

# KLASTERISASI POTENSI MAHASISWA PROGRAM STUDI MATEMATIKA UNIVERSITAS MULAWARMAN MENGGUNAKAN METODE *K-MEANS* UNTUK MENYARING MAHASISWA BERPRESTASI

I. Farha, N.F. Saputri, M. Elvita, A.A. Fauzi, dan \*F.D.T. Amijaya

Program Studi S-1 Matematika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Jl. Kuaro, Gn. Kelua Samarinda.

[izzatyfarha25@gmail.com](mailto:izzatyfarha25@gmail.com), [nolafebrianasaputri@gmail.com](mailto:nolafebrianasaputri@gmail.com), [elvitamelati@gmail.com](mailto:elvitamelati@gmail.com), [andriazmul161022@fmipa.unmul.ac.id](mailto:andriazmul161022@fmipa.unmul.ac.id), [fidiadta@fmipa.unmul.ac.id](mailto:fidiadta@fmipa.unmul.ac.id) \*corresponding author

## Abstrak

Dengan menggunakan metode *K-Means* penelitian ini mengelompokkan mahasiswa Program Studi Matematika Universitas Mulawarman berdasarkan prestasi akademik mereka. Mengelompokkan mahasiswa ke dalam beberapa klaster berdasarkan kesamaan nilai akademik, diharapkan dapat membantu dalam penyaringan mahasiswa berprestasi secara lebih objektif. Data yang digunakan terdiri dari nilai mata kuliah mahasiswa angkatan 2023 selama dua semester, dengan pendekatan kuantitatif eksploratif dan metode *data mining*. Proses analisis dilakukan menggunakan *Google Colab* dan *Elbow Method* untuk menentukan jumlah klaster optimal, yang menghasilkan tiga klaster mahasiswa: mahasiswa dengan nilai tinggi, mahasiswa dengan nilai sedang, dan mahasiswa dengan nilai rendah. Hasil penelitian menunjukkan bahwa metode *K-Means* efektif dalam mengelompokkan data prestasi akademik, serta membantu Program Studi Matematika dalam penyaringan mahasiswa berprestasi. Visualisasi hasil klasterisasi menunjukkan distribusi mahasiswa dalam tiga kelompok utama dengan *centroid* sebagai representasi rata-rata kelompok, sehingga diharapkan proses identifikasi dan pengembangan potensi mahasiswa dapat dilakukan secara lebih sistematis dan terukur.

**Kata kunci:** *Data Mining*, *K-Means*, Klasterisasi, Prestasi Akademik

## 1 Pendahuluan

Mahasiswa adalah bagian penting dalam perkembangan suatu institusi pendidikan tinggi, terutama dalam meningkatkan kualitas akademik. Salah satu indikator keberhasilan program studi adalah prestasi akademik mahasiswa. Di Program Studi Matematika Universitas Mulawarman, identifikasi mahasiswa berprestasi memiliki peran penting, baik dalam memberikan apresiasi kepada mereka yang berprestasi maupun sebagai dasar dalam pengambilan keputusan terkait pengembangan akademik.

Namun, proses identifikasi mahasiswa berprestasi yang selama ini dilakukan seringkali bersifat subjektif dan terbatas pada penilaian formal. Oleh karena itu, diperlukan pendekatan yang lebih objektif dan berbasis data. Untuk tujuan ini metode *K-Means* adalah salah satu metode yang dapat digunakan, sebuah teknik klustering yang dapat digunakan untuk mengelompokkan mahasiswa berdasarkan kesamaan prestasi akademik mereka. Dengan metode ini, kelompok mahasiswa yang berprestasi dapat diidentifikasi secara lebih jelas dan terukur.

*2020 Mathematics Subject Classification:* 62H30, 68T10.

Diajukan: 31/10/2024, diterima: 31/12/2024. DOI: <https://doi.org/10.29244/milang.20.2.89-99>

MILANG Journal of Mathematics and Its Applications, Vol.20, No.2, pp.89-99

ISSN: 2963-5233

Penelitian ini berjudul "Klasterisasi Potensi Mahasiswa Program Studi Matematika Universitas Mulawarman Menggunakan Metode *K-Means* untuk Menyaring Mahasiswa Berprestasi," tujuan penelitian ini adalah untuk mengelompokkan mahasiswa berdasarkan prestasi akademik mereka menggunakan metode *K-Means*. Oleh karena itu, temuan penelitian ini diharapkan dapat membantu Program Studi Matematika dalam mengidentifikasi mahasiswa yang berprestasi dengan lebih efisien dan efektif.

Berdasarkan uraian yang dijelaskan sebelumnya, pada penelitian ini akan dibahas bagaimana cara mengelompokkan prestasi akademik mahasiswa Program Studi Matematika Universitas Mulawarman dengan menggunakan metode *K-Means* dan kedua, apakah metode *K-Means* efektif dalam membantu menyaring mahasiswa berprestasi berdasarkan prestasi akademik.

## 2 Tinjauan Pustaka

### 2.1 Data Mining

*Data mining* mengacu pada konsep menambang basis data besar untuk mengekstrak informasi atau pengetahuan berharga dari data yang terkandung di dalamnya. Pengetahuan berharga tersebut bersifat implisit dan belum diketahui sehingga dapat juga disebut sebagai informasi harta karun. Penggunaan *data mining* sering memiliki keterkaitan dengan berbagai bidang, seperti statistik, *machine learning*, *database technology*, dan *computing algorithms* [3].

Karena *data mining* adalah salah satu proses dalam *Knowledge Discovery in Database* (KDD), *data mining* juga sering dikaitkan dengan KDD. KDD adalah teknik untuk mendapatkan informasi atau pengetahuan dari basis data, dan *data mining* menggunakan berbagai teknik untuk menghasilkan informasi penting seperti deskripsi, estimasi, prediksi, klasifikasi, asosiasi, dan klasterisasi [5].

Fungsi utama *data mining* meliputi deskriptif, prediktif, asosiasi, klasifikasi, Klasterisasi, *forecasting*, dan *sequencing*, yang semuanya bertujuan untuk memahami karakteristik data, memprediksi hasil di masa depan, serta mengelompokkan dan menganalisis perilaku konsumen. Teknik-teknik ini membantu bisnis meningkatkan efisiensi dan daya saing dengan mendeteksi tren pasar, mengetahui preferensi pelanggan, dan membuat keputusan bisnis yang lebih baik [4].

### 2.2 Klasterisasi

Tujuan dari klasterisasi, suatu alat bantu data mining, adalah untuk membagi objek ke dalam beberapa klaster tertentu. Teknik ini mengelompokkan objek yang memiliki banyak kesamaan satu sama lain. Sehingga tingkat similaritas objek yang berbeda klaster akan lebih rendah dibandingkan dengan tingkat objek yang berada dalam klaster yang sama. Akibatnya, klaster dapat didefinisikan sebagai sekumpulan objek data yang memiliki karakteristik yang serupa [1].

Analisis pengelompokan, juga dikenal sebagai klasterisasi, adalah proses membagi data ke dalam kelompok berdasarkan tingkat kesamaan yang lebih tinggi dalam satu kelompok dibandingkan dengan kelompok lain. Klasterisasi memiliki potensi besar untuk mengungkap struktur dalam data, dan dapat digunakan untuk berbagai tujuan seperti klasifikasi, pengolahan gambar, dan pengenalan pola. Klasterisasi memiliki dua teknik utama: *hierarchical clustering*, yang mengelompokkan data secara bertahap membentuk struktur pohon (dendrogram) berdasarkan kedekatan antar objek, dan *non-hierarchical*

*clustering*, yang biasanya mengelompokkan data tanpa struktur hierarki yang jelas. Meskipun pada akhirnya semua objek akan tergabung dalam satu klaster, proses pengelompokan ini memberikan wawasan mendalam tentang hubungan antar data [6].

Teknik klasterisasi dilakukan secara otomatis tanpa diberitahukan label kelasnya pada data. Hal itu dapat memberikan label baru pada kelas data yang belum diketahui, oleh karenanya tergolong dalam pembelajaran yang tidak diawasi. *K-Medoids*, *Fuzzy C-Means*, *Self-Organizing Map* (SOM), dan *K-Means* adalah metode klasterisasi yang sering digunakan [3].

### 2.3 Metode *Elbow*

Metode *Elbow* untuk analisis klasterisasi digunakan untuk menentukan jumlah klaster yang ideal untuk suatu dataset. Sebuah representasi grafik dari hasil perbandingan dari jumlah total  $k$  dengan hasil *Sum of Squared Error* (SSE) digunakan untuk menentukan proses penentuan jumlah klaster. Representasi grafik dari  $k$  dengan nilai SSE yang semakin rendah dapat menunjukkan klasterisasi dengan jumlah klaster yang optimal dan terbaik [7].

### 2.4 Algoritma *K-Means*

Peneliti menganggap algoritma *K-Means* sebagai salah satu algoritma *data mining* yang paling terkenal. Penggunaan algoritma *K-Means* sendiri dinilai cukup tangguh dalam menghadapi berbagai jenis data. Langkah-langkah klasterisasi menggunakan algoritma *K-Means* secara umum:

- a. Menentukan jumlah  $k$  yang akan digunakan untuk membentuk klaster.
- b. *Centroid* awal, yang juga dikenal sebagai titik pusat klaster, dikumpulkan secara acak, sebanyak jumlah  $k$  klaster. Selanjutnya, gunakan persamaan berikut untuk menghitung *centroid* suatu klaster:

$$v = \frac{1}{k} \sum_{i=1}^k x_i \quad (1)$$

dengan

$v$  = *centroid* pada klaster,

$x_i$  = objek ke- $i$  pada klaster,

$k$  = banyaknya jumlah objek yang menjadi anggota klaster.

- c. Untuk masing-masing data klaster, dapat digunakan persamaan jarak geometris untuk menghitung jarak antara setiap objek dan masing-masing *centroid*, yaitu:

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

dengan

$x_i$  = atribut ke- $i$  pada objek  $x$ ,

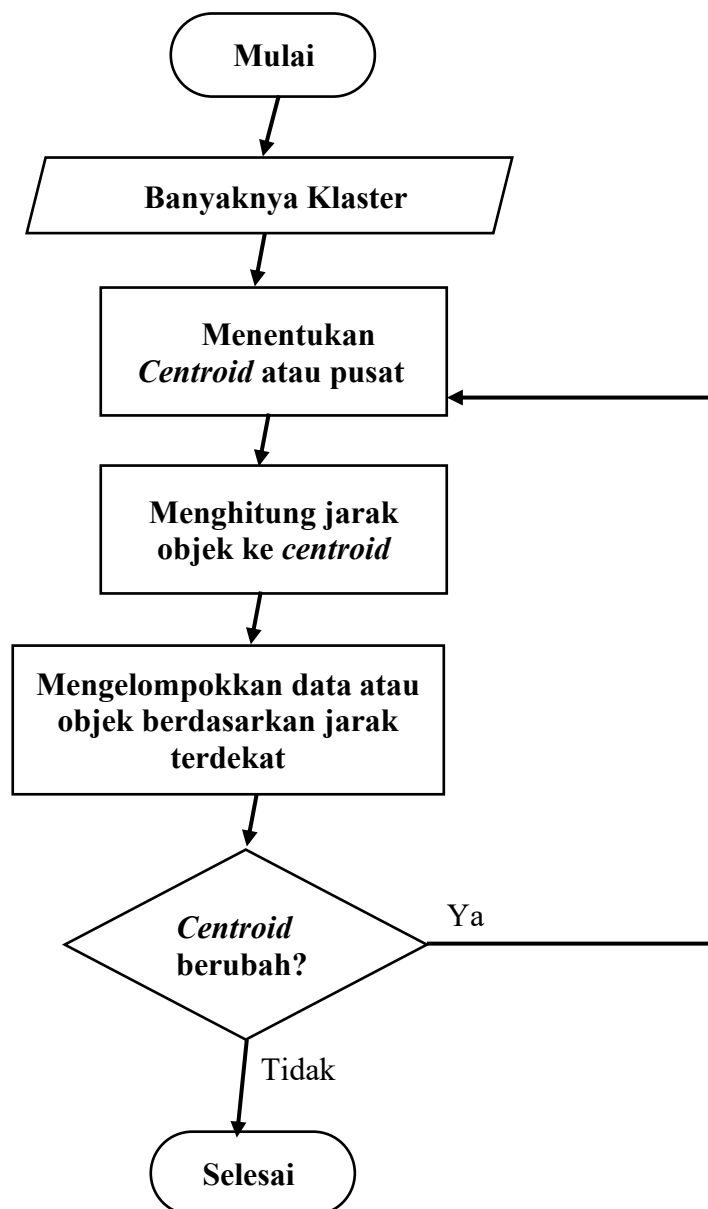
$y_i$  = atribut ke- $i$  pada objek  $y$ ,

$n$  = banyaknya atribut.

- d. Menempatkan semua objek ke *centroid* yang paling terdekat.
- e. Setelah melakukan iterasi ke- $i$ , gunakan persamaan yang ditemukan di langkah b untuk menentukan posisi *centroid* baru.

- f. Jika posisi *centroid* baru tidak sama, ulangi langkah pada poin c. Selain itu, jika rasio perbandingan jarak objek dengan *centroid* baru tidak lebih besar dari nilai rasio perbandingan jarak objek dengan *centroid* sebelumnya, iterasi atau perulangan akan berhenti. Hal ini terjadi hingga hasil perhitungan di masing-masing data konvergen [2].

Berikut gambaran langkah-langkah algoritma *K-Means* dengan menggunakan *flowchart*:



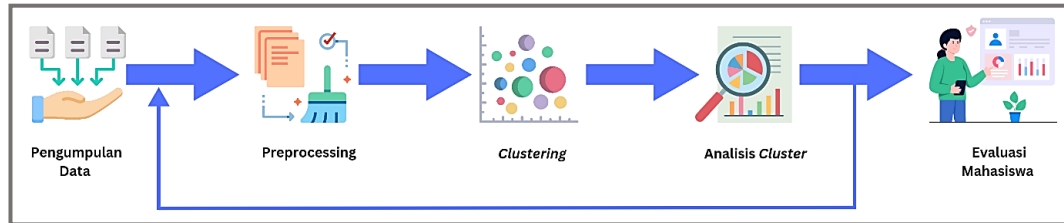
### 3 Metode Penelitian

Penelitian kuantitatif eksploratif ini menggunakan metode *data mining* (pengelompokan data) untuk menyaring dan menemukan mahasiswa yang potensial. Penerapan algoritma *K-Means* pada dataset mahasiswa adalah fokus utama penelitian.

Subjek penelitian adalah mahasiswa Program Studi Matematika Universitas Mulawarman dari angkatan tertentu yaitu 2023 yang merupakan salah satu angkatan aktif.

Jumlah sampel akan disesuaikan dengan data yang tersedia, yang mencakup data akademik.

Data kuantitatif termasuk data akademik per mata kuliah yang akan digunakan dalam penelitian ini. Data ini dikumpulkan melalui sistem akademik Universitas Mulawarman dan hasil survei.



Gambar 1. Diagram Sistem

Blok diagram sistem ditunjukkan pada Gambar 1. Penelitian ini menggunakan data dari mahasiswa Universitas Mulawarman angkatan 2023 selama dua semester, yang mencakup nilai mahasiswa untuk setiap mata kuliah.

Salah satu langkah dalam *preprocessing* adalah mengisi data yang hilang. Sebelum mengisi data kosong akan diperiksa untuk memastikan apakah mahasiswa cuti, mengundurkan diri, atau kehilangan data.

Dalam penelitian ini mahasiswa dikelompokkan ke dalam tiga kluster: mahasiswa dengan nilai tinggi, mahasiswa dengan nilai sedang dan mahasiswa dengan nilai rendah. Untuk melakukan klustering dengan metode *K-Means*, langkah-langkah berikut diikuti:

1. Menentukan nilai  $k = 3$ , yaitu jumlah kluster yang ingin dibentuk.
2. Inisialisasi pusat kluster dapat dilakukan dengan berbagai metode. Namun dalam penelitian ini metode inisialisasi dilakukan secara acak berdasarkan data yang tersedia. Prosesnya dimulai dengan menentukan jumlah kluster, kemudian memilih pusat kluster awal secara acak dari dataset. Pusat-pusat tersebut digunakan sebagai titik awal untuk proses pengelompokan yang dilanjutkan dengan iterasi perhitungan ulang pusat berdasarkan rata-rata data di setiap kluster hingga mencapai konvergensi.
3. Gunakan rumus jarak geometri (*Euclidean Distance*) untuk menghitung jarak terdekat antara data dan *centroid*.

$$d(x, c_j) = \sqrt{\sum_{i=1}^n (x_i - c_{ji})^2} \quad (3)$$

dengan

$d$  = jarak antara nilai data dan nilai pada pusat kluster (*centroid*),

$c_j$  = *centroid* dari kluster ke- $j$ ,

$x_i$  = atribut ke- $i$  pada objek  $x$ ,

$c_{ji}$  = atribut ke- $i$  pada objek  $c_j$ .

4. Kelompokkan setiap data menurut jarak terdekat dengan *centroid*.
5. Perbarui nilai *centroid*. Nilai *centroid* baru dihitung dengan menggunakan rumus berikut untuk menghitung rata-rata data dari kluster yang terkait:

$$c_j = \frac{1}{k_j} \sum_{x \in S_j} x \quad (4)$$

Keterangan:

$c_j$  = *centroid* baru untuk kluster  $S_j$ , yang dihitung sebagai rata-rata dari semua data dalam kluster tersebut,

$k_j$  = banyak objek pada kluster  $S_j$

$S_j$  = himpunan semua objek yang termasuk dalam kluster ke- $j$

$x$  = objek pada kluster  $S_j$

6. Ulangi langkah dari poin tiga sampai lima hingga setiap anggota kluster tidak berubah atau hingga *centroid* yang lama sama dengan *centroid* yang baru.


#### 4 Hasil dan Pembahasan

Penelitian ini menggunakan alat *Google Colab* untuk mempelajari pengelompokan algoritma *K-Means* menggunakan dua puluh data nilai mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam, Program Studi Matematika Angkatan 2023. Tabel 1 merupakan data nilai mahasiswa, dimana PLMH: Pengantar Logika Matematika dan Himpunan, Kael: Kalkulus Elementer, PK1: Pemrograman Komputer I, Ale: Aljabar Linier Elementer, PK2: Pemrograman Komputer II, dan Kalin: Kalkulus Integral.

Tabel 1. Data nilai mahasiswa

Mahasiswa	PLMH	Kael	PK1	Ale	PK2	Kalin
A	74.95	74	77.9	67.45	82.85	81
B	70.6	72.1	73.5	67.9	82.37	78.25
C	70.15	75	78	75.1	83.04	83.2
D	70.6	73	74.7	70.15	79.68	79.5
E	74.95	70.6	72.7	65.2	80.81	79
F	89.5	80.1	86.2	85.45	84.92	82.75
G	63.85	72.52	75.3	64.75	82.44	79.5
H	78.25	74.06	76.1	89.95	80.35	83.95
I	65.2	70.08	79.3	66.1	79.45	79.5
J	76	74.4	84	72.85	86.06	82.1
K	63.85	72.94	74.5	66.55	80.78	79.5
L	64.75	73.54	75.9	62.05	82.5	80
M	83.2	79.3	84.4	84.55	86.19	84
N	81.4	74.36	84.9	86.8	83.33	80.45
O	71.95	73.2	79.3	71.95	83.71	78.25
P	89.5	84.34	83.9	85.9	81.82	84
Q	81.4	73	75.1	76	81.26	80
R	71.95	70.8	79.3	66.55	82.09	79.2
S	82.2	74.3	81.8	92.4	81.5	78.4
T	72.85	73.54	79.7	65.65	80	81.2

```

0 d  import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

```

Gambar 2. Impor Pustaka.

Gambar 3 untuk membaca file CSV bernama 'DataMahasiswa.csv' dan menyimpannya sebagai DataFrame dalam variabel 'dm'. Lima baris pertama dari DataFrame ditampilkan dengan menggunakan fungsi 'head()'.

```

0 d  dm=pd.read_csv('DataMahasiswa.csv')
dm.head()

```

	Inisial	PLMH	Kael	PK1	Ale	PK2	Kalin
0	A	74.95	74.0	77.9	67.45	82.85	81.00
1	B	70.60	72.1	73.5	67.90	82.37	78.25
2	C	70.15	75.0	78.0	75.10	83.04	83.20
3	D	70.60	73.0	74.7	70.15	79.68	79.50
4	E	74.95	70.6	72.7	65.20	80.81	79.00

Langkah berikutnya: [Buat kode dengan dm](#) [Lihat plot yang direkomendasikan](#) [New interactive sheet](#)

Gambar 3. Memanggil data.

Gambar 4 mengambil kolom kedua hingga keenam dari DataFrame dan mengonversi hasilnya menjadi array NumPy

```


0 d  x = dm.iloc[:, [1, 2, 3, 4, 5, 6]].values

```

Gambar 4. Memilih kolom.

Gambar 5 mengimpor algoritma *K-Means* dari scikit-learn library

```

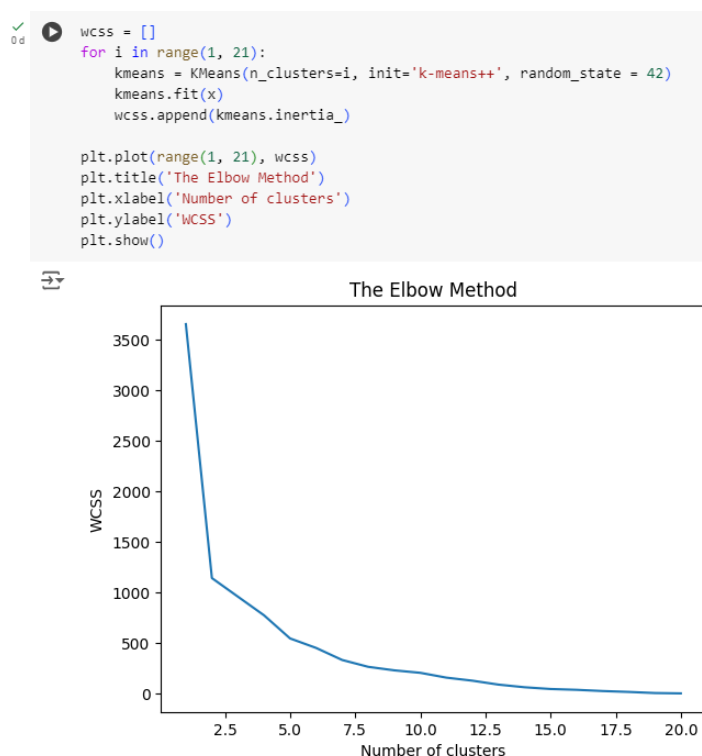
0 d  from sklearn.cluster import KMeans

```

Gambar 5. Impor *K-Means*.

Gambar 6 menggambarkan penggunaan metode *Elbow* untuk menentukan jumlah kluster optimal dalam analisis *K-Means*. Dalam pendekatan ini algoritma *K-Means* diterapkan dengan jumlah kluster yang bervariasi dari 1 hingga 20 dan nilai WCSS (*Within-Cluster Sum of Squares*) dihitung serta disimpan untuk setiap jumlah kluster. WCSS mengukur sejauh mana data dapat dikelompokkan dengan baik dalam setiap kluster di mana nilai WCSS yang lebih kecil menunjukkan pemisahan kluster yang lebih efektif. Grafik yang dihasilkan menggambarkan hubungan antara jumlah kluster dan nilai WCSS. Penurunan yang tajam pada WCSS di awal diikuti dengan penurunan yang lebih

perlahan menunjukkan titik *elbow* yang mengindikasikan jumlah kluster optimal. Dalam penelitian ini jumlah kluster yang telah ditetapkan sebanyak 3 kluster, yang sesuai dengan titik *elbow* yang teridentifikasi. Titik ini menunjukkan bahwa penambahan jumlah kluster lebih lanjut tidak memberikan pengurangan WCSS yang signifikan, yang mengindikasikan bahwa 3 kluster sudah cukup untuk mengelompokkan data secara efektif tanpa menyebabkan *overfitting*. Dengan demikian jumlah kluster 3 memberikan hasil terbaik dalam hal pemisahan data yang optimal sambil mempertahankan efisiensi komputasi.



Gambar 6. Metode *Elbow*.

Gambar 7 menginisialisasi model *K-Means* dengan jumlah kluster yang ditentukan sebanyak 3, menggunakan metode 'k-means++' untuk inisialisasi *centroid*, dan menetapkan 'random\_state' untuk memastikan hasil yang konsisten. Fungsi 'fit\_predict()' kemudian digunakan untuk melatih model dengan data 'x' dan menghasilkan prediksi kluster untuk setiap data, yang disimpan dalam variabel 'y\_kmeans'

```

0d ▶ kmeans = KMeans(n_clusters=3, init='k-means++', random_state = 42)
    y_kmeans = kmeans.fit_predict(x)

```

Gambar 7. Inisialisasi *K-Means*.

Gambar 8 'y\_kmeans' berisi label kluster untuk setiap data dalam array 'x' untuk menunjukkan kluster mana data tersebut termasuk kemudian mengelompokkan data



```

0d y_kmeans
array([1, 1, 1, 1, 1, 2, 1, 0, 1, 1, 1, 1, 2, 2, 1, 2, 1, 1, 0, 1],
      dtype=int32)

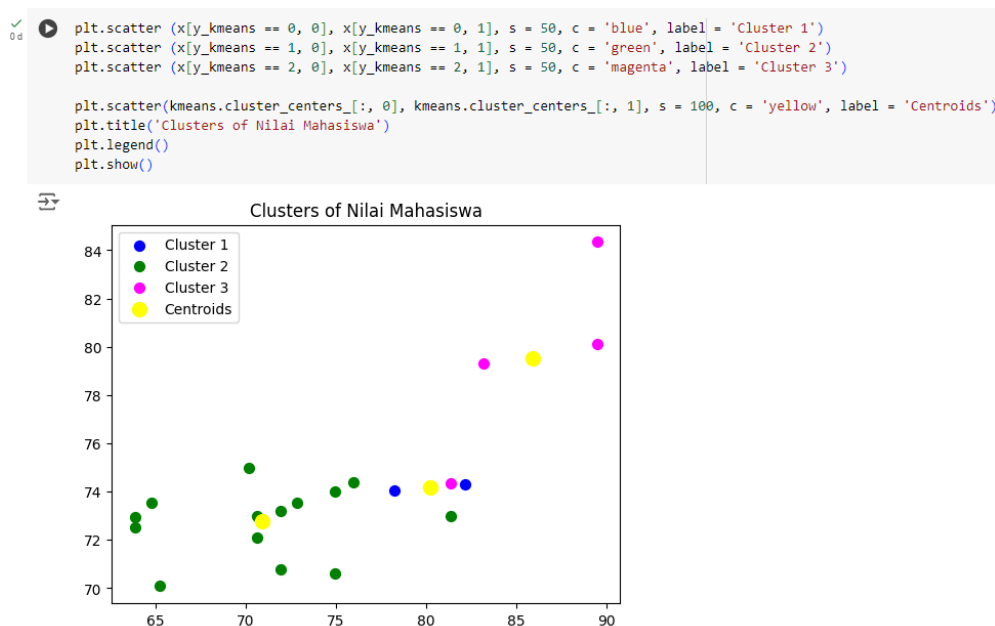
```

Gambar 8. Kluster data.

Tabel 2. Data hasil kluster nilai mahasiswa.

Mahasiswa	Kluster
A	1
B	1
C	1
D	1
E	1
F	2
G	1
H	0
I	1
J	1
K	1
L	1
M	2
N	2
O	1
P	2
Q	1
R	1
S	0
T	1

Gambar 9 menunjukkan grafik dengan tiga kluster, yang menunjukkan hasil klustering dari data nilai siswa yang digunakan menggunakan metode *K-Means*. Sumbu- $x$  pada grafik menunjukkan nilai dari kolom PLMH (sebuah fitur yang diambil dari data mentah), sedangkan sumbu- $y$  menunjukkan nilai dari kolom Kael. Data awal memiliki beberapa fitur, yaitu PLMH, Kael, PK1, Ale, PK2 dan Kalin, tetapi untuk visualisasi grafik ini hanya dua fitur pertama (PLMH dan Kael) yang digunakan agar dapat divisualisasikan dalam dua dimensi. Kluster satu ditunjukkan dengan warna biru, Kluster dua ditunjukkan dengan warna hijau dan Kluster tiga ditunjukkan dengan warna magenta. Titik-titik pada grafik mewakili mahasiswa yang dikelompokkan berdasarkan nilai mereka, sementara titik berwarna kuning menunjukkan posisi *centroid* masing-masing kluster, yang merupakan titik pusat dari kelompok tersebut.



Gambar 9. Visualisasi hasil kluster, dengan sumbu  $x$  menandakan nilai PLMH dan sumbu  $y$  menandakan nilai Kael.

## 5 Simpulan dan Saran

### 5.1 Simpulan

Pengelompokan prestasi akademik mahasiswa dilakukan menggunakan metode *K-Means* yang membagi data nilai mahasiswa ke dalam tiga kluster, yaitu tinggi, sedang, dan rendah, berdasarkan jarak terdekat ke *centroid* kluster. Proses ini diulang hingga posisi *centroid* mencapai kestabilan, memastikan pembentukan kluster yang akurat dan konsisten. Dalam konteks penelitian ini, metode *K-Means* menunjukkan kinerjanya yang mampu mengelompokkan data secara terstruktur dan objektif, mencerminkan perbedaan signifikan dalam prestasi akademik mahasiswa. Hasil klusterisasi ini memberikan gambaran yang jelas mengenai distribusi prestasi mahasiswa dan menjadi alat yang bermanfaat bagi Program Studi Matematika untuk mengidentifikasi mahasiswa berprestasi serta memantau perkembangan akademik secara lebih sistematis dan berbasis data. Dengan demikian, hasil penelitian ini diharapkan dapat mendukung pengambilan keputusan yang strategis untuk meningkatkan kualitas pendidikan.

### 5.2 Saran

Untuk penelitian lebih lanjut, metode klusterisasi ini harus diterapkan pada lebih banyak data, seperti mahasiswa dari berbagai angkatan atau faktor selain prestasi akademik, seperti kegiatan ekstrakurikuler. Selain itu, program studi dapat menggunakan hasil klusterisasi ini untuk merancang program pengembangan yang lebih spesifik dan sesuai dengan kebutuhan setiap kluster, seperti program mentoring untuk kluster dengan nilai sedang dan rendah, serta pengembangan kemampuan lanjutan bagi mahasiswa dengan nilai tinggi.

## Daftar Pustaka

- [1] Mahartika IR, Arief W. 2019. Data mining klasterisasi dengan algoritma K-Means untuk pengelompokan provinsi berdasarkan konsumsi bahan bakar Minyak Nasional. *Prosiding Seminar Nasional Sisfotek (Sistem Informasi dan Teknologi Informasi)*. 3(1):87-91.
- [2] Muliono R, Zulfikar S. 2019. Data mining klastering menggunakan algoritma K-Means untuk klasterisasi tingkat tridarma pengajaran dosen. *Journal of Computer Engineering System and Science*. 4(2):272-279.
- [3] Muslim MA, Prasetyo B, Mawarni EL, Herowati AJ, Mirqotussa'dah M, Rukmana SH, Nurzahputra A. 2019. Data mining algoritma C4. 5 disertai contoh kasus dan penerapannya dengan program computer. *Nucl. Phys*. 13(1):104-116.
- [4] Mustika, Yunita A, Abraham M, Nazaruddin A, Imanuddin H, Guntoro, Melda AM, Mohamad R, Hozairi, Anindya KW, Syariful A, Ikhsan R, Yoga R, D Tri O, Unggul US, lin E. 2021. *Data Mining dan Aplikasinya*. Bandung: WIDINA BHAKTI PERSADA BANDUNG.
- [5] Nanda WS, Akim MHP, Magdalena S. 2023. Analisis data mining untuk klasterisasi data rekam medis menggunakan algoritma K-Means pada Rumah Sakit Sylvani Binjai. *Indonesian Journal of Education and Computer Science*. 1(3):82-88.
- [6] Novia EA, Woro IR, Cahyo P. 2020. *Sistem Perbandingan K-Means dan Naïve Bayes untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan*. Bandung: Kreatif Industri Nusantara.
- [7] Wahyudi, I., Sulthan, M. B., Suhartini, L. 2021. Analisa penentuan Cluster terbaik pada metode K-Means menggunakan Elbow terhadap sentra industri produksi di Pamekasan. *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*. 2(2):72-81.