

Implementasi Algoritma *Clustering K-Means* Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika)

¹Aslam Fatkhudin, ²Ahmad Khambali, ³Fenilinas A. Artanto, ⁴Mundriyah, ⁵Nabil A. Putra Zade
^{1, 2, 3, 4, 5} Universitas Muhammadiyah Pekajangan Pekalongan

¹aslamfatkhudin@gmail.com, ²ahmadkhambali@gmail.com, ³fenilinasadi@gmail.com,
⁴mundriyah@umpp.ac.id, ⁵nabil.a.p.z@umpp.ac.id

ABSTRAK

Tugas Akhir menjadi syarat lulus mahasiswa. Tetapi sering menjadi masalah pada mahasiswa karena gagal dalam menyelesaikan tugas akhirnya. Yang menjadi masalah dalam penyelesaian tugas akhir adalah tidak cocok antara dosen pembimbing dan mahasiswa. Dalam mengatasi hal tersebut maka diperlukan pengelompokan mahasiswa untuk dapat dijadikan acuan strategi dalam pembagian dosen pembimbing. Untuk mengelompokan mahasiswa digunakan metode data mining. Salah satu metode data mining adalah clustering. Pada clustering terdapat algoritma K-means. Algoritma K-means menjadi algoritma yang paling banyak digunakan dalam clustering. Karena Algoritma K-means merupakan algoritma yang efisien dan sederhana. Dari pembagian cluster menjadi 4, 5, dan 6 cluster. Didapatkan bahwa pembagian cluster menjadi 5 cluster adalah yang terbaik karena memiliki nilai *davies boulding* yang terkecil. Dari pembagian menjadi 5 cluster terlihat bahwa pada cluster₂ dan cluster₄ memiliki nilai atribut centroid yang berada dibawah yang terindikasi memiliki banyak nilai dalam mata kuliah yang kecil, sehingga untuk dapat memberikan kualitas dalam membina dan membimbing mahasiswa dalam Tugas Akhir diperlukan perhatian yang khusus agar mahasiswa pada cluster₂ dan cluster₄ dapat menyelesaikan Tugas Akhirnya.

Kata Kunci: Clustering, Data mining, K-Means, Mahasiswa, Tugas Akhir

PENDAHULUAN

Tugas akhir pada mahasiswa menjadi syarat umum untuk menjadi syarat kelulusan mahasiswa. Banyak mahasiswa yang sering mengalami kendala dalam penyusunan ataupun menyelesaikan tugas akhir. Kendala yang sering terjadi adalah hilangnya semangat dalam mengerjakan tugas akhir yang biasanya terjadi karena tidak cocok antara mahasiswa dengan dosen pembimbing. Selain itu kelulusan mahasiswa juga menjadi point penting dalam peningkatan kualitas akreditasi sebuah perguruan tinggi. Semakin banyak mahasiswa yang dapat menyelesaikan studinya dengan tepat waktu akan memberikan point penilaian yang baik dalam meningkatkan akreditasi perguruan tinggi.

Untuk mengatasi masalah dalam ketidak cocokan antara mahasiswa dan pembimbing maka diperlukan pembagian dosen pembimbing dan mahasiswa bimbingan yang sesuai. Oleh karena itu dilakukan pengelompokan mahasiswa agar dalam pembagian semakin mudah. Dalam mengelompokan diperlukanya metode data mining. Data mining merupakan sebuah metode dalam mengumpulkan dan menganalisa data (Kusumawardani et al. 2022). Data mining dapat digunakan untuk melakukan pengelompokan data atau biasa disebut dengan clustering (Fatkhudin et al. 2022). Dalam penelitian (Damanik and Sigiro 2021) digunakan algoritma K-means untuk mengelompokan mahasiswa baru yang digunakan untuk pengambilan strategi dalam mempromosikan kampus. Sedangkan pada penelitian (Putra and Yuniarti 2022) menggunakan metode K-Means yang digunakan untuk mengelompokan mahasiswa dengan gaya belajar agar dapat daat memberikan

metode pembelajaran yang tepat. Lalu pada penelitian (Udariansyah and Rahmat Ibrahim 2022) juga menggunakan algoritma K-means untuk mengelompokkan penerimaan mahasiswa untuk digunakan dalam mempromosikan mahasiswa.

Sedangkan dalam penelitian ini akan digunakan metode klustering K-means untuk digunakan dalam pengelompokan mahasiswa yang nantinya hasil dari pengelompokan tersebut akan digunakan untuk pembagian dalam tugas akhir. Digunakanya algoritma K-means karena merupakan algoritma yang paling efisien dan sederhana dalam pengelompokan (Gustientiedina, Adiya, and Desnelita 2019). Akan dilakukan perbandingan dari jumlah kluster mana yang paling cocok dalam pembagian mahasiswa.

TINJAUAN PUSTAKA

Data Mining

Data mining merupakan sebuah proses dalam menambang pengetahuan dari sekumpulan data (Han and Kamber 2006). Data mining merupakan sebuah langkah dalam membersihkan, meintegrasikan, memilih, mentransformasi, evaluasi pola, dan penyajian pengetahuan dari data (Rosyadi et al. 2022). Dalam proses data mining tersusun dalam tiga tahap yaitu pengumpulan, transformasi, dan analisis data (Ayub 2018).

Clustering K-Means

Analisis Kluster merupakan penerapan dalam pengelompokan, atau clustering sesuai dengan karakteristik yang diukur berdasarkan kemiripan satu sama lain. Clustering merupakan pengelompokan menggunakan teknik *unsupervised learning* dimana tidak diperlukan *fase learning* serta tidak menggunakan pelabelan pada setiap kelompok (Priyatman, Sajid, and Haldivany 2019). Algoritma K-means merupakan algoritma yang banyak digunakan dalam pengelompokan karena K-means dikenal sebagai algoritma yang efisien dan sederhana yang merupakan bagian dari 10 algoritma data mining yang teratas (Gustientiedina et al. 2019).

K-Means merupakan algoritma pengklasteran yang akan mengelompokkan data dengan melihat pada nilai pada *centroid* (titik pusat *cluster*). K-Means menggunakan fungsi jarak pada pengklasteran data terhadap *centroid* (Udariansyah and Rahmat Ibrahim 2022). Berikut ini metode dalam algoritma K-Means:

Menentukan nilai k (jumlah kluster yang akan dibentuk)

Membangkitkan nilai centroid secara acak sesuai dengan jumlah kluster (k)

Menghitung jarak pada setiap data dengan centroid dengan menggunakan rumus :

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Dimana:

x_i adalah data

μ_j adalah *centroid* pada *cluster j*

Data akan dikelompokkan pada centroid dengan jarak terkecil

Memperbarui nilai *centroid* dengan nilai *centroid* yang baru didapatkan dari rata-rata *cluster centroid* tersebut dengan rumus:

$$\mu_j(t + 1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j$$

Dimana:

$\mu_j(t + 1)$ = *centroid* baru pada iterasi (t+1)

N_{sj} = data pada *cluster S_j*

Selanjutnya akan kembali pada proses awal, iterasi akan diulang sampai didapatkan titik centroid yang tidak berubah lagi nilainya atau iterasi dihentikan berdasarkan dengan jumlah iterasi maksimal yang telah ditentukan.

METODE PENELITIAN

Populasi dan Sampel

1. Populasi

Populasi sasaran penelitian ini adalah 40 Mahasiswa dari Program Studi Manajemen Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah Pekajangan Pekalongan.

2. Sampel

Dari seluruh Mahasiswa dari Program Studi Manajemen Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah Pekajangan Pekalongan diambil sampel mahasiswa semester 4 sebanyak 22 Mahasiswa.

Teknik Model Analisis

Penelitian ini menggunakan metode algoritma K-means. Untuk dapat melakukan pengelompokan dengan menggunakan algoritma K-means diperlukan proses:

Menentukan nilai K sebagai jumlah cluster yang akan dibentuk dengan rumus:

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|N_i - C_k\| \quad 1$$

Tentukan titik k pusat cluster (centroid) awal yang dilakukan secara random. Penentuan pada centroid awal dilakukan secara acak atau random dari objek yang tersedia sebanyak k cluster untuk menghitung centroid cluster ke-i digunakan rumus:

$$v = \frac{\sum_{i=1}^n X_i}{n} : i = 1, 2, 3, \dots, n \quad 2$$

Menghitung Jarak setiap objek ke masing-masing centroid.

Alokasikan masing-masing objek ke dalam centroid terdekat.

Lakukan iterasi dan tentukan posisi centroid baru

Ulangi langkah tiga jika posisi centroid baru tidak sama.

Dalam menggunakan algoritma K-means akan dibantu dengan menggunakan aplikasi data mining yaitu rapidminer.

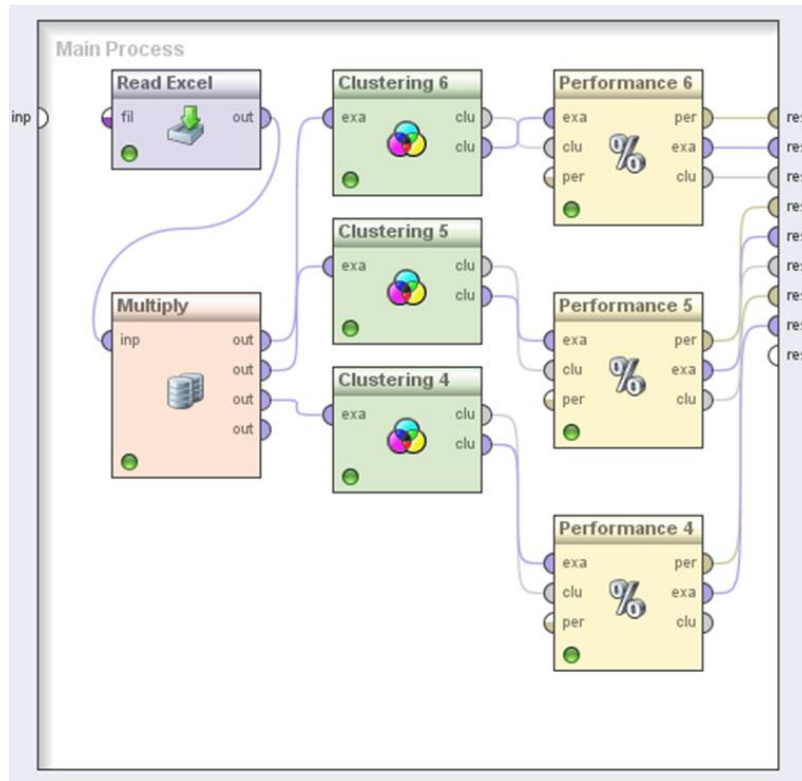
HASIL DAN PEMBAHASAN

Hasil

Dari data yang sudah didapat maka dibuatlah pemodelan algoritma K-Means Clustering dengan menggunakan *software RapidMiner* dengan desain seperti berikut:

Dengan menggunakan algoritma K-Means Clustering dengan jumlah data mahasiswa

Didapatkan perbandingan nilai Davies Boildin sebagai berikut:



Gambar 1 Pemodelan K-Means dengan RapidMiner

Tabel 1 Nilai Davies Boulding

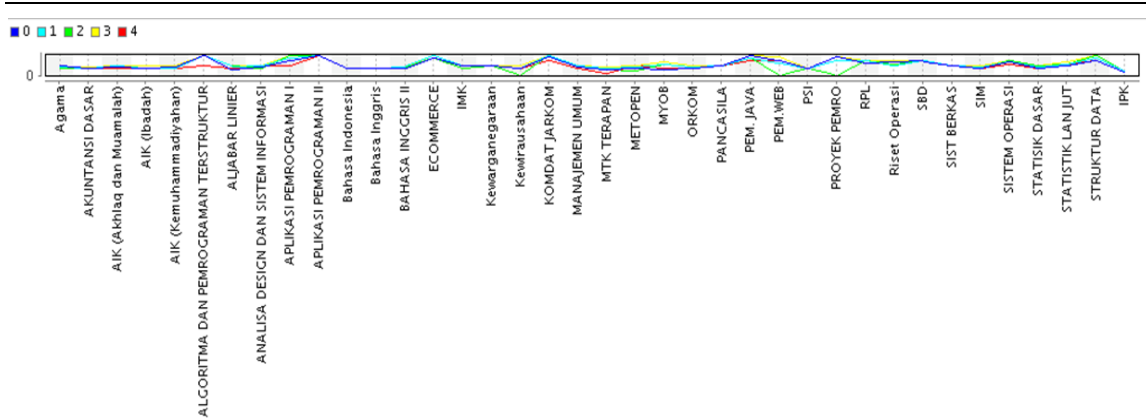
| Jumlah Clustering | Nilai Davies Boulding |
|-------------------|-----------------------|
| 4 | -1,561 |
| 5 | -1,126 |
| 6 | -1,148 |

Dari perbandingan nilai davies boulding terlihat bahwa nilai pada pembagian 5 kluster ($k=5$) memiliki nilai terendah, sehingga pembagian kelompok menjadi 5 cluster adalah yang terbaik. Oleh karena itu diambilah pembagian cluster menjadi 5 cluster dan dari pembagian 5 cluster dihasilkan pembagian cluster:

Tabel 2 Hasil Pembagian Clustering K-Means

| Clustering | Items |
|------------|-------|
| 0 | 4 |
| 1 | 2 |
| 2 | 2 |
| 3 | 13 |
| 4 | 1 |

Dari 5 cluster yang telah dibagi didapatkan hasil sebaran cluster sebagai berikut:



Gambar 2 Persebaran Clustering dengan K-Means

Dari hasil persebaran dibagi sesuai dengan 5 kelompok cluster, terlihat dari hasil cluster terlihat bahwa clustering ke 2 terlihat banyak tersebar pada baris bawah, jadi untuk clustering ke 2 merupakan kumpulan mahasiswa yang perlu di waspadai.

Dan dari pembagian cluster menghasilkan tabel centroid sebagai berikut:

Tabel 3 Nilai Centroid Tiap Atribut

| Attribute | Cluster_0 | Cluster_1 | Cluster_2 | Cluster_3 | Cluster_4 |
|---------------------------------------|-----------|-----------|-----------|-----------|-----------|
| Agama | 7.50 | 7.00 | 6.00 | 6.15 | 6.00 |
| AKUNTANSI DASAR | 6.00 | 6.00 | 6.00 | 6.62 | 6.00 |
| AIK (Akhlak dan Muamalah) | 6.50 | 8.00 | 7.00 | 7.54 | 6.00 |
| AIK (Ibadah) | 6.00 | 6.00 | 6.00 | 7.69 | 6.00 |
| AIK (Kemuhammadiyah) | 6.50 | 6.00 | 6.00 | 7.54 | 6.00 |
| ALGORITMA DAN PEMROGRAMAN TERSTRUKTUR | 16.00 | 16.00 | 16.00 | 16.00 | 8.00 |
| ALJABAR LINIER | 4.50 | 7.00 | 8.00 | 7.54 | 6.00 |
| ANALISA DESIGN DAN SISTEM INFORMASI | 6.50 | 8.00 | 6.00 | 8.00 | 8.00 |
| APLIKASI PEMROGRAMAN I | 12.00 | 14.00 | 16.00 | 16.00 | 8.00 |
| APLIKASI PEMROGRAMAN II | 16.00 | 16.00 | 16.00 | 16.00 | 16.00 |
| Bahasa Indonesia | 5.50 | 6.00 | 6.00 | 6.00 | 6.00 |
| Bahasa Inggris | 6.00 | 6.00 | 6.00 | 6.00 | 6.00 |
| BAHASA INGGRIS II | 6.00 | 7.00 | 7.00 | 6.92 | 6.00 |
| ECOMMERCE | 14.00 | 16.00 | 14.00 | 16.00 | 16.00 |
| INTERAKSI MANUSIA DAN KOMPUTER | 7.50 | 8.00 | 6.00 | 8.00 | 8.00 |
| Kewarganegaraan | 8.00 | 8.00 | 8.00 | 8.00 | 8.00 |
| Kewirausahaan | 6.00 | 6.00 | 0.00 | 7.54 | 6.00 |
| KOMUNIKASI DATA DAN JARINGAN KOMPUTER | 15.00 | 16.00 | 16.00 | 16.00 | 12.00 |
| MANAJEMEN UMUM | 6.50 | 8.00 | 8.00 | 7.69 | 6.00 |
| MATEMATIKA TERAPAN | 5.00 | 6.00 | 5.00 | 6.46 | 2.00 |

| Attribute | Cluster_0 | Cluster_1 | Cluster_2 | Cluster_3 | Cluster_4 |
|----------------------------|-----------|-----------|-----------|-----------|-----------|
| METODOLOGI PENELITIAN | 5.50 | 7.00 | 4.00 | 7.54 | 8.00 |
| MYOB | 5.25 | 9.00 | 9.00 | 11.08 | 6.00 |
| ORGANISASI KOMPUTER | 6.00 | 7.00 | 7.00 | 7.69 | 6.00 |
| PANCASILA | 8.00 | 8.00 | 8.00 | 8.00 | 8.00 |
| PEMROGRAMAN JAWA | 16.00 | 14.00 | 14.00 | 16.00 | 12.00 |
| PEMROGRAMAN WEB | 12.00 | 10.00 | 0.00 | 14.46 | 12.00 |
| PENGANTAR SISTEM INFORMASI | 6.00 | 6.00 | 6.00 | 6.00 | 6.00 |
| PROYEK PEMROGRAMAN | 15.00 | 12.00 | 0.00 | 15.08 | 12.00 |
| REKAYASA PERANGKAT LUNAK | 9.75 | 12.00 | 12.00 | 12.00 | 12.00 |
| RISET OPERASI | 10.50 | 7.50 | 9.00 | 12.00 | 12.00 |
| SISTEM BASIS DATA | 11.25 | 12.00 | 12.00 | 12.00 | 12.00 |
| SISTEM BERKAS | 8.00 | 8.00 | 8.00 | 8.00 | 8.00 |
| SISTEM INFORMASI MANAJEMEN | 6.00 | 7.00 | 6.00 | 8.00 | 6.00 |
| SISTEM OPERASI | 10.50 | 10.50 | 12.00 | 11.77 | 9.00 |
| STATISIK DASAR | 6.00 | 7.00 | 8.00 | 7.38 | 6.00 |
| STATISTIK LAN JUT | 8.25 | 9.00 | 9.00 | 10.85 | 9.00 |
| STRUKTUR DATA | 12.00 | 14.00 | 16.00 | 15.38 | 12.00 |
| IPK | 3.28 | 3.50 | 3.44 | 3.79 | 3.15 |

Dari nilai atribut tiap centroid maka nantinya dalam pengelompokan mahasiswa Manajemen Informatika yang perlu diwaspadai adalah yang berada di cluster 2 karena memiliki beberapa nilai atribut dengan nilai centroid yang mendekati 0, selain itu juga pada cluster 4 karena memiliki nilai ipk terendah dari pada cluster yang lainnya.

Pembahasan

Hasil penelitian ini membuktikan bahwa pembagian mahasiswa lebih baik dibagi menjadi 5 cluster karena dari perbandingan 6, 5 dan 4 cluster menunjukkan bahwa pembagian dengan 5 cluster memiliki nilai *davies boulding* yang terkecil, sehingga pembagian 5 cluster adalah yang terbaik.

Dari pembagian 5 cluster didapatkan bahwa persebaran mahasiswa sebagai berikut. Cluster_0 terdapat 4 mahasiswa, Cluster_1 terdapat 2 mahasiswa, Cluster_2 terdapat 2 mahasiswa, Cluster_3 terdapat 13 mahasiswa, dan Cluster_4 terdapat 1 mahasiswa.

Pada pembagian 5 cluster mahasiswa dapat dilihat bahwa mahasiswa pada cluster_2 dan cluster_4 perlu diwaspadai karena pada cluster_2 memiliki beberapa atribut yang nilai centroidnya mendekati 0. Sedangkan pada cluster_4 memiliki nilai atribut ipk yang terkecil dari cluster lainnya. Sehingga pada cluster_2 dan cluster_4 perlu mendapat perhatian lebih dalam pembagian pembimbing saat tugas akhir nantinya.

KESIMPULAN

Dari pengelompokan mahasiswa pada program studi manajemen informatika, telah dibandingkan pengelompokan mahasiswa dengan membandingkan 6, 5 dan 4 clustering mahasiswa. Dari perbandingan tersebut terlihat bahwa pembagian cluster menjadi 5 kelompok adalah yang terbaik karena memiliki nilai *davies boulding* terkecil daripada 2 pembagian cluster lainnya. Oleh karena itu mahasiswa manajemen informatika dibagi menjadi 5 cluster. Dari ke 5

cluster mahasiswa manajemen informatika tersebut ada 2 cluster yang perlu di waspadai yaitu cluster_2 dan cluster_4. Kedua cluster tersebut memiliki nilai pada atribut mata kuliah yang lebih kecil daripada cluster lainnya. Sehingga perlu adanya perhatian khusus pada kedua kluster tersebut.

REFERENSI

- Ayub, Mewati. 2018. "Proses Data Mining Dalam Sistem Pembelajaran Berbantuan Komputer." (May):21–30.
- Damanik, Nurafni, and Mula Sigiro. 2021. "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru Sebagai Metode Promosi." *Jurnal Teknik Informatika Komputer Universal (Jutisal)* 4(2):158.
- Fatkhudin, Aslam, M. Yusuf Febrianto, Fenilinas Adi Artanto, M. Waffa Najib Hadinata, and Riza Fahlevi. 2022. "ALGORITMA DECISION TREE C.45 DALAM ANALISA KELULUSAN MAHASISWA PROGRAM STUDI MANAJEMEN INFORMATIKA UMPP." *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar* 8(2):83–86.
- Gustientiedina, Gustientiedina, M. Hasmil Adiya, and Yenny Desnelita. 2019. "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan." *Jurnal Nasional Teknologi Dan Sistem Informasi* 5(1):17–24. doi: 10.25077/teknosi.v5i1.2019.17-24.
- Han, Jiawei, and Micheline Kamber. 2006. *Mining Stream, Time-Series and Sequence Data*. Vol. 54.
- Kusumawardani, Hadwitya Handayani, Imam Rosyadi, Fenilinas Adi Artanto, Fauzan Iryan Arzha, and Niar Ajeng Rachmayani. 2022. "Analisis Decision Tree Dalam Pengaruh Digital Marketing Terhadap Penerimaan Siswa Baru." *Remik* 6(April):225–31.
- Priyatman, Hendro, Fahmi Sajid, and Dannis Haldivany. 2019. "Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Waktu Kelulusan Mahasiswa." *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)* 5(1):62. doi: 10.26418/jp.v5i1.29611.
- Putra, Berlian Juliartha Martin, and Dwi Ariani Finda Yuniarti. 2022. "Analisis Gaya Belajar Terhadap Nilai Mahasiswa Dengan Menggunakan Metode K-Means." *Techno.Com* 21(2):343–54. doi: 10.33633/tc.v21i2.5837.
- Rosyadi, Imam, Fenilinas Adi Artanto, Siti Elisa Rahmawati, Haifan Tri, and Buwono Joyo. 2022. "Decision Tree Dalam Analisis Keputusan Pembelian Program Pada Perkumpulan Penggiat Programmer Indonesia." *Jurnal Fasilkom XII(III)*:141–44.
- Udariansyah, Devi, and Deny Rahmat Ibrahim. 2022. "Klasifikasi Data Penerimaan Mahasiswa Baru Pada Universitas Bina Darma Menggunakan Algoritma K-Means Clustering." *Jurnal Pendidikan Dan Konseling* 4(4):2692–2701.