JSRCS Journal of Students' Research in Computer Science

e-ISSN: 2722 – 290X
Vol. 7 No. 1 (Mei 2026), Page: 37 – 50
https://doi.org/10.31599/kv0g1g94
Available Online at http://ejurnal.ubharajaya.ac.id/index.php/JSRCS

Analisis *Clustering* Data Mahasiswa Berdasarkan Nilai Akademik Menggunakan *K-Means*

Muhammad Aliq Aulia [1,*], Kresno Ario Tri Wibowo [1], Irfan Nugraha [1], Ilham Wahyu Analta [1]

* Korespondensi: e-mail: muhammadaliq18@gmail.com

[1] Informatika; Universitas Kusuma Husada; Jl. Jaya Wijaya No..11, Kadipiro, Kec. Banjarsari, Kota Surakarta, Jawa Tengah, (0271) 857724/856832; e-mail: muhammadaliq18@ukh.ac.id, kresnoario@ukh.ac.id, irfannugraha@ukh.ac.id, ilham.wahyu@ukh.ac.id

Submitted : **16 Maret 2026**
Revised : **7 April 2026**
Accepted : **5 Mei 2026**
Published : **30 Mei 2026**

Abstract

Academic data in higher education are mainly used for administrative purposes, rather than for meaningful insights. Yet, analyzing student grade data can reveal patterns that help institutions evaluate and improve development strategies. This study grouped student data by academic grades using the K-Means Clustering method. Grades from core courses underwent data collection, preprocessing, cluster number selection, and Clustering using K-Means. The results showed K-Means successfully clustered students by performance level. Each cluster reflected a category of academic ability: high, medium, or low. These results can help institutions monitor progress and design better academic guidance. Thus, applying K-Means may be effective for analyzing student academic data in higher education.

Keywords: *Academic Performance, Clustering, Data Mining, K-Means, Students*

Abstrak

Data akademik mahasiswa yang tersimpan pada perguruan tinggi umumnya hanya digunakan sebagai arsip administrasi dan belum dimanfaatkan secara optimal untuk memperoleh informasi yang lebih bermakna. Padahal, data nilai mahasiswa dapat dianalisis untuk mengetahui pola performa akademik sehingga dapat membantu pihak akademik dalam melakukan evaluasi dan pembinaan mahasiswa. Penelitian ini bertujuan untuk mengelompokkan data mahasiswa berdasarkan nilai akademik menggunakan metode *K-Means Clustering*. Data yang digunakan berupa nilai beberapa mata kuliah inti mahasiswa yang kemudian diproses melalui tahapan pengumpulan data, praproses data, penentuan jumlah *cluster*, serta proses pengelompokan menggunakan algoritma *K-Means*. Hasil penelitian menunjukkan bahwa metode *K-Means* mampu mengelompokkan mahasiswa ke dalam beberapa kelompok berdasarkan tingkat performa akademiknya. Setiap *cluster* merepresentasikan kategori performa akademik mahasiswa yang berbeda, seperti kategori tinggi, sedang, dan rendah. Hasil pengelompokan ini dapat memberikan gambaran bagi pihak akademik dalam melakukan pemantauan perkembangan mahasiswa serta menyusun strategi pembinaan yang lebih tepat. Dengan demikian, penerapan teknik data mining menggunakan metode *K-Means* dapat menjadi salah satu pendekatan yang efektif dalam menganalisis data akademik mahasiswa di lingkungan perguruan tinggi.

Kata kunci: Akademik, *Clustering*, Data Mining, *K-Means*, Nilai Mahasiswa

JSRCS is a nationally accredited journal ranked SINTA 5 by the Directorate General of Research Strengthening and Development under Decree No. 225/E/KPT/2022, covering publications from Volume 1, Issue 2 (2020) to Volume 6, Issue 1 (2025). The journal was re-accredited under Decree No. 295/C/C3/KPT/2026 and retains its SINTA 5 status for the period from Volume 6, Issue 2 (2025) through Volume 11, Issue 1 (2030).

1. Pendahuluan

Perkembangan teknologi informasi telah menghasilkan pertumbuhan data yang sangat besar di berbagai bidang, termasuk dalam lingkungan pendidikan tinggi. Perguruan tinggi saat ini menyimpan berbagai jenis data akademik mahasiswa, seperti nilai mata kuliah, indeks prestasi, serta riwayat akademik lainnya. Data tersebut umumnya hanya dimanfaatkan sebagai arsip administratif atau laporan akademik, sehingga potensi informasi yang terkandung di dalamnya belum dimanfaatkan secara optimal. Padahal, data akademik mahasiswa dapat dianalisis lebih lanjut untuk menemukan pola tertentu yang dapat digunakan sebagai dasar dalam pengambilan keputusan akademik. Salah satu pendekatan yang dapat digunakan untuk menganalisis data dalam jumlah besar adalah teknik *data mining*. *Data mining* merupakan proses penggalian informasi atau pengetahuan baru dari kumpulan data yang besar dengan menggunakan metode statistik, matematika, dan pembelajaran mesin. Dalam bidang pendidikan, penerapan *data mining* dapat dimanfaatkan untuk menganalisis performa akademik mahasiswa, mengidentifikasi pola pembelajaran, serta membantu institusi pendidikan dalam meningkatkan kualitas proses pembelajaran. Penerapan educational data mining dalam bidang pendidikan telah banyak digunakan untuk menganalisis performa akademik mahasiswa, memprediksi hasil pembelajaran, serta membantu institusi pendidikan dalam pengambilan keputusan berbasis data. Menurut (Dutt et al., 2017), educational data mining mampu membantu institusi pendidikan dalam memahami pola akademik mahasiswa secara lebih efektif melalui proses analisis data yang sistematis. Selain itu, (Witten et al., 2016) menjelaskan bahwa teknik *Clustering* dapat digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik tertentu sehingga mempermudah proses identifikasi pola dalam data akademik.

Salah satu teknik dalam *data mining* yang sering digunakan adalah metode *Clustering*. *Clustering* merupakan teknik pengelompokan data yang bertujuan untuk mengelompokkan objek berdasarkan kemiripan karakteristik yang dimiliki sehingga objek yang berada dalam satu kelompok memiliki tingkat kemiripan yang signifikan dibandingkan dengan objek pada kelompok lainnya. Metode ini banyak digunakan dalam analisis data pendidikan untuk mengidentifikasi pola performa akademik mahasiswa. *Algoritma K-Means* merupakan salah satu metode *Clustering* yang banyak digunakan karena memiliki konsep yang sederhana serta mampu mengelompokkan data secara efektif berdasarkan kedekatan jarak antara data dengan pusat kelompok (*centroid*). Beberapa penelitian sebelumnya menunjukkan bahwa metode *K-Means* dapat digunakan untuk mengelompokkan data akademik siswa maupun mahasiswa berdasarkan tingkat performa akademiknya. Penelitian yang dilakukan oleh (Pamungkas, 2024) menunjukkan bahwa algoritma *K-Means* mampu mengelompokkan data nilai mahasiswa sehingga dapat membantu dalam analisis performa akademik. Penelitian lain yang dilakukan oleh (Akram et al., 2024) juga menunjukkan bahwa metode *Clustering* menggunakan *K-Means* dapat digunakan untuk memberikan rekomendasi pengelompokan kelas berdasarkan performa akademik siswa. Penelitian lain yang dilakukan oleh Kurniawan dan Nugroho (2021) menunjukkan bahwa algoritma *K-Means* mampu digunakan untuk menganalisis performa akademik mahasiswa

melalui proses pengelompokan data berdasarkan tingkat kemiripan nilai akademik. Selain itu, Bunkers dan Miller (2020) menyatakan bahwa penerapan *Clustering* pada data pendidikan dapat membantu institusi akademik dalam melakukan evaluasi performa belajar mahasiswa secara lebih terstruktur.

Selain itu, penelitian yang dilakukan oleh (Sujana, 2025) menunjukkan bahwa penerapan teknik data mining menggunakan algoritma *K-Means* dapat membantu dalam mengidentifikasi kelompok mahasiswa berdasarkan performa akademik yang dimiliki. Penelitian lainnya oleh (Alalawi et al., 2023) juga menyatakan bahwa metode *K-Means* efektif digunakan untuk mengelompokkan data performa mahasiswa sehingga memudahkan institusi pendidikan dalam melakukan analisis data akademik. Selanjutnya, penelitian yang dilakukan oleh (Chairunisa et al., 2025) menunjukkan bahwa metode *Clustering* mampu memberikan gambaran mengenai distribusi performa akademik siswa berdasarkan hasil pengelompokan nilai. Berdasarkan beberapa penelitian sebelumnya, dapat diketahui bahwa metode *K-Means* memiliki potensi yang besar untuk digunakan dalam analisis data akademik. Namun demikian, pemanfaatan teknik data mining khususnya metode *Clustering* dalam pengolahan data akademik mahasiswa masih belum banyak diterapkan secara optimal di berbagai perguruan tinggi. Banyak institusi pendidikan yang masih memanfaatkan data akademik hanya sebagai data administratif tanpa dilakukan analisis lebih lanjut untuk memperoleh informasi yang lebih bermakna.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk melakukan analisis *Clustering* terhadap data nilai akademik mahasiswa menggunakan algoritma *K-Means*. Melalui proses pengelompokan ini diharapkan dapat diperoleh kelompok mahasiswa berdasarkan tingkat performa akademiknya, seperti kategori tinggi, sedang, dan rendah. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai distribusi performa akademik mahasiswa serta membantu pihak akademik dalam menyusun strategi pembinaan dan evaluasi pembelajaran yang lebih tepat. Nilai kebaruan dari penelitian ini terletak pada penerapan metode *K-Means* untuk menganalisis data nilai akademik mahasiswa sebagai dasar dalam pengelompokan performa akademik secara sistematis. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pemanfaatan teknik data mining di bidang pendidikan, khususnya dalam pengolahan data akademik mahasiswa di perguruan tinggi.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan data mining dengan metode *Clustering* menggunakan algoritma *K-Means* untuk mengelompokkan data akademik mahasiswa berdasarkan nilai beberapa mata kuliah. Metode *K-Means* dipilih karena memiliki konsep yang sederhana serta mampu mengelompokkan data secara efektif berdasarkan tingkat kemiripan karakteristik data (Han et al., 2022). Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *K-Means* mampu mengelompokkan data performa akademik mahasiswa secara efisien sehingga dapat membantu institusi pendidikan dalam melakukan analisis performa belajar mahasiswa (Jain, 2021). Selain itu, metode *Clustering* juga banyak digunakan dalam analisis data

pendidikan untuk mengidentifikasi pola belajar dan karakteristik kelompok mahasiswa (Aggarwal & Reddy, 2019). Konsep dasar data mining dan knowledge discovery merupakan proses *Clustering* merupakan salah satu teknik penting dalam pengelompokan data berdasarkan kemiripan atribut tertentu mahasiswa (Maimon & Rokach, 2010). Metode *Clustering* dapat digunakan untuk menemukan pola tersembunyi dalam kumpulan data berukuran besar (Larose, 2014). Dalam implementasinya, teknik *Clustering* menggunakan algoritma *K-Means* banyak digunakan karena memiliki proses perhitungan yang relatif sederhana dan efisien dalam pengolahan data multidimensi (Prasetyo, 2014).

2.1. Desain Penelitian

Desain penelitian yang digunakan dalam penelitian ini terdiri dari beberapa tahapan utama:

a. Pengumpulan Data

   Tahap ini dilakukan dengan mengumpulkan data nilai akademik mahasiswa dari beberapa mata kuliah yang dijadikan sebagai atribut dalam proses *Clustering*.

b. Pra-pemrosesan Data

   Tahap ini dilakukan untuk membersihkan data (data cleaning) serta memastikan data yang digunakan dalam penelitian memiliki kualitas yang baik sehingga dapat diproses lebih lanjut.

c. Penentuan Jumlah *Cluster*

   Pada penelitian ini jumlah *cluster* ditentukan sebanyak tiga kelompok yang merepresentasikan kategori performa akademik mahasiswa, yaitu kategori tinggi, sedang, dan rendah.

d. Proses *Clustering* Menggunakan *K-Means*

   Data yang telah diproses kemudian dianalisis menggunakan algoritma *K-Means* untuk mengelompokkan data mahasiswa berdasarkan tingkat kemiripan nilai akademiknya.

e. Analisis Hasil *Clustering*

   Hasil pengelompokan data kemudian dianalisis untuk mengetahui distribusi performa akademik mahasiswa pada setiap *cluster* yang terbentuk.

2.2. Akuisisi Data

Data yang digunakan dalam penelitian ini berupa data nilai akademik mahasiswa yang diperoleh dari data akademik mahasiswa yang dianonimkan untuk keperluan penelitian yang merepresentasikan nilai beberapa mata kuliah inti pada program studi informatika. *Dataset* yang digunakan terdiri dari 50 data mahasiswa dengan beberapa atribut nilai mata kuliah yang digunakan sebagai parameter dalam proses *Clustering*. Atribut yang digunakan dalam penelitian ini meliputi nilai mata kuliah Algoritma, Basis Data, Jaringan Komputer, dan Kecerdasan Buatan. Setiap atribut memiliki rentang nilai antara 0 hingga 100 yang menggambarkan performa akademik mahasiswa pada masing-masing mata kuliah. Data tersebut kemudian diproses dan dianalisis menggunakan algoritma *K-Means* untuk mengelompokkan mahasiswa berdasarkan tingkat kemiripan nilai akademik yang dimiliki. Hasil pengelompokan ini diharapkan dapat

memberikan gambaran mengenai kategori performa akademik mahasiswa. Untuk memberikan gambaran mengenai dataset yang digunakan, data mahasiswa ditampilkan pada Tabel 1.

Tabel 1. *Dataset* Nilai Akademik Mahasiswa

No	Algoritma	Basis Data	Jaringan Komputer	Kecerdasan Buatan
1	85	88	80	90
2	78	82	75	80
3	92	90	88	94
4	65	70	68	72
5	74	76	72	78
6	88	91	85	89
7	69	72	70	68
8	90	92	89	93
9	76	79	74	77
10	82	85	80	84
11	67	69	65	70
12	91	94	90	92
13	73	75	70	74
14	80	83	78	82
15	87	89	84	88
16	62	65	60	64
17	79	81	76	80
18	93	95	91	94
19	71	74	70	72
20	84	86	82	85
21	77	78	75	79
22	90	91	88	92
23	66	68	65	69
24	81	84	79	83
25	75	77	72	76
26	89	90	87	91
27	70	73	68	71
28	83	85	80	86
29	72	74	70	73
30	94	93	91	95
31	76	78	74	77
32	85	87	83	88
33	68	70	66	69
34	91	92	90	93
35	74	76	72	75
36	80	82	78	81
37	87	89	86	90

No	Algoritma	Basis Data	Jaringan Komputer	Kecerdasan Buatan
38	63	66	62	65
39	79	80	77	81
40	92	94	89	93
41	71	73	69	72
42	84	86	82	85
43	75	77	74	76
44	88	90	85	89
45	67	69	65	68
46	90	91	88	92
47	73	75	71	74
48	82	84	80	83
49	78	79	76	80
50	95	93	92	94

Sumber : Hasil Penelitian (2026)

Dataset yang digunakan dalam penelitian ini terdiri dari 50 data mahasiswa dengan empat atribut nilai mata kuliah yaitu Algoritma, Basis Data, Jaringan Komputer, dan Kecerdasan Buatan. Data tersebut kemudian digunakan sebagai input dalam proses pengelompokan menggunakan algoritma *K-Means*.

2.3. Prosedur Algoritma *K-Means*

Algoritma *K-Means* merupakan salah satu metode *Clustering* yang digunakan untuk mengelompokkan data berdasarkan kedekatan karakteristik yang dimiliki. Metode ini bekerja dengan cara menentukan pusat *cluster* (centroid) dan mengelompokkan data berdasarkan jarak terdekat terhadap centroid tersebut.

Langkah-langkah algoritma *K-Means* dalam penelitian ini adalah sebagai berikut:

a. Menentukan jumlah *cluster* (k).
b. Menentukan centroid awal secara acak.
c. Menghitung jarak setiap data terhadap centroid.
d. Mengelompokkan data ke dalam *cluster* yang memiliki jarak terdekat.
e. Menghitung ulang centroid baru dari setiap *cluster*.
f. Mengulangi proses hingga posisi centroid tidak berubah atau mencapai kondisi konvergen.

Perhitungan jarak antara data dan centroid dilakukan menggunakan metode Euclidean Distance sebagai berikut:

$$d(x,y) = \sqrt{\{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2\}} \quad (1)$$

Keterangan:

a. d(x,y) = jarak antara data dan centroid
b. $x_1, x_2, \ldots, x_n$ = nilai atribut dari data mahasiswa
c. $y_1, y_2, \ldots, y_n$ = nilai atribut dari centroid
d. n = jumlah atribut yang digunakan dalam proses *Clustering*

Selain menggunakan perhitungan jarak menggunakan *Euclidean Distance*, proses pada algoritma *K-Means* juga melibatkan perhitungan *centroid* baru pada setiap iterasi. *Centroid* merupakan titik pusat dari suatu *cluster* yang diperoleh dari rata-rata nilai seluruh data yang berada dalam *cluster* tersebut.

Perhitungan centroid baru dapat dirumuskan sebagai berikut:

$$C_k = \left(\frac{1}{n}\right) \Sigma\, xi \tag{2}$$

Keterangan:

a. Ck = centroid pada *cluster* ke-k
b. xi = nilai data ke-i yang berada dalam *cluster*
c. n = jumlah data pada *cluster*

Rumus tersebut digunakan untuk menghitung nilai pusat *cluster* baru setelah seluruh data dikelompokkan berdasarkan jarak terdekat terhadap centroid sebelumnya. Proses ini dilakukan secara berulang hingga nilai centroid tidak mengalami perubahan yang signifikan.

Dalam penelitian ini, proses *Clustering* dilakukan dengan menentukan jumlah *cluster* sebanyak 3 *cluster*, yaitu:

Cluster 1 : Mahasiswa dengan performa akademik tinggi

Cluster 2 : Mahasiswa dengan performa akademik sedang

Cluster 3 : Mahasiswa dengan performa akademik rendah

Pengelompokan ini bertujuan untuk mengetahui pola performa akademik mahasiswa berdasarkan kemiripan nilai pada beberapa mata kuliah yang dianalisis.

2.4 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang sistematis untuk menghasilkan proses pengelompokan data yang optimal. Tahapan penelitian dimulai dari pengumpulan data hingga analisis hasil *Clustering* menggunakan algoritma *K-Means*.

Adapun tahapan penelitian yang dilakukan adalah sebagai berikut:

a. Pengumpulan Data

   Tahap awal penelitian adalah mengumpulkan dataset nilai akademik mahasiswa yang terdiri dari beberapa atribut mata kuliah, yaitu Algoritma, Basis Data, Jaringan Komputer, dan Kecerdasan Buatan.

b. Pra-pemrosesan Data

   Data yang telah diperoleh kemudian dipersiapkan sebelum proses *Clustering* dilakukan. Tahapan ini meliputi pengecekan kelengkapan data serta penyusunan data agar dapat digunakan sebagai input pada algoritma *K-Means*.

c. Penentuan Jumlah *Cluster*

Pada penelitian ini jumlah *cluster* yang digunakan adalah sebanyak tiga *cluster* yang merepresentasikan kategori performa akademik mahasiswa, yaitu tinggi, sedang, dan rendah.

d. Proses *Clustering* Menggunakan *K-Means*

Dataset yang telah dipersiapkan kemudian diproses menggunakan algoritma *K-Means* dengan melakukan perhitungan jarak antara data dan centroid menggunakan metode Euclidean Distance.

e. Analisis Hasil *Clustering*

Hasil pengelompokan data mahasiswa kemudian dianalisis untuk mengetahui pola performa akademik mahasiswa berdasarkan nilai mata kuliah yang digunakan dalam penelitian.

3. Hasil dan Pembahasan

3.1. *Dataset* Penelitian

Dataset yang digunakan dalam penelitian ini terdiri dari 50 data mahasiswa yang memiliki empat atribut nilai mata kuliah, yaitu Algoritma, Basis Data, Jaringan Komputer, dan Kecerdasan Buatan. Setiap atribut memiliki rentang nilai antara 0–100 yang merepresentasikan performa akademik mahasiswa pada masing-masing mata kuliah. *Dataset* tersebut digunakan sebagai input dalam proses *Clustering* menggunakan algoritma *K-Means* untuk mengelompokkan mahasiswa berdasarkan tingkat kemiripan nilai akademik yang dimiliki.

Sebagian contoh dataset yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2.

Tabel 2. *Dataset* Nilai Akademik Mahasiswa

No	Algoritma	Basis Data	Jaringan Komputer	Kecerdasan Buatan
1	85	88	80	90
2	78	82	75	80
3	92	90	88	94
4	65	70	68	72
5	74	76	72	78
…	…	…	…	…
…	…	…	…	…
50	95	93	92	94

Sumber : Hasil Penelitian (2026)

3.2. Penentuan *Centroid* Awal

Pada algoritma *K-Means*, langkah awal yang dilakukan adalah menentukan jumlah *cluster* serta menentukan nilai centroid awal. Dalam penelitian ini jumlah *cluster* yang digunakan adalah k = 3, yang merepresentasikan kategori performa akademik mahasiswa yaitu:

Cluster 1 : Mahasiswa dengan performa akademik tinggi

Cluster 2 : Mahasiswa dengan performa akademik sedang

Cluster 3 : Mahasiswa dengan performa akademik rendah

Centroid awal ditentukan secara acak dari data mahasiswa yang tersedia. Nilai centroid awal yang digunakan dalam penelitian ini ditunjukkan pada Tabel 3.

Tabel 3. Data *Centroid* Awal

Cluster	Algoritma	Basis Data	Jaringan Komputer	Kecerdasan Buatan
C1	90	92	88	91
C2	78	82	76	79
C3	65	68	64	66

Sumber : Hasil Penelitian (2026)

Centroid tersebut kemudian digunakan sebagai acuan dalam menghitung jarak antara setiap data mahasiswa dengan pusat *cluster* menggunakan metode *Euclidean Distance*.

3.3. Proses Iterasi *K-Means*

Setelah centroid awal ditentukan, langkah berikutnya adalah menghitung jarak setiap data mahasiswa terhadap masing-masing centroid menggunakan rumus *Euclidean Distance*. Sebagai contoh perhitungan jarak pada data mahasiswa pertama dengan nilai:

a. Algoritma = 85
b. Basis Data = 88
c. Jaringan = 80
d. Kecerdasan Buatan = 90
e. terhadap centroid **C1** adalah sebagai berikut:
   $d = \sqrt{((85-90)^2 + (88-92)^2 + (80-88)^2 + (90-91)^2)}$
   $d = \sqrt{(25 + 16 + 64 + 1)}$
   $d = \sqrt{106}$
   $d = 10.29$
f. terhadap centroid **C2** adalah sebagai berikut:
   $d = \sqrt{((85-78)^2 + (88-80)^2 + (80-76)^2 + (90-79)^2)}$
   $d = \sqrt{(49 + 64 + 16 + 121)}$
   $d = \sqrt{250}$
   $d = 15.81$
g. terhadap centroid **C3** adalah sebagai berikut:
   $d = \sqrt{((85-65)^2 + (88-68)^2 + (80-64)^2 + (90-66)^2)}$
   $d = \sqrt{(400 + 400 + 256 + 576)}$
   $d = \sqrt{1632}$
   $d = 40.40$

Perhitungan yang sama dilakukan terhadap centroid C2 dan C3. Data mahasiswa kemudian dimasukkan ke dalam *cluster* yang memiliki jarak paling kecil. Proses ini dilakukan pada seluruh dataset hingga semua data mahasiswa memperoleh kelompok *cluster* masing-masing.

3.4. Perhitungan Pembaruan *Centroid*

Setelah data mahasiswa dikelompokkan ke dalam *cluster* berdasarkan jarak terdekat terhadap centroid, langkah berikutnya adalah menghitung centroid baru untuk setiap *cluster*. *Centroid* baru diperoleh dengan menghitung nilai rata-rata dari seluruh data yang berada pada *cluster* tersebut. Sebagai contoh, misalkan pada iterasi pertama *Cluster* C1 berisi data mahasiswa:

Tabel 4. Hasil *Clustering* Mahasiswa

No	Algoritma	Basis Data	Jaringan Komputer	Kecerdasan Buatan
1	85	88	80	90
3	92	90	88	94
6	88	91	85	89

Sumber : Hasil Penelitian (2026)

Maka centroid baru dihitung sebagai berikut.

Algoritma

C1 = (85 + 92 + 88) / 3

C1 = 88.33

Basis Data

C1 = (88 + 90 + 91) / 3

C1 = 89.67

Jaringan Komputer

C1 = (80 + 88 + 85) / 3

C1 = 84.33

Kecerdasan Buatan

C1 = (90 + 94 + 89) / 3

C1 = 91

Sehingga centroid baru untuk *cluster* **C1** adalah:

C1 = (88.33, 89.67, 84.33, 91)

Centroid baru ini kemudian digunakan kembali pada iterasi berikutnya hingga nilai centroid tidak mengalami perubahan yang signifikan.

3.5. Hasil Pengelompokan *Cluster*

Setelah proses iterasi dilakukan beberapa kali, centroid akan mengalami perubahan hingga mencapai kondisi konvergen, yaitu kondisi ketika nilai centroid tidak mengalami perubahan signifikan. Hasil akhir proses *Clustering* menggunakan algoritma *K-Means* menghasilkan pengelompokan data mahasiswa ke dalam tiga *cluster* yang ditunjukkan pada Tabel 5.

Cluster C1 terdiri dari 17 mahasiswa yang termasuk dalam kategori performa akademik tinggi. *Cluster* C2 merupakan kelompok dengan jumlah anggota terbanyak, yaitu 21 mahasiswa, yang dikategorikan sebagai mahasiswa dengan performa akademik sedang. *Cluster* C3 terdiri dari 12 mahasiswa yang termasuk dalam kategori performa akademik rendah.

Tabel 5. Hasil *Clustering* Mahasiswa

Cluster	Jumlah Mahasiswa	Kategori
C1	17	Performa Akademik Tinggi
C2	21	Performa Akademik Sedang
C3	12	Performa Akademik Rendah

Sumber : Hasil Penelitian (2026)

Secara keseluruhan, hasil pengelompokan menggunakan algoritma K-Means berhasil mengidentifikasi pola performa akademik mahasiswa berdasarkan karakteristik data yang dimiliki.

3.6 *Centroid* Akhir Hasil *Clustering*

Setelah proses iterasi algoritma *K-Means* dilakukan beberapa kali, nilai centroid akan mengalami perubahan hingga mencapai kondisi konvergen, yaitu kondisi ketika nilai centroid tidak lagi mengalami perubahan yang signifikan. Nilai centroid akhir menunjukkan karakteristik rata-rata dari setiap *cluster* yang terbentuk. *Centroid* akhir yang diperoleh dari proses *Clustering* pada penelitian ini dapat dilihat pada Tabel 6.

Tabel 6. *Centroid* Akhir Hasil *Clustering*

Cluster	Algoritma	Basis Data	Jaringan Komputer	Kecerdasan Buatan
C1	90.6	92.1	88.5	91.7
C2	79.4	81.2	77.3	80.1
C3	67.3	69.5	65.8	68.4

Sumber : Hasil Penelitian (2026)

Berdasarkan nilai centroid akhir tersebut dapat diketahui bahwa *cluster* pertama memiliki nilai rata-rata tertinggi pada seluruh atribut, sedangkan *cluster* ketiga memiliki nilai rata-rata paling rendah. Hal ini menunjukkan bahwa hasil pengelompokan yang dihasilkan oleh algoritma *K-Means* mampu memisahkan data mahasiswa berdasarkan tingkat performa akademik yang dimiliki.

3.7 Analisis Hasil *Cluster*

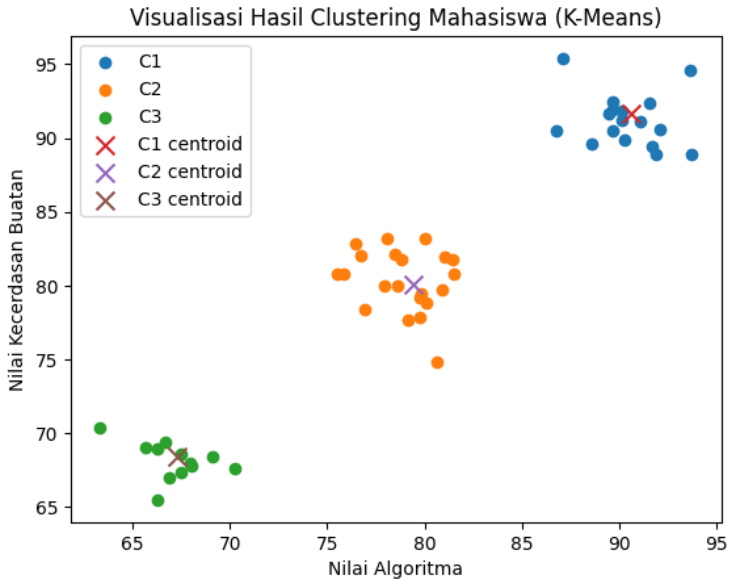
Berdasarkan hasil proses *Clustering* yang dilakukan menggunakan algoritma *K-Means*, diperoleh tiga kelompok mahasiswa yang memiliki karakteristik nilai akademik yang berbeda.

1. *Cluster* Pertama menunjukkan kelompok mahasiswa dengan nilai rata-rata yang tinggi pada seluruh mata kuliah yang dianalisis. Mahasiswa dalam kelompok ini dapat dikategorikan sebagai mahasiswa dengan performa akademik yang sangat baik.
2. *Cluster* Kedua merupakan kelompok mahasiswa dengan nilai akademik pada kategori sedang. Kelompok ini menunjukkan bahwa mahasiswa memiliki performa yang cukup baik namun masih terdapat beberapa mata kuliah yang perlu ditingkatkan.
3. *Cluster* Ketiga merupakan kelompok mahasiswa dengan nilai akademik yang relatif lebih rendah dibandingkan dengan *cluster* lainnya. Hasil pengelompokan ini dapat digunakan sebagai bahan evaluasi bagi pihak program studi untuk memberikan perhatian atau strategi pembelajaran yang lebih tepat bagi mahasiswa pada kategori ini.

Dengan demikian, penerapan algoritma *K-Means* dalam penelitian ini dapat membantu mengidentifikasi pola performa akademik mahasiswa berdasarkan kemiripan nilai yang dimiliki.

3.8 Visualisasi Hasil *Clustering* Mahasiswa

Visualisasi hasil *Clustering* menggunakan algoritma *K-Means* ditunjukkan pada Gambar 3. Grafik scatter plot tersebut menggambarkan distribusi data mahasiswa berdasarkan dua atribut nilai, yaitu nilai mata kuliah Algoritma dan Kecerdasan Buatan. Setiap titik pada grafik merepresentasikan satu data mahasiswa, sedangkan tanda centroid menunjukkan pusat *cluster* yang terbentuk. Berdasarkan hasil visualisasi, terlihat bahwa data mahasiswa terbagi menjadi tiga kelompok yang berbeda. *Cluster* pertama berada pada area nilai yang tinggi, yang menunjukkan kelompok mahasiswa dengan performa akademik tinggi. *Cluster* kedua berada pada area nilai sedang yang menunjukkan kelompok mahasiswa dengan performa akademik sedang. Sementara itu, *cluster* ketiga berada pada area nilai yang lebih rendah yang menunjukkan kelompok mahasiswa dengan performa akademik rendah.



Sumber : Hasil Penelitian (2026)

Gambar 3. Visualisasi Hasil *Clustering*

Visualisasi ini menunjukkan bahwa algoritma *K-Means* mampu memisahkan data mahasiswa ke dalam kelompok yang memiliki karakteristik nilai yang relatif homogen dalam setiap *cluster*.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, penerapan algoritma *K-Means* berhasil digunakan untuk mengelompokkan data mahasiswa berdasarkan kemiripan nilai akademik pada mata kuliah Algoritma, Basis Data, Jaringan Komputer, dan Kecerdasan Buatan. Proses *Clustering* dilakukan menggunakan dataset sebanyak 50 data mahasiswa dengan jumlah *cluster* sebanyak tiga kelompok, yaitu *cluster* performa akademik tinggi, sedang, dan rendah. Hasil pengolahan data menunjukkan bahwa sebanyak 17 mahasiswa berada pada *cluster* performa akademik tinggi, 21 mahasiswa berada pada *cluster* performa akademik sedang, dan 12 mahasiswa berada pada *cluster* performa akademik rendah. Pengelompokan ini menunjukkan

bahwa algoritma *K-Means* mampu mengidentifikasi pola performa akademik mahasiswa berdasarkan kemiripan karakteristik nilai yang dimiliki. Hasil penelitian ini dapat digunakan sebagai bahan evaluasi bagi pihak program studi dalam melakukan pemantauan perkembangan akademik mahasiswa serta sebagai dasar dalam menentukan strategi pembelajaran yang lebih tepat. Untuk penelitian selanjutnya, metode *Clustering* ini dapat dikembangkan dengan menggunakan jumlah dataset yang lebih besar, penambahan atribut akademik lainnya, serta membandingkan algoritma *Clustering* lain seperti DBSCAN atau Hierarchical *Clustering* untuk memperoleh hasil pengelompokan yang lebih optimal.

Ucapan Terima Kasih

Peneliti mengucapkan terima kasih kepada Program Studi Informatika Universitas Kusuma Husada Surakarta yang telah memberikan dukungan dalam pelaksanaan penelitian ini. Peneliti juga mengucapkan terima kasih kepada seluruh pihak yang telah membantu dalam proses pengumpulan data serta penyusunan penelitian ini sehingga penelitian dapat diselesaikan dengan baik.

Daftar Pustaka

Aggarwal, C., & Reddy, C. (2019). Data *Clustering*: Algorithms and Applications. CRC Press

Akram, A., Risal, N., Maryani, D., Fadillah, N., Alviadi, A., & Id, A. (2024). Implementasi *K-Means Clustering* untuk rekomendasi kelas unggulan di SMK 1 Teknologi dan Rekayasa Mimika. *JESSI (Journal of Embedded System Security and Intelligent System),* 5(3).

Alalawi, S. J., Shaharanee, I. N., & Jamil, J. (2023). *Clustering* Student Performance Data Using *K-Means* Algorithms. *Journal of Computational Innovation and Analytics (JCIA),* 2(1), 41–55. https://doi.org/10.32890/jcia2023.2.1.3

Azzahra, S. C. (2025). *Clustering* Of High School Students Academic Scores Using *K-Means* Algorithm. Journal of Information Systems and Informatics. *Journal of Information Systems and Informatics,* 7(1), 572–586. https://doi.org/10.51519/journalisi.v7i1.1029

Bunkers, M., & Miller, J. (2020). Educational Data Mining And Student Performance Analysis Using *Clustering* Techniques. *International Journal of Educational Technology*, 15(2), 45–56.

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review On Educational Data Mining. *IEEE Access*, 5, 15991–16005. https://doi.org/10.1109/ACCESS.2017.2654247

Han, J., Kamber, M., & Pei, J. (2022). *Data Mining: Concepts and Techniques (4th ed.) Morgan Kaufmann.* 585–631.

Jain, A. (2021). Data *Clustering*: 50 Years Beyond *K-Means*. In *King-Sun Fu Prize Lecture At The 19th International Conference On Pattern Recognition,* 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Kurniawan, A., & Nugroho, Y. (2021). Implementation of *K-Means Clustering* for student academic performance analysis. *Journal of Information Systems Engineering and Business Intelligence*, 7(1), 12–20.

Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons.

Maimon, O., & Rokach, L. (2010). Data Mining And Knowledge Discovery Handbook. Springer.

Pamungkas, L., D. N. A., & P. N. A. (2024). Implementation of *K-Means Clustering* Algorithm For Grouping Student Academic Performance Data. *Jurnal Sisfokom (Sistem Informasi Dan Komputer), 13(1), 45–52.*

Prasetyo, E. (2014). Data mining: Konsep dan aplikasi menggunakan MATLAB. Andi Publisher.

Sujana, T., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). Implementation Of Data Mining Using The *K-Means* Algorithm To Group Students Based On Academic Performance. *Journal of Artificial Intelligence and Engineering Applications (JAIEA),* 4(2), 1527–1531. https://doi.org/10.59934/jaiea.v4i2.936

Witten, I. H., Frank, E., & Hall, M. A. (2016). Data mining: Practical machine learning tools and techniques. Morgan Kaufmann.