu pekerjaan mengelompokkan sekumpulan objek data sehingga objek-objek dalam suatu kelompok memiliki kemiripan yang tinggi, tetapi sangat berbeda dengan objek-objek dalam kelompok lain. Proses pada *clustering* akan mempartisi sekumpulan objek data menjadi subset yang dapat dimanfaatkan untuk mengatur hasil pencarian ke dalam kelompok-kelompok dan menyajikan hasilnya secara ringkas dan mudah diakses [1]. *Clustering* umumnya digunakan di berbagai bidang dengan berbagai aplikasi

yang sangat penting termasuk riset pasar, sistem rekomendasi, sistem keamanan, dan mesin pencari.

2.3. Metode Pengumpulan Data

Tahap analisis dilakukan untuk menentukan kebutuhan perangkat lunak yang akan dibangun. Hal ini terkait dengan penentuan perangkat keras, perangkat lunak, tampilan program dan bentuk yang akan digunakan dalam pembuatan prototipe. Metode pengumpulan data yang digunakan untuk mendapatkan data yang dibutuhkan dalam penelitian ini adalah:

- **Studi Literatur**
Studi literatur dilakukan dengan mempelajari buku-buku, literatur yang ada di perpustakaan dan jurnal mengenai penelitian sebelumnya yang berkaitan dengan sistem pakar dan khususnya yang menggunakan *clustering K-Means*.
- **Wawancara**
Melakukan wawancara dengan staff IIB Darmajaya di lokasi penelitian. Wawancara dilakukan dengan tujuan untuk memperoleh data, informasi dan informasi mengenai objek penelitian yang dipilih dan permasalahan apa yang dimiliki oleh perguruan tinggi IIB Darmajaya. Dari proses wawancara ini akan diperoleh data apa saja yang dibutuhkan oleh subjek penelitian dan sistem seperti apa yang diinginkan oleh pengguna.

2.4. Metode Pengembangan Perangkat Lunak

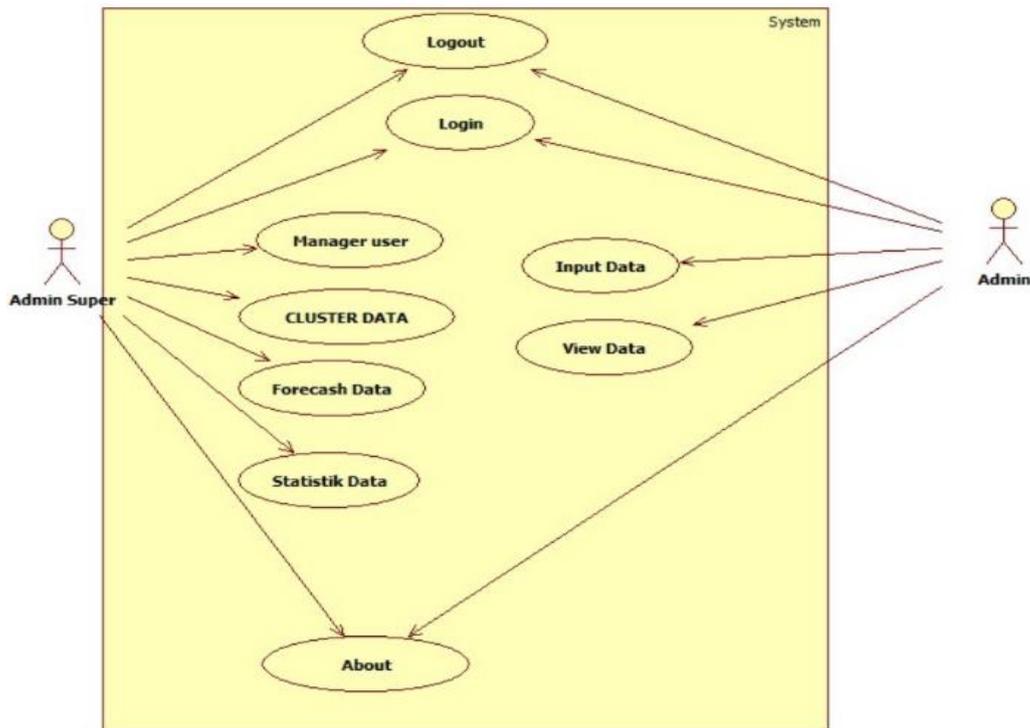
Metode pengembang perangkat lunak yang digunakan dalam aplikasi ini adalah metode *prototype*, dengan tahapan sebagai berikut [10] [11]:

- Komunikasi dan pengumpulan data awal
- *Quick Planning*, yaitu membuat rancangan umum untuk pengembangan lebih lanjut.
- Pembentukan Prototipe, yaitu pembuatan perangkat prototipe termasuk pengujian dan penyempurnaan.
- Evaluasi prototipe, yaitu mengevaluasi prototipe dan menyempurnakan analisis kebutuhan pengguna.
- *Prototype improvement*, yaitu membuat tipe yang sebenarnya berdasarkan hasil evaluasi *prototype*.
- Produksi akhir, yaitu memproduksi perangkat dengan benar sehingga dapat digunakan oleh pengguna.

2.5. Use case diagram

Use case diagram merupakan pemodelan untuk aktivitas pada sistem yang akan dibuat. Rancangan *use case diagram* dapat dilihat pada Gambar 3.

Aktor yang terlibat dalam penentuan pengelompokan judul skripsi IIB Darmajaya terdiri dari admin dan super admin. Super Admin berperan dalam mengolah data dalam sistem pengelompokan judul skripsi yang ada di IIB Darmajaya. Sedangkan peran admin adalah memasukkan data skripsi yang sudah diisi.



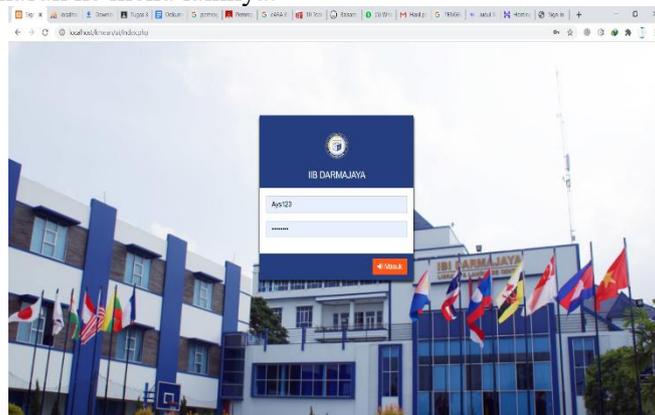
Gambar 1. Use Case Diagram

3. HASIL DAN PEMBAHASAN

Berdasarkan desain antarmuka yang dibuat, dihasilkan output yang terdiri dari: halaman login, halaman dashboard, dan user manager. Tampilan outputnya ditunjukkan pada gambar berikut:

3.1. Tampilan Halaman Login

Pada halaman ini terdapat form login dimana user harus memasukkan user dan password sebelum masuk ke menu lainnya.



Gambar 2. Halaman Login

3.2. Tampilan Halaman Dasbor

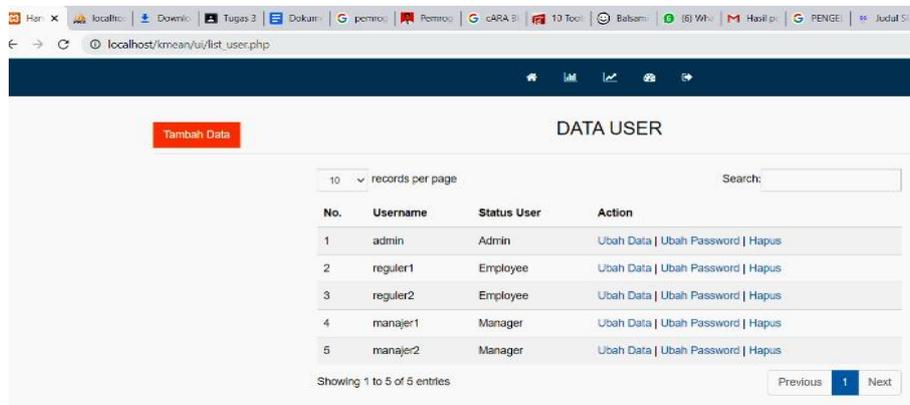
Pada halaman dashboard terdapat beberapa menu yang dapat digunakan oleh user dalam menggunakan aplikasi.



Gambar 3. Halaman Dasbor

3.3. Tampilan Halaman Manajer Pengguna

Halaman Pengelola Pengguna berisi pengelola untuk mengubah atau menambah pengguna baru. Dapat dilihat pada gambar dibawah ini.



Gambar 4. Halaman Manajer Pengguna

3.4. Tampilan Halaman Data *Clustering*

Halaman Data *Clustering*, untuk melihat hasil *cluster* data yang didapat sehingga dapat menganalisa data judul skripsi yang ada untuk dianalisis menggunakan metode *k-means*.

HASIL AKHIR CLUSTERING DATA DENGAN K-MEANS

Jumlah Data = 104
 Jumlah Iterasi = 17
 Jumlah Cluster = 3

[Lihat Detail Iterasi](#)

Pola Ke-1

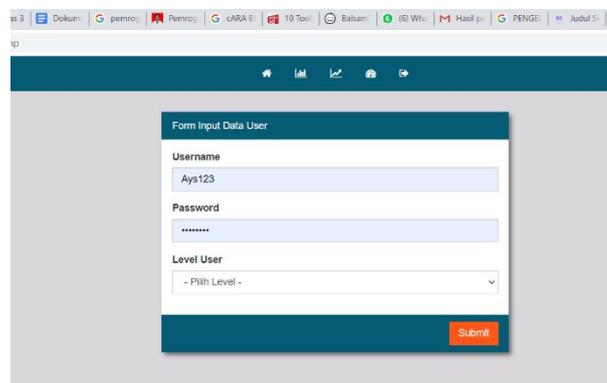
| Prodi | Nama Judul | Dosen Pembimbing | Semester | Status Disetujui |
|--------------------|---|------------------|----------|------------------|
| Teknik Informatika | Rancang Bangun Sistem Penentuan Kualitas Getah Karet Menggunakan Metode K Means Clustering | Hendro | Genap | Y |
| Teknik Informatika | Perancangan Aplikasi Pembelajaran dan Mengirim Pesan Sandi Morse pada smartphone phone berbasis android | Hendro | Genap | Y |
| Teknik Informatika | Aplikasi Perhitungan Key Performance Indicators (KPI) Jurusan Berbasis Website Pada Institut Informatika dan Bisnis Darma | nizar | Ganjil | Y |
| Teknik Informatika | Rancang Bangun Aplikasi Penyedia Layanan Guru Privat Berbasis Android | Hendro | Ganjil | Y |

Gambar 5. Halaman Data *Cluster*

Sistem pengelompokan judul skripsi IIB Darmajaya, akan digunakan tiga klaster, yaitu klaster pertama (C1), klaster kedua (C2) dan klaster ketiga (C3).

3.5. Tampilan Halaman Formulir Pengguna

Berikut ini adalah tampilan form tambah user dimana pada saat super admin ingin menambah data user baru diperlihatkan pada Gambar 6.



Gambar 6. Halaman Data Cluster

3.6. Pengujian

Pengujian dilakukan untuk menjalankan sistem aplikasi ini dapat berjalan seperti yang diharapkan tanpa error atau kesalahan di dalamnya. Pengujian dilakukan pada fungsi menu dan fungsi tombol. Pengujian fungsi menu terdiri dari menu *login* dan menu *manager user*. Tes tombol meliputi: menu *login*, manajer pengguna, cluster data, Forach Data, statistik, menu aktif, dan *logout*. Hasil pengujian fungsi menu dan fungsi tombol berhasil 100%.

4. KESIMPULAN

Hasil percobaan yang dilakukan pada penelitian ini menunjukkan bahwa algoritma *clustering K-Means* melakukan pengelompokan skripsi dengan membagi data menjadi sejumlah *k cluster* yang ditentukan, dan menggunakan perhitungan jarak untuk mengukur kemiripan antar data. Data mining menggunakan *k-mean* mampu melakukan pengelompokan data dalam jumlah besar dengan sangat cepat sehingga dapat membantu mempercepat proses pengelompokan.

5. SARAN

Pada penelitian ini masih memiliki banyak kekurangan sehingga diperlukan rencana pengembangan di penelitian selanjutnya. Beberapa saran yang dilakukan adalah melakukan integrasi dengan aplikasi lain yang ada di perpustakaan IIB Darmajaya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada UPT Perpustakaan, Lembaga penelitian dan pihak manajemen IIB Darmajaya yang telah memfasilitasi penelitian ini.

DAFTAR PUSTAKA

- [1] P. D. J. Wu, *Advances in K-means Clustering A Data Mining Thinking*. 2012.
- [2] R. Goejantoro, "Perbandingan Pengelompokan K-Means dan K-Medoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus : Data Titik Panas Di Indonesia Pada 28 April 2018) Comparison," vol. 10, no. April 2018, pp.

- 143–152, 2019.
- [3] A. Bastian *et al.*, “Penerapan Algoritma K-MEANS Clustering Analisis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka),” no. 1, pp. 26–32.
- [4] I. Nasution, A. P. Windarto, and M. Fauzan, “Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi,” vol. 2, no. 2, pp. 76–83, 2020.
- [5] D. Sunia, P. A. Jusia, T. Informatika, T. Informatika, J. Subdistrict, and J. Selatan, “Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-MEANS,” no. 2016, pp. 121–134, 2017.
- [6] M. K. Canggih Ajika Pamungkas, “Aplikasi Penghitung Jarak Koordinat Berdasarkan Latitude Dan Longitude Dengan Metode Euclidean Distance Dan Metode Haversine,” *J. Inf. Politek. Indonusa Surakarta*, vol. 5, pp. 8–13, 2019.
- [7] L. Studi, K. Pada, U. D. Toko, and B. Yd, “Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan.”
- [8] A. P. Windarto, “Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method,” vol. 1, no. 2, 2017.
- [9] Y. Irawan, “Implementation of Data Mining for Determining for Determining Majors Using K-Means AlgoritmIn Student of SMA Negeri 1 Pangkalan Kerinci,” vol. 1, no. 1, pp. 17–29, 2019.
- [10] N. Kunicina, A. Patlins, and J. Peksa, “Prototyping process in education and science,” 2020.
- [11] S. Alsaqqa, S. Sawalha, and H. Abdel-nabi, “Agile Software Development : Methodologies and Trends Agile Software Development : Methodologies and Trends,” no. July, 2020.