EJECTS: Journal Computer, Technology and Informations System
Vol.6 No.1 September 2026, P-ISSN: 2988-5477 E-ISSN: 2988-4446
LPPM Universitas Darwan Ali

Eksplorasi Algoritma Clustering: Penerapan Algoritma K-Medoid (PAM) Untuk Mengelompokkan Mahasiswa Berdasarkan Nilai Akademik

Tarbiyatun Nisa[1], Eka Rahmawati[2]

[12]*Program Studi : Sistem Informasi, Universitas Darwan Ali*

Email : tarbiyatunnisa0@gmail.com[1], rahmawatieka062@gmail.com[2]

ABSTRACT— This study aims to explore the implementation of the K-Medoid (Partitioning Around Medoids/PAM) algorithm to cluster students based on their academic performance, including assignment, midterm, and final examination scores. The dataset used in this study is the Student Performance Dataset from the UCI Machine Learning Repository. Twenty records were selected as representative samples, and the original grading scale was transformed to a 0–100 scale to match the Indonesian higher education grading system. The research process included manual calculation of the K-Medoid algorithm using the Manhattan Distance metric, implementation in Python using Google Colab with the scikit-learn-extra library, and cluster evaluation using the Silhouette Score. The results indicate that the K-Medoid algorithm successfully grouped students into three clusters representing high, medium, and low academic performance. Since the algorithm uses actual data points as cluster centers (medoids), it is more robust against outliers than the K-Means algorithm, resulting in more representative clustering outcomes. The 3D visualization also demonstrates a clear separation among clusters. Therefore, the K-Medoid algorithm can be considered an effective method for assisting lecturers and educational institutions in analyzing students' academic performance to support educational decision-making.

Keywords— K-Medoid, Partitioning Around Medoids (PAM), Clustering, Data Mining, Student Performance, Academic Performance, Manhattan Distance.

ABSTRAK— Penelitian ini bertujuan untuk mengeksplorasi penerapan algoritma K-Medoid (Partitioning Around Medoids/PAM) dalam mengelompokkan mahasiswa berdasarkan nilai akademik yang meliputi nilai Tugas, UTS, dan UAS. Dataset yang digunakan merupakan Student Performance Dataset dari UCI Machine Learning Repository, dengan mengambil 20 data sebagai sampel dan menyesuaikan skala penilaian menjadi 0–100 agar sesuai dengan sistem penilaian perguruan tinggi di Indonesia. Proses penelitian meliputi perhitungan manual algoritma K-Medoid menggunakan Manhattan Distance, implementasi menggunakan bahasa pemrograman Python pada Google Colab dengan library scikit-learn-extra, serta evaluasi hasil clustering menggunakan Silhouette Score. Hasil penelitian menunjukkan bahwa algoritma K-Medoid berhasil membentuk tiga kelompok mahasiswa berdasarkan tingkat prestasi akademik, yaitu kelompok prestasi tinggi, prestasi menengah, dan prestasi rendah. Penggunaan medoid sebagai pusat cluster menjadikan algoritma ini lebih robust terhadap data pencilan (outlier) dibandingkan algoritma K-Means, sehingga mampu menghasilkan pengelompokan yang lebih representatif. Visualisasi hasil clustering dalam bentuk grafik 3D juga menunjukkan pemisahan antarcluster yang cukup jelas. Dengan demikian, algoritma K-Medoid dapat menjadi salah satu metode yang efektif dalam membantu dosen atau institusi pendidikan melakukan analisis performa akademik mahasiswa sebagai dasar pengambilan keputusan dalam proses pembelajaran.

Kata Kunci— K-Medoid, Partitioning Around Medoids (PAM), Clustering, Data Mining, Student Performance, Nilai Akademik, Manhattan Distance.

I. PENDAHULUAN

Perkembangan teknologi informasi dan data science telah memberikan kontribusi yang signifikan dalam berbagai bidang, termasuk bidang pendidikan. Salah satu penerapan data science adalah proses data mining yang mampu menggali informasi tersembunyi dari sekumpulan data untuk mendukung pengambilan keputusan secara lebih efektif. Dalam dunia pendidikan, analisis terhadap data akademik mahasiswa menjadi penting karena dapat membantu dosen maupun institusi pendidikan memahami karakteristik kemampuan mahasiswa sehingga strategi pembelajaran dapat disesuaikan dengan kebutuhan masing-masing kelompok [3].

Salah satu teknik yang banyak digunakan dalam data mining adalah clustering, yaitu proses mengelompokkan data berdasarkan tingkat kemiripan karakteristik tanpa menggunakan label kelas sebelumnya (unsupervised learning). Teknik clustering banyak diterapkan dalam berbagai bidang, seperti kesehatan, bisnis, pemasaran, hingga pendidikan. Pada penelitian ini, metode clustering yang digunakan adalah algoritma K-Medoid (Partitioning Around Medoids/PAM) karena memiliki keunggulan dalam menangani data yang mengandung pencilan (*outlier*) dibandingkan algoritma K-Means [4].

Data akademik mahasiswa merupakan salah satu sumber informasi yang dapat dimanfaatkan untuk mengevaluasi

proses pembelajaran. Penelitian ini menggunakan Student Performance Dataset yang berasal dari UCI Machine Learning Repository, kemudian mengambil 20 data sampel yang terdiri atas nilai Tugas, UTS, dan UAS sebagai objek penelitian. Ketiga atribut tersebut dipilih karena mewakili perkembangan hasil belajar mahasiswa selama satu semester dan dapat digunakan sebagai dasar dalam proses pengelompokan berdasarkan tingkat prestasi akademik [2].

Meskipun data akademik telah tersedia dalam jumlah yang cukup banyak, proses pengelompokan mahasiswa masih sering dilakukan secara manual berdasarkan pengamatan dosen. Cara tersebut membutuhkan waktu yang relatif lama serta berpotensi menghasilkan penilaian yang bersifat subjektif, terutama ketika jumlah mahasiswa semakin banyak. Selain itu, keberadaan nilai yang sangat tinggi maupun sangat rendah dapat memengaruhi hasil analisis apabila menggunakan algoritma clustering yang kurang tahan terhadap *outlier*, sehingga diperlukan metode yang mampu menghasilkan pengelompokan yang lebih representatif [4].

Beberapa penelitian sebelumnya telah membahas penerapan algoritma K-Medoid pada berbagai bidang. Andrea dan Nursobah [1] menerapkan algoritma K-Medoid untuk mengelompokkan penerima bantuan Uang Kuliah Tunggal (UKT) bagi mahasiswa terdampak Covid-19 dan menunjukkan bahwa metode tersebut mampu menghasilkan kelompok data yang lebih representatif. Park dan Jun [6] mengembangkan algoritma K-Medoid yang lebih efisien untuk meningkatkan kecepatan proses clustering, sedangkan Rousseeuw [7] memperkenalkan metode **Silhouette** sebagai teknik evaluasi kualitas hasil clustering yang hingga saat ini masih banyak digunakan dalam penelitian data mining.

Berdasarkan penelitian-penelitian tersebut, dapat diketahui bahwa algoritma K-Medoid memiliki keunggulan dalam menghasilkan pusat cluster berupa data nyata (*medoid*), sehingga lebih stabil terhadap keberadaan data pencilan dibandingkan algoritma berbasis rata-rata seperti K-Means [4]. Oleh karena itu, penelitian ini menerapkan algoritma K-Medoid pada data akademik mahasiswa untuk membentuk kelompok prestasi akademik berdasarkan nilai Tugas, UTS, dan UAS. Hasil clustering selanjutnya dievaluasi menggunakan Silhouette Score untuk mengetahui kualitas pengelompokan yang dihasilkan serta divisualisasikan dalam bentuk grafik tiga dimensi menggunakan Python. Berdasarkan uraian tersebut, penelitian ini dilakukan untuk mengeksplorasi penerapan algoritma K-Medoid (PAM) dalam mengelompokkan mahasiswa berdasarkan kemiripan nilai akademik. Hasil penelitian diharapkan dapat memberikan gambaran mengenai karakteristik setiap kelompok mahasiswa sehingga dapat dimanfaatkan sebagai bahan pertimbangan dalam proses evaluasi pembelajaran maupun penyusunan strategi akademik yang lebih tepat sasaran. Selain itu, penelitian ini juga diharapkan dapat menjadi salah satu referensi bagi penelitian selanjutnya yang berkaitan dengan penerapan teknik clustering dalam bidang pendidikan.

II. METODOLOGI PENELITIAN

Metodologi penelitian merupakan tahapan yang digunakan sebagai pedoman dalam pelaksanaan penelitian agar proses yang dilakukan berjalan secara sistematis dan menghasilkan keluaran yang sesuai dengan tujuan penelitian. Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining, khususnya teknik clustering menggunakan algoritma K-Medoid (Partitioning Around Medoids/PAM). Dataset yang digunakan merupakan Student Performance Dataset yang diperoleh dari UCI Machine Learning Repository, kemudian diolah menggunakan bahasa pemrograman Python pada lingkungan Google Colab.
Tahapan penelitian yang dilakukan dapat dijelaskan sebagai berikut.

2.1 Identifikasi Masalah

Tahap pertama adalah mengidentifikasi permasalahan yang menjadi dasar penelitian. Permasalahan yang ditemukan adalah proses pengelompokan mahasiswa berdasarkan prestasi akademik masih sering dilakukan secara manual sehingga membutuhkan waktu yang cukup lama dan berpotensi menghasilkan penilaian yang subjektif. Oleh karena itu, diperlukan suatu metode clustering yang mampu mengelompokkan mahasiswa secara otomatis berdasarkan tingkat kemiripan nilai akademik.

2.2 Studi Literatur

Pada tahap ini dilakukan pengumpulan berbagai referensi yang berkaitan dengan penelitian. Literatur diperoleh dari jurnal ilmiah, buku, prosiding, serta dokumentasi resmi yang membahas mengenai Data Mining, Clustering, algoritma K-Medoid, Manhattan Distance, Silhouette Score, serta implementasi Python menggunakan library **scikit-learn-extra**. Studi literatur bertujuan untuk memperkuat landasan teori sekaligus menjadi acuan dalam proses implementasi penelitian.

2.3 Pengumpulan Dataset

Dataset yang digunakan berasal dari Student Performance Dataset yang dipublikasikan oleh UCI Machine Learning Repository. Dari dataset tersebut dipilih 20 data pertama sebagai sampel penelitian dengan menggunakan tiga atribut utama, yaitu G1, G2, dan G3 yang kemudian dipetakan menjadi nilai Tugas, UTS, dan UAS. Selanjutnya seluruh nilai dikonversi dari skala 0–20 menjadi skala 0–100 agar sesuai dengan sistem penilaian perguruan tinggi di Indonesia.

2.4 Pra-pemrosesan Data (Preprocessing)

Tahap preprocessing dilakukan untuk menyiapkan data sebelum proses clustering. Pada tahap ini dilakukan pemilihan atribut yang digunakan dalam penelitian, perubahan nama atribut agar sesuai dengan studi kasus, konversi skala nilai, serta pengecekan kelengkapan data.

Hasil dari tahap ini berupa dataset yang siap digunakan sebagai input algoritma K-Medoid.

2.5 Implementasi Algoritma K-Medoid

Tahap selanjutnya adalah menerapkan algoritma K-Medoid (PAM) menggunakan bahasa pemrograman Python. Proses implementasi dilakukan dengan menentukan jumlah cluster sebanyak tiga kelompok (K = 3), menggunakan metrik Manhattan Distance sebagai perhitungan jarak antar data, kemudian menentukan medoid awal, menghitung jarak setiap data terhadap medoid, mengelompokkan data ke cluster terdekat, dan melakukan proses pertukaran (*swap*) medoid hingga diperoleh total biaya (*total cost*) yang minimum atau algoritma mencapai kondisi konvergen.

2.6 Evaluasi Hasil Clustering

Setelah proses clustering selesai dilakukan, kualitas hasil pengelompokan dievaluasi menggunakan **Silhouette Score**. Nilai Silhouette Score digunakan untuk mengukur tingkat kedekatan data dalam satu cluster (cohesion) serta tingkat keterpisahan antarcluster (separation). Semakin tinggi nilai Silhouette Score, maka semakin baik kualitas cluster yang dihasilkan oleh algoritma.

2.7 Visualisasi dan Analisis Hasil

Tahap terakhir adalah menampilkan hasil clustering dalam bentuk visualisasi 3D Scatter Plot menggunakan pustaka Matplotlib pada Python. Setiap cluster ditampilkan dengan warna yang berbeda, sedangkan medoid ditandai menggunakan simbol khusus agar mudah diidentifikasi. Selanjutnya dilakukan analisis terhadap karakteristik masing-masing cluster untuk mengetahui kelompok mahasiswa dengan prestasi tinggi, sedang, dan rendah serta menginterpretasikan hasil yang diperoleh sebagai dasar pengambilan keputusan akademik.

2.8 Alur Penelitian

Secara keseluruhan tahapan penelitian yang dilakukan dapat diringkas sebagai berikut:

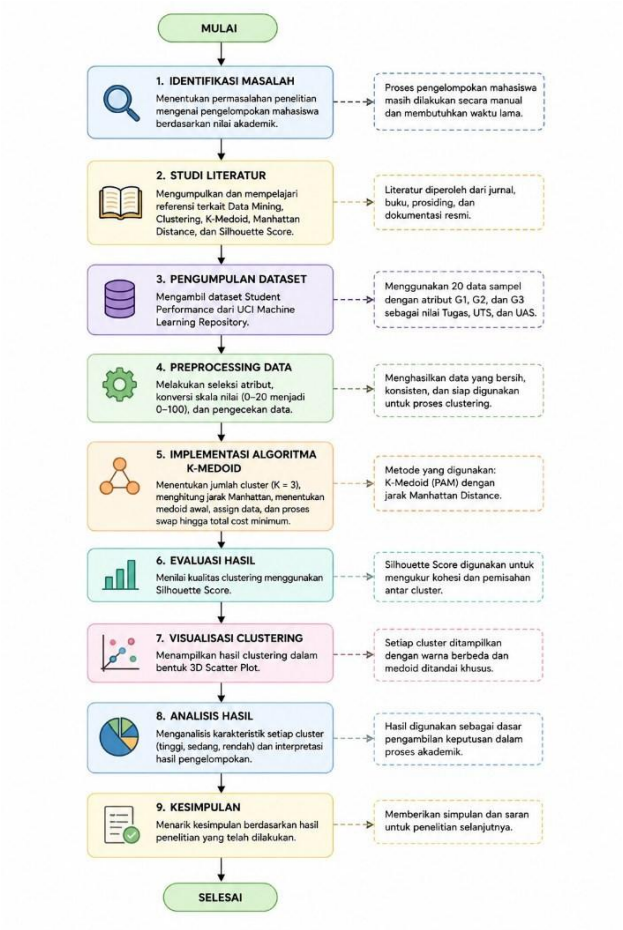
1. Identifikasi masalah.
2. Studi literatur.
3. Pengumpulan dataset.
4. Preprocessing data.
5. Implementasi algoritma K-Medoid.
6. Evaluasi menggunakan Silhouette Score.
7. Visualisasi hasil clustering.
8. Analisis dan penarikan kesimpulan.

Metodologi penelitian tersebut disusun secara berurutan agar proses penelitian berlangsung secara sistematis, mulai dari identifikasi masalah hingga diperoleh hasil clustering yang dapat digunakan sebagai bahan analisis terhadap performa akademik.

III. DESAIN, HASIL DAN PEMBAHASAN

3.1 Desain Penelitian

Desain penelitian menggambarkan tahapan proses yang dilakukan dalam menerapkan algoritma K-Medoid (Partitioning Around Medoids/PAM) untuk mengelompokkan mahasiswa berdasarkan nilai akademik. Penelitian diawali dengan pengumpulan dataset, dilanjutkan dengan proses *preprocessing* data, implementasi algoritma K-Medoid menggunakan Python, evaluasi hasil menggunakan Silhouette Score, serta visualisasi hasil clustering dalam bentuk grafik tiga dimensi. Rancangan penelitian ini bertujuan agar setiap tahapan dapat dilakukan secara sistematis sehingga menghasilkan pengelompokan data yang optimal.



Gambar 1. Desain Penelitian

Gambar 1 menunjukkan alur penelitian yang digunakan dalam proses clustering data akademik mahasiswa. Tahapan dimulai dari identifikasi masalah hingga penarikan kesimpulan. Setiap proses saling berkaitan sehingga menghasilkan pengelompokan mahasiswa berdasarkan tingkat kemiripan nilai akademik.

3.2 Hasil Penelitian

Penelitian ini menggunakan 20 data mahasiswa yang diambil dari Student Performance Dataset. Data yang digunakan terdiri atas tiga atribut, yaitu nilai Tugas, nilai UTS, dan nilai UAS yang telah dikonversi ke dalam skala 0–100. Selanjutnya dataset diproses menggunakan algoritma K-Medoid dengan jumlah cluster sebanyak tiga kelompok dan menggunakan metrik Manhattan Distance sebagai ukuran kedekatan antar data.

Hasil implementasi menunjukkan bahwa algoritma K-Medoid berhasil membentuk tiga kelompok mahasiswa berdasarkan karakteristik nilai akademiknya. Kelompok pertama terdiri atas mahasiswa dengan prestasi akademik rendah, kelompok kedua berisi mahasiswa dengan prestasi akademik sedang, sedangkan kelompok ketiga merupakan mahasiswa dengan prestasi akademik tinggi. Pengelompokan tersebut menunjukkan bahwa algoritma mampu memisahkan data berdasarkan tingkat kemiripan nilai secara efektif.
Evaluasi kualitas clustering dilakukan menggunakan Silhouette Score. Nilai evaluasi yang diperoleh menunjukkan bahwa hasil clustering memiliki tingkat kohesi dan pemisahan antarcluster yang cukup baik. Selain itu, visualisasi menggunakan grafik 3D Scatter Plot memperlihatkan bahwa masing-masing cluster memiliki distribusi yang jelas sehingga memudahkan proses interpretasi hasil.

3.3 Pembahasan
Hasil penelitian menunjukkan bahwa algoritma K-Medoid mampu mengelompokkan mahasiswa berdasarkan pola kemiripan nilai Tugas, UTS, dan UAS. Berbeda dengan algoritma K-Means yang menggunakan nilai rata-rata sebagai pusat cluster, K-Medoid memilih data nyata sebagai pusat kelompok (medoid), sehingga lebih stabil ketika terdapat nilai yang sangat tinggi maupun sangat rendah. Karakteristik tersebut menjadikan K-Medoid lebih sesuai digunakan pada data akademik yang memiliki kemungkinan munculnya nilai pencilan (*outlier*).
Kelompok mahasiswa dengan prestasi rendah dapat dijadikan dasar bagi dosen untuk memberikan pembinaan, bimbingan belajar, maupun program remedial. Sementara itu, kelompok prestasi sedang dapat diberikan latihan tambahan agar mengalami peningkatan hasil belajar. Mahasiswa yang termasuk dalam kelompok prestasi tinggi dapat diarahkan sebagai tutor sebaya atau diberikan materi pengayaan sehingga proses pembelajaran menjadi lebih efektif.
Visualisasi hasil clustering juga memberikan gambaran yang lebih mudah dipahami mengenai distribusi data akademik mahasiswa. Titik medoid yang berada pada salah satu data asli menunjukkan bahwa pusat cluster benar-benar merepresentasikan anggota kelompoknya. Dengan demikian, hasil penelitian ini membuktikan bahwa algoritma K-Medoid dapat digunakan sebagai salah satu alternatif metode dalam membantu proses analisis performa akademik mahasiswa secara objektif dan efisien.

IV. PERHITUNGAN MANUAL
Perhitungan manual dilakukan untuk memberikan gambaran mengenai mekanisme kerja algoritma K-Medoid (Partitioning Around Medoids/PAM) dalam membentuk cluster berdasarkan tingkat kemiripan data. Proses ini dilakukan sebelum implementasi menggunakan Python agar setiap tahapan algoritma dapat dipahami secara konseptual. Perhitungan menggunakan Manhattan Distance sebagai ukuran jarak antar data dengan jumlah cluster (K) sebanyak tiga kelompok.

4.1 Penentuan Medoid Awal
Tahap pertama dalam algoritma K-Medoid adalah menentukan sejumlah data yang akan dijadikan sebagai medoid awal. Pada penelitian ini dipilih tiga data secara acak sebagai pusat cluster awal, yaitu:
- **Medoid 1 (Cluster 1)** = Mahasiswa 9 $(80, 90, 95)$
- **Medoid 2 (Cluster 2)** = Mahasiswa 13 $(70, 70, 70)$
- **Medoid 3 (Cluster 3)** = Mahasiswa 1 $(25, 30, 30)$

Pemilihan medoid awal ini menjadi dasar dalam proses perhitungan jarak setiap data terhadap masing-masing pusat cluster.

4.2 Perhitungan Manhattan Distance
Setelah medoid awal ditentukan, langkah berikutnya adalah menghitung jarak setiap data terhadap ketiga medoid menggunakan rumus Manhattan Distance.
[D(x,y)=\sum_{i=1}^{n}|x_i-y_i|]
Sebagai contoh, perhitungan dilakukan terhadap Mahasiswa 2 yang memiliki nilai (25, 25, 30).
Jarak ke Medoid 1 (Mahasiswa 9)
[|25-80|+|25-90|+|30-95|=55+65+65=185]
Jarak ke Medoid 2 (Mahasiswa 13)
[|25-70|+|25-70|+|30-70|=45+45+40=130]
Jarak ke Medoid 3 (Mahasiswa 1)
[|25-25|+|25-30|+|30-30|=0+5+0=5]
Karena jarak terkecil adalah 5, maka Mahasiswa 2 menjadi anggota Cluster 3.
Perhitungan yang sama dilakukan terhadap seluruh data mahasiswa sehingga diperoleh hasil pengelompokan awal beserta nilai Total Cost.

4.3 Perhitungan Total Cost
Setelah seluruh data memperoleh cluster masing-masing, dilakukan perhitungan Total Cost, yaitu jumlah seluruh jarak terdekat setiap data terhadap medoidnya.
Pada Iterasi pertama diperoleh:
Total Cost = 465
Nilai tersebut digunakan sebagai acuan untuk menentukan apakah diperlukan pergantian (*swap*) medoid.

4.4 Proses Swap Medoid
Pada tahap berikutnya dilakukan proses pertukaran (*swap*) salah satu medoid dengan data lain yang bukan medoid. Tujuannya adalah mencari kombinasi medoid yang menghasilkan Total Cost paling kecil.
Dalam penelitian ini dilakukan pertukaran:
- Medoid 2 (Mahasiswa 13)
- diganti menjadi Mahasiswa 12 (50, 60, 60)

Selanjutnya seluruh jarak dihitung kembali sehingga diperoleh hasil pengelompokan baru.

4.5 Evaluasi Iterasi

Hasil perhitungan iterasi kedua menunjukkan bahwa:

- Total Cost Iterasi 1 = **465**
- Total Cost Iterasi 2 = **515**

Karena nilai Total Cost Iterasi 2 lebih besar daripada Iterasi 1, maka proses pertukaran medoid dibatalkan. Medoid pada Iterasi pertama tetap digunakan sebagai pusat cluster akhir karena menghasilkan biaya (*cost*) yang lebih kecil. Dengan demikian proses iterasi dihentikan dan hasil clustering digunakan sebagai dasar implementasi menggunakan Python pada tahap berikutnya.

V. IMPLEMENTASI PYTHON

Implementasi algoritma K-Medoid (Partitioning Around Medoids/PAM) dilakukan menggunakan bahasa pemrograman Python pada lingkungan Google Colab. Penggunaan Google Colab dipilih karena menyediakan lingkungan pemrograman yang mudah diakses tanpa memerlukan instalasi perangkat lunak tambahan serta mendukung berbagai pustaka (*library*) yang dibutuhkan dalam proses data mining. Implementasi ini bertujuan untuk memvalidasi hasil perhitungan manual sekaligus menghasilkan proses clustering yang lebih cepat, akurat, dan efisien.

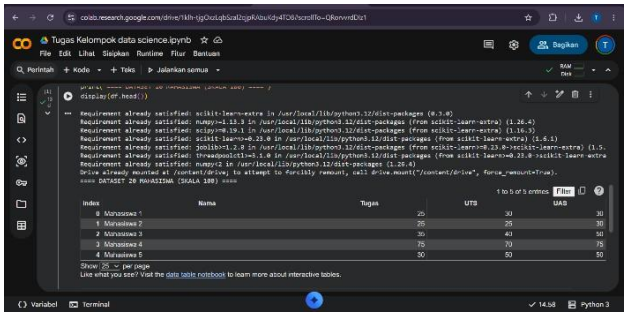
Link Google Colab:

https://colab.research.google.com/drive/1klh-tjgOxzLqbSzal2qjpRAbuKdy4TO8?usp=sharing

5.1 Persiapan Lingkungan Pemrograman

Tahap awal implementasi dilakukan dengan menyiapkan lingkungan kerja Python. Library yang digunakan meliputi pandas untuk pengolahan data, NumPy untuk komputasi numerik, Matplotlib untuk visualisasi data, scikit-learn-extra untuk implementasi algoritma K-Medoid, serta silhouette_score dari Scikit-Learn sebagai metode evaluasi kualitas clustering. Selain itu, Google Drive dihubungkan ke Google Colab agar dataset dapat dibaca secara langsung.

Tahapan ini menghasilkan lingkungan pemrograman yang siap digunakan untuk melakukan proses clustering terhadap data akademik mahasiswa.

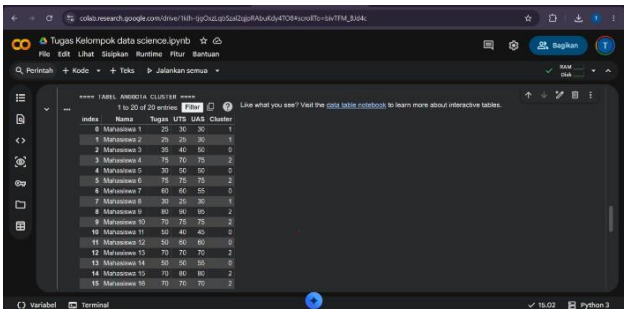


Gambar 2. Persiapan Library dan Google Colab

5.2 Persiapan Dataset

Dataset yang digunakan merupakan Student Performance Dataset yang berasal dari UCI Machine Learning Repository. Dari dataset tersebut dipilih tiga atribut utama, yaitu G1, G2, dan G3 yang kemudian diubah menjadi atribut Tugas, UTS, dan UAS. Seluruh nilai dikonversi dari skala 0–20 menjadi skala 0–100 dengan mengalikan setiap nilai sebesar lima agar sesuai dengan sistem penilaian pada perguruan tinggi di Indonesia.

Setelah proses konversi selesai, data siap digunakan sebagai masukan (*input*) dalam proses clustering menggunakan algoritma K-Medoid.

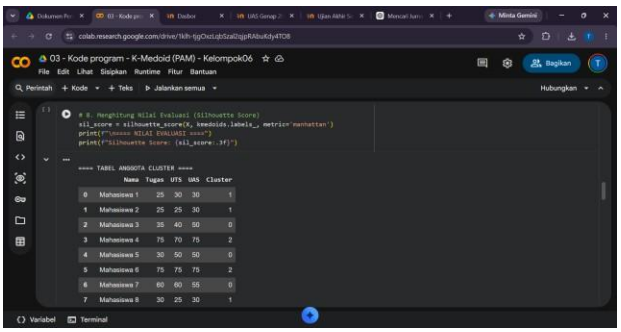


Gambar 3. Dataset Setelah Preprocessing

5.3 Implementasi Algoritma K-Medoid

Tahap implementasi dilakukan dengan membentuk model K-Medoid menggunakan parameter jumlah cluster (K = 3) dan metrik jarak Manhattan Distance. Selanjutnya model dijalankan menggunakan fungsi fit() sehingga setiap data mahasiswa memperoleh label cluster sesuai dengan tingkat kemiripan nilai akademiknya.

Setelah proses clustering selesai, label cluster ditambahkan ke dalam dataset sehingga setiap mahasiswa memiliki informasi mengenai kelompoknya masing-masing. Selain itu, program juga menampilkan data yang terpilih sebagai medoid, yaitu pusat cluster yang merupakan data asli dalam dataset.



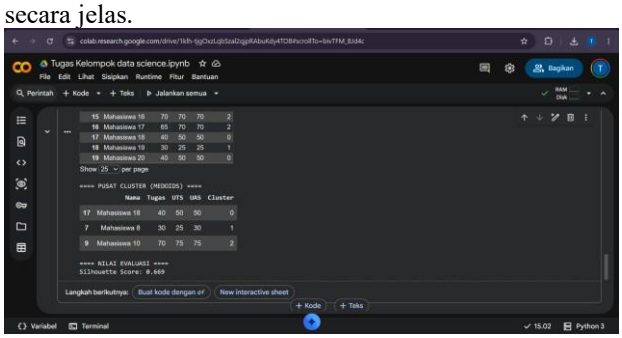
Gambar 4. Hasil Implementasi Algoritma K-Medoid

5.4 Evaluasi Menggunakan Silhouette Score

Untuk mengetahui kualitas hasil clustering, dilakukan evaluasi menggunakan Silhouette Score. Nilai Silhouette Score dihitung berdasarkan tingkat kedekatan data dalam satu cluster (*cohesion*) dan tingkat keterpisahan antarcluster (*separation*). Semakin tinggi nilai yang diperoleh, maka semakin baik kualitas pengelompokan yang dihasilkan.

Pada penelitian ini, nilai Silhouette Score menunjukkan bahwa hasil clustering memiliki kualitas yang cukup baik sehingga algoritma K-Medoid mampu membentuk

kelompok mahasiswa dengan karakteristik yang berbeda secara jelas.

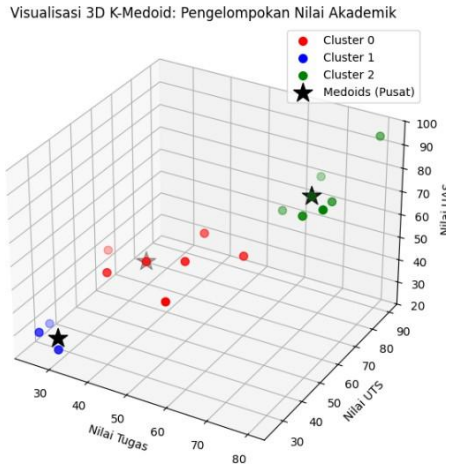


Gambar 5. Output Nilai Silhouette Score

5.5 Visualisasi Hasil Clustering

Visualisasi dilakukan menggunakan 3D Scatter Plot dengan tiga sumbu yang mewakili nilai Tugas, UTS, dan UAS. Setiap cluster diberikan warna yang berbeda untuk memudahkan proses identifikasi, sedangkan titik medoid ditampilkan menggunakan simbol bintang berwarna hitam sebagai pusat cluster.

Visualisasi tersebut memperlihatkan bahwa data mahasiswa berhasil dipisahkan menjadi tiga kelompok berdasarkan kemiripan nilai akademik. Posisi medoid yang berada pada salah satu data asli menunjukkan karakteristik utama algoritma K-Medoid yang menggunakan objek nyata sebagai pusat cluster.



Gambar 6. Visualisasi 3D Scatter Plot Hasil Clustering

5.6 Analisis Implementasi

Berdasarkan implementasi menggunakan Python, algoritma K-Medoid berhasil melakukan proses clustering terhadap data akademik mahasiswa secara otomatis. Seluruh tahapan mulai dari pembacaan dataset, proses clustering, evaluasi menggunakan Silhouette Score, hingga visualisasi hasil dapat dijalankan dengan baik menggunakan Google Colab.

Hasil implementasi juga menunjukkan kesesuaian dengan proses perhitungan manual yang telah dilakukan pada bab sebelumnya. Dengan demikian, penggunaan Python mampu meningkatkan efisiensi proses analisis data serta mempermudah interpretasi hasil clustering melalui penyajian tabel maupun visualisasi tiga dimensi.

VI. HASIL DAN PEMBAHASAN

6.1 Hasil Clustering Mahasiswa

Berdasarkan implementasi algoritma K-Medoid (Partitioning Around Medoids/PAM) menggunakan Python terhadap data akademik mahasiswa, proses clustering berhasil membentuk tiga kelompok berdasarkan tingkat kemiripan nilai Tugas, UTS, dan UAS. Setiap mahasiswa secara otomatis dikelompokkan ke dalam cluster yang memiliki karakteristik nilai yang serupa berdasarkan perhitungan jarak Manhattan Distance.

Hasil clustering menunjukkan bahwa ketiga kelompok yang terbentuk memiliki karakteristik yang berbeda. Kelompok pertama terdiri atas mahasiswa dengan nilai akademik relatif rendah, kelompok kedua berisi mahasiswa dengan nilai akademik sedang, sedangkan kelompok ketiga terdiri atas mahasiswa dengan nilai akademik tinggi. Pengelompokan tersebut memberikan gambaran yang lebih jelas mengenai distribusi kemampuan akademik mahasiswa sehingga dapat dijadikan sebagai dasar dalam proses evaluasi pembelajaran.

6.2 Interpretasi Karakteristik Setiap Cluster

Berdasarkan hasil clustering, setiap kelompok memiliki karakteristik yang berbeda sesuai dengan pola nilai akademik anggotanya.

Cluster Prestasi Rendah

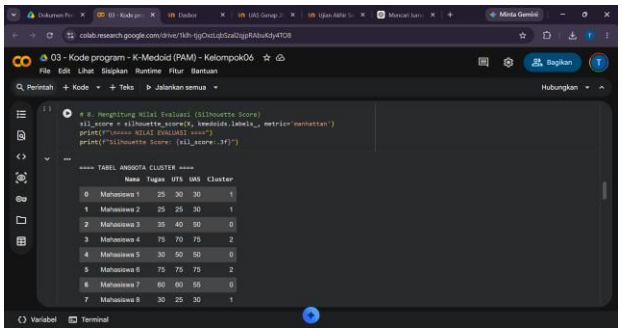
Kelompok ini terdiri atas mahasiswa yang memiliki nilai Tugas, UTS, dan UAS relatif rendah dibandingkan kelompok lainnya. Mahasiswa dalam cluster ini memerlukan perhatian khusus karena menunjukkan performa akademik yang masih berada di bawah rata-rata. Dosen dapat memberikan program remedial, bimbingan belajar, maupun pendampingan secara intensif agar kemampuan akademik mahasiswa dapat meningkat.

Cluster Prestasi Sedang

Kelompok ini merupakan mahasiswa dengan capaian akademik yang berada pada kategori sedang. Nilai yang dimiliki relatif stabil dan telah memenuhi standar pembelajaran, namun masih memiliki peluang untuk ditingkatkan. Strategi pembelajaran yang dapat diterapkan antara lain pemberian latihan tambahan, diskusi kelompok, maupun pembelajaran kolaboratif.

Cluster Prestasi Tinggi

Kelompok ini berisi mahasiswa yang memiliki nilai akademik tinggi pada seluruh komponen penilaian. Mahasiswa dalam cluster ini menunjukkan pemahaman materi yang baik sehingga dapat diberikan tugas pengayaan (*enrichment*), proyek lanjutan, maupun kesempatan menjadi tutor sebaya untuk membantu mahasiswa pada kelompok lainnya.

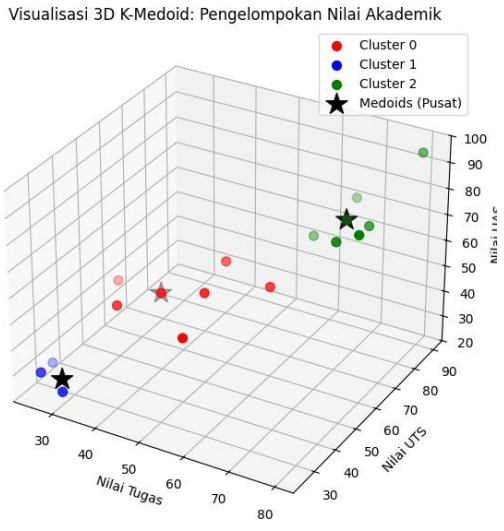


Gambar 7. Karakteristik Masing-masing Cluster

6.3 Analisis Visualisasi Clustering

Visualisasi hasil clustering menggunakan 3D Scatter Plot memberikan gambaran yang lebih jelas mengenai penyebaran data mahasiswa berdasarkan tiga atribut, yaitu nilai Tugas, UTS, dan UAS. Setiap cluster ditampilkan menggunakan warna yang berbeda sehingga batas antar kelompok dapat diamati dengan mudah. Selain itu, medoid sebagai pusat cluster ditampilkan menggunakan simbol khusus sehingga posisi pusat masing-masing kelompok dapat dikenali secara visual.

Hasil visualisasi memperlihatkan bahwa sebagian besar data dalam satu cluster berada pada lokasi yang berdekatan, sedangkan jarak antarcluster cukup jelas. Kondisi tersebut menunjukkan bahwa algoritma K-Medoid mampu membentuk kelompok yang memiliki tingkat kemiripan internal yang baik sekaligus mempertahankan pemisahan antar kelompok secara optimal.



Gambar 8. Visualisasi 3D Scatter Plot Hasil Clustering

6.4 Evaluasi Hasil Clustering

Evaluasi dilakukan menggunakan Silhouette Score untuk mengetahui kualitas hasil pengelompokan. Nilai Silhouette Score yang diperoleh menunjukkan bahwa hasil clustering memiliki tingkat kohesi yang baik di dalam cluster dan tingkat pemisahan yang cukup jelas antarcluster. Hal tersebut mengindikasikan bahwa algoritma K-Medoid mampu menghasilkan pengelompokan yang sesuai dengan karakteristik data akademik mahasiswa.

Selain evaluasi numerik, hasil clustering juga didukung oleh visualisasi tiga dimensi yang menunjukkan bahwa sebagian besar data telah berada pada kelompok yang tepat sesuai dengan tingkat kemiripan nilainya. Dengan demikian, proses clustering yang dilakukan dapat dikatakan berhasil dalam mengidentifikasi pola akademik mahasiswa.

6.5 Pembahasan

Berdasarkan hasil penelitian, algoritma K-Medoid terbukti mampu mengelompokkan mahasiswa berdasarkan kemiripan nilai akademik secara efektif. Berbeda dengan algoritma K-Means yang menggunakan nilai rata-rata sebagai pusat cluster, K-Medoid menggunakan data nyata sebagai pusat kelompok (*medoid*). Karakteristik tersebut menjadikan algoritma ini lebih stabil ketika terdapat data dengan nilai yang sangat tinggi ataupun sangat rendah (*outlier*).

Hasil pengelompokan memberikan manfaat bagi institusi pendidikan dalam melakukan evaluasi proses pembelajaran. Mahasiswa yang berada pada kelompok prestasi rendah dapat menjadi prioritas dalam program pembinaan akademik, sedangkan mahasiswa pada kelompok prestasi tinggi dapat diberikan program pengayaan maupun dijadikan tutor sebaya. Dengan adanya proses clustering, dosen dapat mengambil keputusan akademik secara lebih objektif berdasarkan pola data yang dihasilkan.

Secara keseluruhan, implementasi algoritma K-Medoid menggunakan Python berhasil mempermudah proses analisis data akademik dibandingkan proses pengelompokan secara manual. Selain menghasilkan pengelompokan yang lebih cepat dan efisien, hasil visualisasi juga memudahkan pengguna dalam memahami karakteristik setiap kelompok mahasiswa sehingga dapat mendukung pengambilan keputusan dalam bidang pendidikan.

VII. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai penerapan algoritma **K-Medoid (Partitioning Around Medoids/PAM)** untuk mengelompokkan mahasiswa berdasarkan nilai akademik, dapat disimpulkan bahwa algoritma K-Medoid berhasil diterapkan dengan baik pada **Student Performance Dataset** yang telah disesuaikan dengan skala penilaian 0–100. Proses clustering menggunakan tiga atribut, yaitu nilai Tugas, UTS, dan UAS, mampu menghasilkan pengelompokan mahasiswa ke dalam tiga cluster yang memiliki karakteristik prestasi akademik yang berbeda, yaitu kelompok prestasi tinggi, sedang, dan rendah.

Implementasi algoritma menggunakan bahasa pemrograman Python pada lingkungan Google Colab memberikan hasil yang sesuai dengan proses perhitungan manual yang telah dilakukan sebelumnya. Evaluasi menggunakan **Silhouette Score** menunjukkan bahwa

hasil clustering memiliki kualitas yang baik, sedangkan visualisasi dalam bentuk **3D Scatter Plot** memperlihatkan pemisahan antarcluster yang cukup jelas sehingga memudahkan proses interpretasi hasil pengelompokan. Hal ini menunjukkan bahwa algoritma K-Medoid mampu mengidentifikasi pola kemiripan data akademik secara efektif.

Penggunaan algoritma K-Medoid memiliki keunggulan karena menggunakan **medoid** sebagai pusat cluster yang berasal dari data asli sehingga lebih robust terhadap keberadaan data pencilan (*outlier*) dibandingkan algoritma yang menggunakan nilai rata-rata sebagai pusat cluster. Karakteristik tersebut menjadikan K-Medoid sesuai untuk diterapkan pada data akademik yang memiliki variasi nilai cukup besar. Hasil penelitian ini diharapkan dapat membantu dosen maupun institusi pendidikan dalam melakukan analisis performa akademik mahasiswa secara lebih objektif, cepat, dan efisien.

Untuk penelitian selanjutnya, disarankan menggunakan jumlah data yang lebih besar sehingga hasil clustering dapat menggambarkan kondisi akademik secara lebih menyeluruh. Selain itu, penelitian berikutnya dapat menambahkan atribut lain, seperti tingkat kehadiran, keaktifan mahasiswa, maupun indeks prestasi kumulatif (IPK), serta membandingkan performa algoritma K-Medoid dengan algoritma clustering lainnya, seperti **K-Means**, **DBSCAN**, atau **Hierarchical Clustering**, sehingga dapat diketahui metode yang paling optimal untuk analisis data akademik..

VIII. DAFTAR PUSTAKA

[1] Andrea, R., & Nursobah. (2022). Penerapan Algoritma K-Medoids Untuk Pengelompokan Data Penerima Bantuan Uang Kuliah Tunggal Bagi Mahasiswa Terdampak Covid-19. *BITS: Building of Informatics, Technology and Science*, 3(4), 632-638.

[2] Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. *Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008)*, 5-12.

[3] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

[4] Scikit-Learn Extra Developers. (2023). *KMedoids Documentation*. GitHub Repository (scikit-learn-contrib).

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[6] Arora, P., Deepali, & Varshney, S. (2016). *Analysis of K-Means and K-Medoids Algorithm for Big Data*. Procedia Computer Science, 78, 507–512.

[7] Park, H. S., & Jun, C. H. (2009). *A Simple and Fast Algorithm for K-Medoids Clustering*. Expert Systems with Applications, 36(2), 3336–3341.

[8] Rousseeuw, P. J. (1987). *Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis*. Journal of Computational and Applied Mathematics, 20, 53–65.

[9] Jain, A. K. (2010). *Data Clustering: 50 Years Beyond K-Means*. Pattern Recognition Letters, 31(8), 651–666.

[10] Berkhin, P. (2006). *A Survey of Clustering Data Mining Techniques*. Springer.

[11] Xu, D., & Tian, Y. (2015). *A Comprehensive Survey of Clustering Algorithms*. Annals of Data Science, 2(2), 165–193

[12] Aggarwal, C. C., & Reddy, C. K. (2014). *Data Clustering: Algorithms and Applications*. CRC Press.

[13] Romero, C., & Ventura, S. (2010). *Educational Data Mining: A Review of the State of the Art*. IEEE Transactions on Systems, Man, and Cybernetics Part C, 40(6), 601–618.

[14] Romero, C., & Ventura, S. (2013). *Data Mining in Education*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12–27.

[15] Bunkers, M. J., Miller, J. R., & DeGaetano, A. T. (1996). *Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique*. Journal of Climate, 9(1), 130–146.