

Perbandingan K-Means dan Hierarchical Clustering untuk Pengelompokan Calon Penerima Beasiswa Berdasarkan Data Akademik Sosial dan Ekonomi Mahasiswa

Septia Harliansyah*, Muhammad Hafizh Al-Ghifari Rangkuti, Muhammad Zikri Ramadhan

Magister Teknologi Informasi, Pascasarjana, Universitas Pembangunan Pancabudi, Medan, Indonesia

Email: ¹*septiaharliansyah@unsam.ac.id, ²hafizhrangkuti04@gmail.com, ³muhammadzikriramadhan7@gmail.com

Email Penulis Korespondensi: septiaharliansyah@unsam.ac.id*

Submitted: 25/05/2026; Accepted: 10/06/2026; Published: 30/06/2026

Abstrak— Penelitian ini bertujuan untuk membandingkan performa algoritma K-Means dan Hierarchical Clustering dalam melakukan pengelompokan calon penerima beasiswa berdasarkan data akademik dan sosial ekonomi. Dataset yang digunakan terdiri dari 3212 data calon penerima beasiswa yang mencakup variabel akademik dan kondisi sosial ekonomi. Tahapan penelitian meliputi preprocessing data melalui normalisasi, penentuan jumlah cluster menggunakan metode Elbow, serta penerapan algoritma K-Means dan Hierarchical Clustering. Evaluasi kualitas cluster dilakukan menggunakan Silhouette Score untuk mengukur tingkat kohesi dan separasi antar cluster. Hasil penelitian menunjukkan bahwa algoritma K-Means menghasilkan Silhouette Score sebesar 0,205, sedangkan Hierarchical Clustering menghasilkan Silhouette Score sebesar 0,181. Nilai tersebut menunjukkan bahwa K-Means mampu membentuk cluster yang lebih kompak dan memiliki pemisahan yang lebih baik dibandingkan Hierarchical Clustering. Selain itu, Hierarchical Clustering cenderung menghasilkan struktur cluster yang kurang optimal pada dataset dengan jumlah data yang relatif besar. Dengan demikian, K-Means dapat direkomendasikan sebagai metode yang lebih efektif untuk pengelompokan calon penerima beasiswa berdasarkan data akademik dan sosial ekonomi. Penelitian ini diharapkan dapat menjadi referensi dalam pemilihan metode clustering yang tepat untuk mendukung pengambilan keputusan yang lebih objektif dan tepat sasaran.

Kata Kunci: Clustering; Data Mining; Hierarchical Clustering; K-Means; Silhouette Score.

Abstract— This study aims to compare the performance of the K-Means and Hierarchical Clustering algorithms in grouping scholarship applicants based on academic and socioeconomic data. The dataset used in this study consisted of 3212 scholarship applicant records, including academic and socioeconomic variables. The research procedure involved data preprocessing through normalization, determining the optimal number of clusters using the Elbow Method, and implementing both K-Means and Hierarchical Clustering algorithms. Cluster quality was evaluated using the Silhouette Score to measure cluster cohesion and separation. The results showed that the K-Means algorithm achieved a Silhouette Score of 0.205, while Hierarchical Clustering obtained a Silhouette Score of 0.181. These findings indicate that K-Means produced relatively more compact clusters with better separation compared to Hierarchical Clustering on the dataset used in this study. In contrast, Hierarchical Clustering tended to generate less distinct cluster structures when applied to a relatively large dataset. However, considering that both Silhouette Score values remain relatively low, the clustering quality obtained cannot yet be considered strong. Therefore, K-Means can be regarded as a relatively better alternative than Hierarchical Clustering for grouping scholarship applicants based on academic and socioeconomic characteristics within the context of this study. The findings of this research are expected to contribute as a reference for selecting appropriate clustering methods in data mining applications and to support more objective and data-driven decision-making processes in scholarship allocation.

Keywords: Clustering; Data Mining; Hierarchical Clustering; K-Means; Silhouette Score.

1. PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan besar dalam berbagai bidang kehidupan, termasuk pada sektor pendidikan. Perguruan tinggi saat ini memanfaatkan sistem informasi akademik untuk menyimpan berbagai data mahasiswa secara terstruktur, seperti data nilai akademik, indeks prestasi kumulatif (IPK), kondisi ekonomi keluarga, hingga riwayat penerimaan bantuan pendidikan. Ketersediaan data dalam jumlah besar tersebut sebenarnya memiliki potensi yang sangat penting dalam mendukung pengambilan keputusan yang lebih objektif dan tepat sasaran, khususnya dalam proses seleksi calon penerima beasiswa. Namun, pada praktiknya proses penentuan calon penerima beasiswa masih sering dilakukan secara manual atau hanya berdasarkan beberapa indikator tertentu sehingga belum mampu memanfaatkan keseluruhan data yang tersedia secara optimal. Kondisi ini dapat menyebabkan proses seleksi menjadi kurang efektif, kurang objektif, dan berpotensi menimbulkan ketidaktepatan sasaran dalam penyaluran beasiswa. Selain itu, banyaknya jumlah data mahasiswa juga menyulitkan pihak pengelola dalam melakukan analisis karakteristik calon penerima beasiswa secara cepat dan akurat. Oleh karena itu, diperlukan suatu metode yang mampu mengolah data dalam jumlah besar untuk menemukan pola atau informasi tersembunyi yang dapat digunakan sebagai dasar pengambilan keputusan yang lebih baik. Salah satu solusi yang dapat diterapkan adalah penggunaan teknik data mining untuk melakukan analisis data calon penerima beasiswa secara otomatis dan sistematis [1][2][3].

Data mining merupakan proses pengolahan data untuk menemukan pola, hubungan, maupun pengetahuan baru yang berguna dari sekumpulan data berukuran besar. Salah satu teknik dalam *data mining* yang banyak digunakan adalah *clustering*, yaitu metode pengelompokan data berdasarkan tingkat kemiripan karakteristik tanpa

menggunakan label kelas sebelumnya atau dikenal sebagai *unsupervised learning*. Teknik *clustering* dapat dimanfaatkan untuk mengelompokkan mahasiswa calon penerima beasiswa berdasarkan data akademik dan sosial ekonomi, seperti IPK, penghasilan orang tua, jumlah tanggungan keluarga, semester, dan nilai akademik. Hasil pengelompokan tersebut dapat membantu pihak kampus dalam memahami karakteristik mahasiswa calon penerima beasiswa serta menjadi bahan evaluasi dalam menentukan kebijakan penyaluran bantuan pendidikan yang lebih tepat sasaran [4][5].

Beberapa penelitian sebelumnya telah menerapkan metode *clustering* dalam bidang pendidikan, khususnya untuk pengelompokan mahasiswa dan calon penerima bantuan pendidikan. Penelitian oleh Nelson Butarbutar dkk. membandingkan algoritma *K-Means* dan *Fuzzy C-Means* dalam pengelompokan data siswa berdasarkan prestasi nilai akademik. Hasil penelitian menunjukkan bahwa algoritma *K-Means* memiliki proses komputasi yang lebih cepat dan sederhana dibandingkan *Fuzzy C-Means*. Namun, *K-Means* memiliki kelemahan pada penentuan *centroid* awal yang dapat memengaruhi hasil *cluster* yang dihasilkan [6]. Menurut *Data Mining: Concepts and Techniques* oleh Han, Kamber, dan Pei, *Hierarchical Clustering* merupakan metode *clustering* yang membentuk kelompok data secara bertahap berdasarkan tingkat kemiripan antar data dan dapat direpresentasikan dalam bentuk struktur hierarki atau dendrogram. [7].

Selanjutnya, Penelitian oleh Diah Nilam Cahya menggunakan algoritma *K-Means* untuk menentukan kelompok calon penerima bantuan pendidikan berdasarkan variabel IPK dan penghasilan orang tua. Penelitian tersebut menghasilkan pengelompokan data yang cukup baik dalam mengidentifikasi kelompok mahasiswa calon penerima bantuan pendidikan. Namun, evaluasi kualitas *cluster* pada penelitian tersebut masih menggunakan pendekatan inerti sehingga belum sepenuhnya menunjukkan kualitas pemisahan antar *cluster* secara menyeluruh [8]. Selain itu, penelitian oleh Sindrawati dkk. menerapkan metode *K-Means* pada data siswa calon penerima beasiswa dengan fokus pada variabel akademik seperti nilai dan prestasi belajar. Penelitian tersebut menunjukkan bahwa *clustering* mampu membantu proses pengelompokan calon penerima beasiswa, namun belum mempertimbangkan faktor sosial ekonomi mahasiswa sehingga hasil pengelompokan masih terbatas pada aspek akademik saja [9].

Berdasarkan penelitian-penelitian terdahulu, dapat diketahui bahwa penerapan *clustering* pada data pendidikan telah banyak dilakukan, namun masih terdapat beberapa keterbatasan. Sebagian penelitian hanya menggunakan satu algoritma *clustering* tanpa melakukan perbandingan performa metode lain. Selain itu, beberapa penelitian lebih berfokus pada variabel akademik tanpa mengombinasikannya dengan variabel sosial ekonomi yang sebenarnya memiliki pengaruh penting dalam proses seleksi beasiswa. Perbedaan metode evaluasi *cluster* juga menjadi salah satu kekurangan pada penelitian sebelumnya karena sebagian besar penelitian hanya menggunakan inerti atau *Davies-Bouldin Index* tanpa menggunakan *Silhouette Score* yang mampu mengukur tingkat kedekatan data dalam satu *cluster* dan pemisahan antar *cluster* secara lebih baik. Berdasarkan kondisi tersebut, terdapat *gap analysis* yang menunjukkan perlunya penelitian yang membandingkan algoritma *K-Means* dan *Hierarchical Clustering* menggunakan kombinasi data akademik dan sosial ekonomi dengan evaluasi menggunakan *Silhouette Score* untuk mengetahui metode yang memberikan kualitas *cluster* terbaik pada data calon penerima beasiswa.

Metode *K-Means* dipilih dalam penelitian ini karena memiliki kemampuan komputasi yang cepat dan efisien dalam mengelompokkan data berdasarkan kedekatan terhadap *centroid cluster*. Namun, metode ini memiliki kelemahan seperti sensitivitas terhadap pemilihan *centroid* awal dan keharusan menentukan jumlah *cluster* sejak awal proses [10]. Sementara itu, *Hierarchical Clustering* dipilih karena mampu membentuk struktur *cluster* secara bertahap tanpa harus menentukan jumlah *cluster* di awal serta dapat memberikan visualisasi hubungan antar data dalam bentuk dendrogram [7]. Dengan karakteristik yang berbeda tersebut, kedua algoritma memiliki potensi menghasilkan kualitas *cluster* yang berbeda pada data calon penerima beasiswa.

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *K-Means* dan *Hierarchical Clustering* dalam mengelompokkan calon penerima beasiswa berdasarkan data akademik dan sosial ekonomi menggunakan *Silhouette Score* sebagai metode evaluasi kualitas *cluster*. Variabel yang digunakan meliputi IPK, penghasilan orang tua, jumlah tanggungan keluarga, semester, dan nilai akademik mahasiswa. Melalui penelitian ini diharapkan dapat diperoleh metode *clustering* yang paling optimal dalam menghasilkan pengelompokan calon penerima beasiswa yang lebih akurat dan objektif. Selain itu, hasil penelitian diharapkan mampu memberikan kontribusi dalam mendukung pengambilan keputusan berbasis data pada proses penyaluran beasiswa sehingga bantuan pendidikan dapat diberikan secara lebih tepat sasaran, efektif, dan transparan.

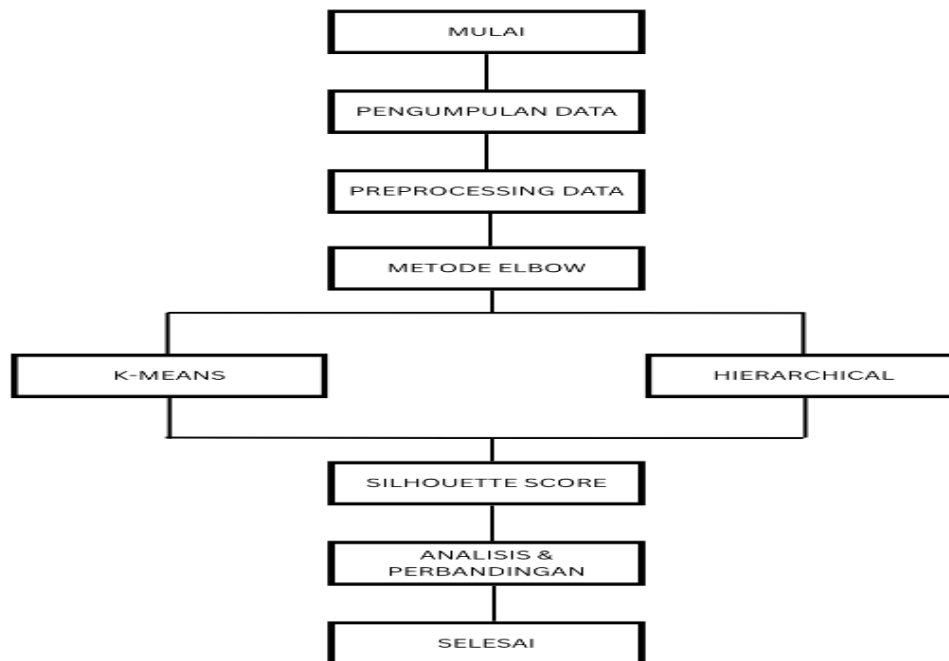
2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan data mining dengan teknik *clustering* untuk melakukan pengelompokan calon penerima beasiswa berdasarkan karakteristik data akademik dan sosial ekonomi. Metode yang digunakan dalam penelitian ini adalah algoritma *K-Means* dan *Hierarchical Clustering*, yang bertujuan untuk mengidentifikasi pola kemiripan antar data tanpa menggunakan label sebelumnya. Selanjutnya, hasil pengelompokan dievaluasi menggunakan *Silhouette Score* untuk menentukan kualitas *cluster* yang dihasilkan serta membandingkan kinerja kedua algoritma dalam menghasilkan pengelompokan yang optimal [2].

2.1. Prosedur Penelitian

Prosedur penelitian dilakukan secara sistematis mulai dari pengumpulan data calon penerima beasiswa dalam format *Excel/CSV* yang diperoleh dari instansi terkait. Tahap selanjutnya adalah *preprocessing* data yang meliputi seleksi fitur dan normalisasi menggunakan metode *standardization* pada atribut IPK, penghasilan orang tua, jumlah tanggungan, semester, dan nilai akademik.

Penentuan jumlah cluster optimal dilakukan menggunakan metode *Elbow* berdasarkan nilai inerti. Selanjutnya, proses *clustering* dilakukan menggunakan algoritma *K-Means* dan *Hierarchical Clustering (agglomerative)*. Hasil clustering kemudian dievaluasi menggunakan *Silhouette Score* untuk mengukur kualitas *cluster*. Tahap akhir adalah analisis dan perbandingan hasil untuk menentukan metode yang paling optimal dalam mengelompokkan calon penerima beasiswa [11]. Diagram prosedur penelitian seperti yang ditunjukkan pada gambar 1.



Gambar 1. Diagram Prosedur Penelitian

2.2. Rancangan Penelitian

Rancangan penelitian ini bersifat kuantitatif dengan pendekatan eksperimen, di mana dilakukan perbandingan dua algoritma *clustering* untuk mengetahui performa terbaik dalam mengelompokkan data calon penerima beasiswa berdasarkan karakteristik akademik dan sosial ekonomi. Proses penelitian dimulai dari pengumpulan data, *preprocessing*, penerapan algoritma *clustering*, hingga evaluasi hasil menggunakan *Silhouette Score* [4].

2.3. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data calon penerima beasiswa yang diperoleh dari bagian kemahasiswaan Universitas X dalam format Microsoft Excel (XLSX) yang kemudian dikonversi ke format CSV untuk proses analisis. Dataset terdiri dari 3212 data mahasiswa calon penerima beasiswa periode Tahun 2021–2025 yang memuat informasi terkait kondisi akademik dan sosial ekonomi mahasiswa. Untuk menjaga kerahasiaan dan privasi data, atribut yang bersifat identitas pribadi seperti NIM dan nama mahasiswa hanya digunakan sebagai data administrasi dan tidak dilibatkan dalam proses *clustering*. Seluruh data yang digunakan dalam analisis telah melalui proses anonimisasi sehingga tidak dapat digunakan untuk mengidentifikasi individu tertentu. Adapun atribut yang terdapat dalam dataset ditunjukkan pada Tabel 1.

Tabel 1. Atribut *dataset*

| No | Nama Atribut |
|----|--------------|
| 1 | NIM |
| 2 | Nama |

| | |
|---|-----------------------|
| 3 | Fakultas |
| 4 | Program Studi |
| 5 | IPK |
| 6 | Penghasilan Orang Tua |
| 7 | Jumlah Tanggungan |
| 8 | Semester |
| 9 | Nilai Akademik |

Dalam penelitian ini, hanya atribut yang memiliki relevansi terhadap proses seleksi beasiswa yang digunakan sebagai variabel utama dalam clustering, sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Variabel utama

| No | Variabel Utama | Deskripsi |
|----|-----------------------|---|
| 1 | IPK | Indikator prestasi akademik mahasiswa yang menunjukkan tingkat keberhasilan studi secara keseluruhan. |
| 2 | Penghasilan Orang Tua | Indikator kondisi ekonomi keluarga mahasiswa yang digunakan untuk mengukur tingkat kemampuan finansial orang tua. |
| 3 | Jumlah Tanggungan | Indikator beban ekonomi keluarga berdasarkan jumlah anggota keluarga yang menjadi tanggungan orang tua. |
| 4 | Semester | Menunjukkan tingkat atau jenjang studi mahasiswa pada saat pengajuan beasiswa. |
| 5 | Nilai Akademik | Indikator capaian akademik mahasiswa yang digunakan untuk melengkapi informasi prestasi selain IPK. |

Variabel tersebut meliputi IPK, penghasilan orang tua, jumlah tanggungan keluarga, semester, dan nilai akademik. Pemilihan variabel tersebut didasarkan pada kemampuannya dalam merepresentasikan dua aspek utama yang menjadi pertimbangan dalam pemberian beasiswa, yaitu prestasi akademik dan kondisi sosial ekonomi mahasiswa. Dengan demikian, hasil clustering diharapkan mampu menggambarkan karakteristik calon penerima beasiswa secara lebih komprehensif dan mendukung proses pengambilan keputusan yang lebih objektif [12][13].

2.4. Preprocessing Data

Tahap *preprocessing* data dilakukan untuk mempersiapkan data sebelum proses *clustering* sehingga menghasilkan analisis yang lebih optimal. Proses ini diawali dengan seleksi fitur dari dataset yang digunakan, di mana atribut yang dipilih adalah IPK, penghasilan orang tua, jumlah tanggungan, semester, dan nilai akademik. Pemilihan atribut tersebut didasarkan pada kemampuannya dalam merepresentasikan kondisi akademik serta latar belakang sosial ekonomi calon penerima beasiswa [14]. Setelah seleksi fitur dilakukan, tahap berikutnya adalah normalisasi data menggunakan metode *standardization* untuk menyamakan skala antar atribut. Normalisasi diperlukan karena perbedaan skala data dapat mempengaruhi perhitungan jarak pada algoritma *clustering*, khususnya yang berbasis jarak seperti *K-Means* dan *Hierarchical Clustering* [2] [15]. Proses normalisasi dilakukan menggunakan (1).

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

dimana Z merupakan nilai hasil normalisasi, X merupakan nilai data, μ adalah rata-rata, dan σ adalah standar deviasi. Dengan demikian, data yang telah melalui proses *preprocessing* menjadi lebih siap untuk digunakan pada tahap analisis berikutnya.

2.5. Penentuan Jumlah Cluster

Penentuan jumlah *cluster* optimal dilakukan menggunakan metode *Elbow*, yang bertujuan untuk menemukan jumlah *cluster* terbaik berdasarkan nilai inerti. Inerti merupakan ukuran yang menunjukkan seberapa dekat data dalam suatu *cluster* terhadap *centroid*-nya, sehingga nilai yang lebih kecil menunjukkan bahwa *cluster* semakin kompak. Proses ini dilakukan dengan menguji beberapa jumlah cluster dan menghitung nilai inerti untuk masing-masing nilai tersebut, kemudian divisualisasikan dalam bentuk grafik. Perhitungan inerti dapat dinyatakan dengan menggunakan (2).

$$Inertia = \sum_{i=1}^k \sum_{x \in C_i} (x - c_i)^2 \tag{2}$$

di mana C_i adalah *cluster* ke- i , c_i adalah *centroid* dari *cluster* tersebut, dan x merupakan data dalam *cluster*, *Inertia* adalah total variasi dalam cluster. Berdasarkan grafik yang dihasilkan, jumlah cluster optimal ditentukan dari titik perubahan yang signifikan atau yang dikenal sebagai *elbow point*, yaitu saat penurunan nilai inerti mulai

melandai. Penentuan jumlah *cluster* yang tepat sangat penting karena akan mempengaruhi kualitas hasil *clustering* yang dihasilkan [16].

2.6. Eksperimen *Clustering*

Tahap eksperimen *clustering* dilakukan dengan menerapkan dua algoritma, yaitu *K-Means* dan *Hierarchical Clustering*, untuk membandingkan performa keduanya dalam mengelompokkan data calon penerima beasiswa. Algoritma *K-Means* digunakan dengan pendekatan berbasis partisi, di mana data dikelompokkan ke dalam sejumlah *cluster* berdasarkan kedekatan terhadap *centroid*. Proses ini dilakukan secara iteratif dengan menghitung jarak antara data dan *centroid* menggunakan metode *Euclidean Distance*, kemudian memperbarui *centroid* hingga mencapai kondisi konvergen [2][17][18]. Perhitungan jarak *Euclidean* dinyatakan dengan (3).

$$d(x, y) = \sqrt{\sum(x_1 - y_1)^2} \quad (3)$$

dimana $d(x, y)$ merupakan jarak *Euclidean* antara dua objek, x_1 merupakan nilai atribut ke- i pada objek pertama, y_1 merupakan nilai atribut ke- i pada objek kedua, Sementara itu, *Hierarchical Clustering* diterapkan dengan pendekatan *agglomerative*, yaitu metode pengelompokan secara bertahap dari bawah ke atas, di mana setiap data awalnya dianggap sebagai *cluster* tersendiri, kemudian digabungkan dengan *cluster* lain yang memiliki jarak terdekat hingga terbentuk jumlah *cluster* yang diinginkan. Evaluasi hasil *clustering* dilakukan menggunakan *Silhouette Score* untuk mengukur kualitas *cluster* yang dihasilkan [12]. Nilai *Silhouette Score* dihitung menggunakan (4).

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

di mana $a(i)$ merupakan rata-rata jarak data terhadap *cluster* yang sama dan $b(i)$ merupakan jarak rata-rata terhadap *cluster* terdekat lainnya, $S(i)$ merupakan nilai *silhouette* untuk data ke- i . Nilai *Silhouette Score* yang mendekati 1 menunjukkan kualitas *cluster* yang baik dan terpisah dengan jelas[1]. Hasil dari kedua algoritma kemudian dibandingkan untuk mengetahui metode yang menghasilkan *cluster* paling optimal.

3. HASIL DAN PEMBAHASAN

3.1. Hasil *Preprocessing Data*

Tahap awal dalam penelitian ini dilakukan melalui proses *preprocessing* data untuk mempersiapkan *dataset* sebelum dilakukan proses *clustering*. Tahapan *preprocessing* sangat penting karena kualitas data akan memengaruhi hasil pengelompokan yang dihasilkan oleh algoritma *clustering*. Data yang belum diproses dengan baik dapat menyebabkan hasil *cluster* menjadi tidak optimal, terutama pada metode yang menggunakan perhitungan jarak seperti *K-Means* dan *Hierarchical Clustering*. Oleh karena itu, *preprocessing* dilakukan agar data menjadi lebih bersih, konsisten, dan memiliki skala yang seragam sehingga proses analisis dapat berjalan dengan lebih baik.

Pada penelitian ini, *preprocessing* data meliputi dua tahapan utama, yaitu seleksi fitur dan normalisasi data. Seleksi fitur dilakukan dengan memilih atribut yang dianggap paling relevan dalam merepresentasikan kondisi akademik dan sosial ekonomi mahasiswa calon penerima beasiswa. Adapun fitur yang digunakan dalam penelitian ini adalah IPK, penghasilan orang tua, jumlah tanggungan, semester, dan nilai akademik. Pemilihan atribut tersebut dilakukan karena variabel-variabel tersebut dianggap mampu menggambarkan tingkat kemampuan akademik mahasiswa sekaligus kondisi ekonomi keluarga yang menjadi faktor penting dalam proses seleksi beasiswa.

Atribut IPK digunakan untuk merepresentasikan prestasi akademik mahasiswa secara umum. Semakin tinggi nilai IPK, maka semakin baik performa akademik mahasiswa. Penghasilan orang tua digunakan sebagai indikator kondisi ekonomi keluarga mahasiswa, di mana mahasiswa dengan penghasilan orang tua yang rendah cenderung lebih membutuhkan bantuan pendidikan. Jumlah tanggungan digunakan untuk melihat beban ekonomi keluarga, sedangkan semester digunakan untuk mengetahui tingkat pendidikan mahasiswa saat ini. Nilai akademik digunakan sebagai indikator tambahan untuk melihat kemampuan akademik mahasiswa secara lebih spesifik.

Setelah proses seleksi fitur dilakukan, tahap berikutnya adalah normalisasi data menggunakan metode *Standard Scaler* atau *standardization*. Normalisasi dilakukan karena setiap atribut memiliki rentang nilai yang berbeda. Sebagai contoh, nilai IPK berada pada rentang 0 hingga 4, sedangkan penghasilan orang tua dapat berada pada rentang jutaan rupiah. Jika data tidak dinormalisasi, maka atribut dengan nilai terbesar akan lebih dominan dalam proses perhitungan jarak sehingga dapat menyebabkan bias pada hasil *clustering*. Oleh karena itu, normalisasi diperlukan agar seluruh atribut memiliki kontribusi yang seimbang dalam proses pengelompokan data.

Metode *standardization* bekerja dengan mengubah data sehingga memiliki rata-rata bernilai 0 dan standar deviasi bernilai 1. Dengan demikian, distribusi data menjadi lebih seragam dan proses *clustering* dapat

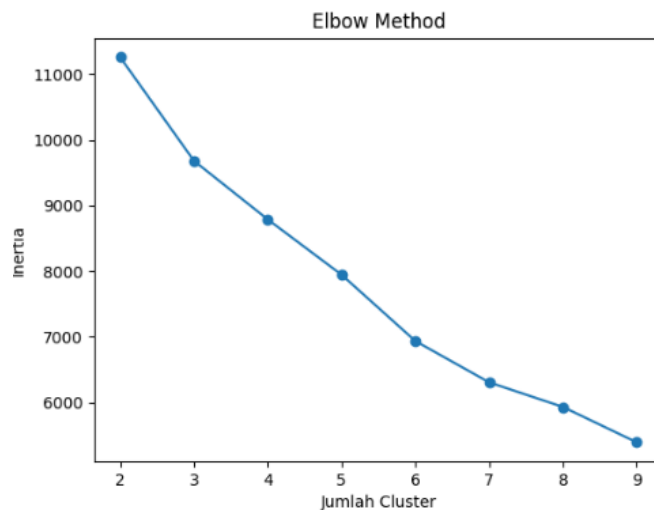
menghasilkan pengelompokan yang lebih optimal. Hasil *preprocessing* menunjukkan bahwa data telah berhasil dinormalisasi dengan baik sehingga siap digunakan pada tahap analisis berikutnya.

3.2. Penentuan Jumlah Cluster

Tahap berikutnya dalam penelitian ini adalah menentukan jumlah *cluster* optimal menggunakan metode *Elbow*. Penentuan jumlah cluster merupakan salah satu langkah penting dalam proses clustering karena jumlah *cluster* yang tidak tepat dapat menyebabkan hasil pengelompokan menjadi kurang optimal. Jika jumlah cluster terlalu sedikit, maka data yang memiliki karakteristik berbeda dapat tergabung dalam satu *cluster*. Sebaliknya, jika jumlah *cluster* terlalu banyak, maka *cluster* yang terbentuk menjadi terlalu spesifik dan sulit diinterpretasikan.

Metode *Elbow* digunakan untuk menentukan jumlah *cluster* terbaik berdasarkan nilai inerti. Inerti merupakan ukuran yang menunjukkan tingkat kedekatan data terhadap *centroid* dalam setiap *cluster*. Semakin kecil nilai inerti, maka semakin kompak *cluster* yang dihasilkan. Namun, penurunan nilai inerti tidak selalu signifikan ketika jumlah *cluster* terus ditambah. Oleh karena itu, metode *Elbow* digunakan untuk mencari titik perubahan yang signifikan atau disebut *elbow point* [13].

Pada penelitian ini, pengujian dilakukan dengan mencoba beberapa jumlah *cluster*, mulai dari $k=1$ hingga $k=10$. Setiap jumlah cluster menghasilkan nilai inerti yang kemudian divisualisasikan dalam bentuk grafik. Grafik tersebut menunjukkan hubungan antara jumlah *cluster* dengan nilai inerti yang dihasilkan oleh algoritma K-Means.



Gambar 2. Grafik *Elbow Method*

Berdasarkan hasil visualisasi pada Gambar 2, terlihat bahwa nilai inerti mengalami penurunan yang cukup signifikan dari $k=1$ hingga $k=4$. Setelah jumlah *cluster* mencapai 4, penurunan nilai inerti mulai melandai sehingga penambahan jumlah cluster di atas 4 tidak memberikan peningkatan kualitas cluster yang signifikan. Titik tersebut disebut sebagai *elbow point* atau titik siku.

Dengan demikian, jumlah *cluster* optimal yang digunakan dalam penelitian ini adalah sebanyak 4 *cluster*. Pemilihan jumlah *cluster* tersebut diharapkan mampu menghasilkan pengelompokan data mahasiswa calon penerima beasiswa yang lebih baik dan mudah diinterpretasikan.

3.3. Hasil Clustering Menggunakan K-Means

Setelah jumlah *cluster* optimal diperoleh, tahap berikutnya adalah penerapan algoritma *K-Means*. Algoritma ini digunakan untuk mengelompokkan data mahasiswa calon penerima beasiswa berdasarkan kemiripan karakteristik akademik dan sosial ekonomi. *K-Means* bekerja dengan membagi data ke dalam sejumlah cluster berdasarkan kedekatan jarak terhadap centroid cluster.

Pada penelitian ini, jumlah cluster yang digunakan adalah sebanyak 4 cluster sesuai dengan hasil metode *Elbow*. Proses *clustering* dilakukan secara iteratif, dimulai dengan menentukan *centroid* awal secara acak, kemudian menghitung jarak setiap data terhadap *centroid* menggunakan *Euclidean Distance*. Data akan ditempatkan pada *cluster* dengan *centroid* terdekat. Setelah seluruh data dikelompokkan, *centroid* diperbarui berdasarkan rata-rata data dalam setiap *cluster*. Proses ini terus dilakukan hingga posisi *centroid* stabil atau tidak mengalami perubahan signifikan [11].

Untuk memperoleh pemahaman yang lebih mendalam terhadap hasil *clustering*, dilakukan analisis karakteristik masing-masing cluster berdasarkan nilai rata-rata setiap variabel. Analisis ini bertujuan untuk

mengidentifikasi perbedaan karakteristik akademik dan sosial ekonomi antar kelompok mahasiswa calon penerima beasiswa [19].

Tabel 3. Karakteristik Cluster Hasil K-Means

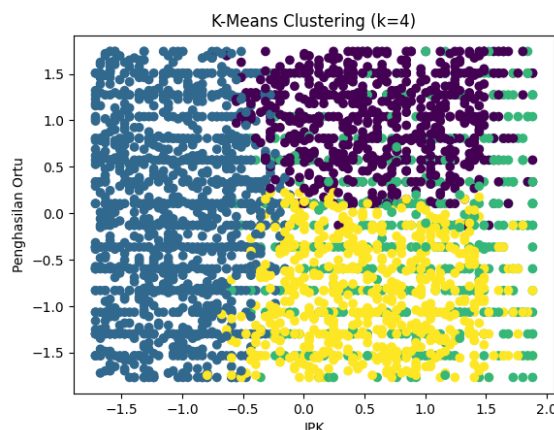
| Cluster | Jumlah Anggota | Rata-rata IPK | Rata-rata Penghasilan Orang Tua (RP) | Rata-rata Jumlah Tanggungan | Rata-rata Semester | Rata-rata Nilai Akademik |
|-----------|----------------|---------------|--------------------------------------|-----------------------------|--------------------|--------------------------|
| Cluster 0 | 712 | 3,68 | 3.178.005,62 | 3,82 | 5,03 | 3,68 |
| Cluster 1 | 1.248 | 3,27 | 2.746.907,05 | 3,25 | 4,51 | 3,27 |
| Cluster 2 | 594 | 3,73 | 2.688.771,04 | 1,96 | 2,57 | 3,73 |
| Cluster 3 | 658 | 3,68 | 2.373.647,42 | 3,86 | 5,60 | 3,68 |

Berdasarkan Tabel 3, setiap cluster menunjukkan karakteristik yang berbeda. Cluster 2 memiliki rata-rata IPK dan nilai akademik tertinggi, yaitu sebesar 3,73, dengan rata-rata penghasilan orang tua sebesar Rp2.688.771,04. Cluster 3 memiliki rata-rata penghasilan orang tua terendah, yaitu Rp2.373.647,42, serta jumlah tanggungan keluarga tertinggi sebesar 3,86. Sementara itu, Cluster 1 memiliki rata-rata IPK dan nilai akademik terendah dibandingkan cluster lainnya. Perbedaan karakteristik tersebut menunjukkan bahwa hasil clustering mampu mengelompokkan mahasiswa berdasarkan kombinasi faktor akademik dan sosial ekonomi.

Tabel 4. Interpretasi dan Prioritas Cluster

| Cluster | Karakteristik Utama | Interprestasi | Prioritas |
|-----------|---|---|---------------------------|
| Cluster 0 | IPK tinggi (3,68), semester menengah, tanggungan relatif tinggi | Mahasiswa berprestasi dengan kondisi ekonomi menengah | Tinggi (Sangat Layak) |
| Cluster 1 | IPK relatif rendah (3,27), ekonomi menengah, tanggungan sedang | Mahasiswa yang memenuhi sebagian kriteria beasiswa | Sedang (Layak) |
| Cluster 2 | IPK tertinggi (3,73), penghasilan orang tua rendah, semester awal | Mahasiswa berprestasi dengan kondisi ekonomi yang membutuhkan dukungan | Sedang (Dipertimbangkan) |
| Cluster 3 | Penghasilan orang tua terendah, tanggungan tertinggi, semester lanjut | Mahasiswa dengan kebutuhan ekonomi tinggi namun memerlukan pertimbangan akademik tambahan | Rendah (Kurang Prioritas) |

Berdasarkan Tabel 4, hasil clustering dapat digunakan sebagai dasar pengambilan keputusan dalam proses seleksi beasiswa. Cluster 0 dapat diprioritaskan sebagai kelompok penerima beasiswa karena memiliki capaian akademik yang baik dengan jumlah tanggungan keluarga yang relatif tinggi. Cluster 1 termasuk kelompok yang layak menerima beasiswa namun memiliki tingkat prioritas yang lebih rendah dibandingkan Cluster 0. Cluster 2 menunjukkan mahasiswa dengan prestasi akademik tertinggi dan kondisi ekonomi yang relatif membutuhkan sehingga dapat menjadi kelompok yang perlu dipertimbangkan sesuai kebijakan program beasiswa. Sementara itu, Cluster 3 memiliki kondisi ekonomi yang kurang mampu, namun perlu dilakukan evaluasi lebih lanjut terhadap aspek akademik sehingga ditempatkan pada prioritas yang lebih rendah. Hasil pengelompokan ini menunjukkan bahwa metode clustering dapat membantu proses seleksi beasiswa secara lebih objektif, transparan, dan berbasis data.



Gambar 3. Grafik K-Means Clustering

Gambar 3 menunjukkan hasil visualisasi *clustering* menggunakan algoritma *K-Means*. Sumbu horizontal (x) merepresentasikan nilai IPK sebagai indikator akademik, sedangkan sumbu vertikal (y) menunjukkan penghasilan orang tua sebagai indikator kondisi sosial ekonomi. Setiap titik pada grafik merepresentasikan satu data mahasiswa, sedangkan warna yang berbeda menunjukkan keanggotaan *cluster*.

Cara membaca grafik tersebut adalah dengan memperhatikan persebaran warna pada setiap titik. Titik dengan warna yang sama menunjukkan bahwa data tersebut berada dalam *cluster* yang sama dan memiliki karakteristik yang serupa. Semakin jelas pemisahan antar warna, maka semakin baik kualitas *cluster* yang dihasilkan.

Berdasarkan hasil visualisasi, terlihat bahwa algoritma *K-Means* mampu membentuk *cluster* yang relatif terpisah dengan baik. Setiap *cluster* menunjukkan kelompok mahasiswa dengan karakteristik tertentu. Berdasarkan karakteristik *cluster* yang diperoleh, *Cluster 2* memiliki rata-rata IPK dan nilai akademik tertinggi dibandingkan *cluster* lainnya, sementara rata-rata penghasilan orang tua berada pada kategori rendah serta jumlah tanggungan keluarga relatif tinggi. Karakteristik tersebut menunjukkan bahwa anggota *Cluster 2* memiliki prestasi akademik yang baik namun berasal dari keluarga dengan kondisi ekonomi yang kurang mampu. Oleh karena itu, *Cluster 2* dapat dianggap sebagai kelompok yang paling layak diprioritaskan sebagai calon penerima beasiswa.

Sebaliknya, *Cluster 1* memiliki rata-rata penghasilan orang tua yang lebih tinggi dibandingkan *cluster* lainnya meskipun memiliki capaian akademik yang relatif baik. Dengan kondisi tersebut, kelompok ini dapat dikategorikan sebagai prioritas yang lebih rendah dalam penyaluran bantuan pendidikan berbasis kebutuhan ekonomi. Sementara itu, *Cluster 0* dan *Cluster 3* menunjukkan karakteristik yang berada di antara kedua kelompok tersebut, sehingga dapat dipertimbangkan sebagai kelompok prioritas menengah sesuai dengan kebijakan dan kriteria beasiswa yang berlaku.

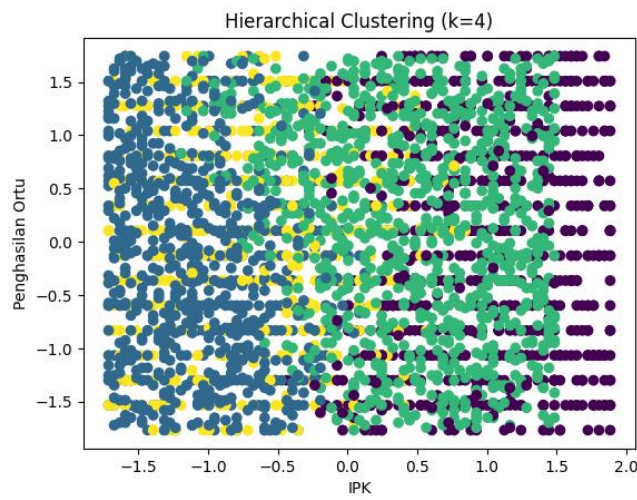
Selain itu, terdapat pula kelompok mahasiswa dengan IPK rendah dan penghasilan orang tua tinggi yang memiliki prioritas lebih rendah.

Hasil ini menunjukkan bahwa *K-Means* mampu mengidentifikasi pola kemiripan antar data secara efektif. Pengelompokan yang dihasilkan dapat membantu pihak pengelola beasiswa dalam memahami karakteristik mahasiswa calon penerima bantuan pendidikan secara lebih objektif.

3.4. Hasil *Clustering* Menggunakan *Hierarchical Clustering*

Selain menggunakan *K-Means*, penelitian ini juga menerapkan algoritma *Hierarchical Clustering* dengan jumlah *cluster* yang sama, yaitu sebanyak 4 *cluster*. *Hierarchical Clustering* merupakan metode *clustering* yang membentuk *cluster* secara bertahap berdasarkan kedekatan antar data.

Pada penelitian ini digunakan pendekatan *agglomerative*, yaitu metode yang dimulai dengan menganggap setiap data sebagai *cluster* tersendiri. Selanjutnya, *cluster* yang memiliki jarak terdekat akan digabungkan hingga terbentuk jumlah *cluster* yang diinginkan. Metode ini menghasilkan struktur hierarki yang dapat divisualisasikan dalam bentuk dendrogram.



Gambar 4. Grafik *Hierarchical Clustering*

Gambar 4 menunjukkan hasil *clustering* menggunakan *Hierarchical Clustering*. Sama seperti pada visualisasi sebelumnya, sumbu horizontal menunjukkan nilai IPK dan sumbu vertikal menunjukkan penghasilan orang tua. Warna yang berbeda menunjukkan hasil pengelompokan *cluster*.

Berdasarkan hasil visualisasi, terlihat bahwa *Hierarchical Clustering* juga mampu membentuk empat kelompok mahasiswa. Namun, dibandingkan dengan *K-Means*, terdapat beberapa *cluster* yang masih saling tumpang tindih sehingga pemisahan antar kelompok belum terlihat secara jelas. Hal ini menunjukkan bahwa kualitas *cluster* yang dihasilkan oleh *Hierarchical Clustering* masih relatif lebih rendah pada *dataset* ini.

Meskipun demikian, metode *Hierarchical Clustering* memiliki kelebihan dalam membentuk hubungan antar data secara bertahap sehingga dapat memberikan gambaran struktur data secara lebih detail. Namun, pada *dataset* dengan jumlah data yang cukup besar, metode ini cenderung memiliki kompleksitas komputasi yang lebih tinggi dibandingkan *K-Means*.

3.5. Evaluasi Menggunakan *Silhouette Score*

Tahap evaluasi dilakukan untuk mengukur kualitas *cluster* yang dihasilkan oleh masing-masing algoritma *clustering*. Evaluasi dalam penelitian ini menggunakan *Silhouette Score*, yaitu metode evaluasi internal yang digunakan untuk mengukur tingkat kedekatan data dalam satu *cluster* dan tingkat pemisahan antar *cluster* yang berbeda.

Nilai *Silhouette Score* berada pada rentang -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa data berada pada *cluster* yang tepat dan memiliki pemisahan yang jelas dengan *cluster* lainnya. Sebaliknya, nilai yang mendekati 0 menunjukkan bahwa *cluster* saling tumpang tindih, sedangkan nilai negatif menunjukkan bahwa data kemungkinan berada pada *cluster* yang salah. Hasil evaluasi menunjukkan bahwa:

- Nilai *Silhouette Score K-Means* sebesar 0,205
- Nilai *Silhouette Score Hierarchical Clustering* sebesar 0,181

```

Kolom: Index(['nim', 'nama', 'fakultas', 'prodi', 'ipk', 'penghasilan_ortu',
             'jumlah_tanggungan', 'semester', 'nilai_akademik'],
       dtype='str')

Rata-rata tiap cluster (K-Means):
      ipk  penghasilan_ortu  jumlah_tanggungan  semester  nilai_akademik
cluster_kmeans
0      3.682921      3.178006e+06      3.821629  5.030899      3.682921
1      3.266218      2.746907e+06      3.245192  4.514423      3.266218
2      3.731498      2.688771e+06      1.964646  2.569024      3.731498
3      3.675486      2.373647e+06      3.858663  5.595745      3.675486

Rata-rata tiap cluster (Hierarchical):
      ipk  penghasilan_ortu  jumlah_tanggungan  semester  nilai_akademik
cluster_hc
0      3.771085      2.762539e+06      3.275328  2.655542      3.771085
1      3.274243      2.685449e+06      3.844141  5.067122      3.274243
2      3.652072      2.810044e+06      3.519084  6.741549      3.652072
3      3.366568      2.763917e+06      1.946274  2.703640      3.366568

Silhouette Score K-Means: 0.20523624012851496
Silhouette Score Hierarchical: 0.18197885493361976
    
```

Gambar 5. Hasil Perbandingan *Cluster*

Berdasarkan hasil tersebut, algoritma *K-Means* memperoleh nilai *Silhouette Score* yang lebih tinggi dibandingkan *Hierarchical Clustering*. Hal ini menunjukkan bahwa *K-Means* mampu menghasilkan *cluster* yang lebih baik dan memiliki tingkat pemisahan antar *cluster* yang lebih jelas.

Meskipun nilai *Silhouette Score* yang diperoleh masih berada pada kategori lemah, hasil tersebut tetap menunjukkan bahwa *K-Means* memiliki performa yang lebih baik dibandingkan *Hierarchical Clustering* pada *dataset* calon penerima beasiswa yang digunakan dalam penelitian ini.

3.6. Analisis dan Pembahasan

Berdasarkan hasil pengujian yang telah dilakukan, kedua algoritma *clustering* mampu mengelompokkan data mahasiswa calon penerima beasiswa ke dalam empat *cluster* yang berbeda. Namun, kualitas *cluster* yang dihasilkan menunjukkan adanya perbedaan performa antara algoritma *K-Means* dan *Hierarchical Clustering*.

Algoritma *K-Means* memperoleh nilai *Silhouette Score* sebesar 0,205, sedangkan *Hierarchical Clustering* memperoleh nilai sebesar 0,181. Nilai tersebut menunjukkan bahwa *K-Means* menghasilkan *cluster* yang lebih kompak dan memiliki pemisahan antar *cluster* yang lebih baik dibandingkan *Hierarchical Clustering*.

Keunggulan *K-Means* pada penelitian ini dipengaruhi oleh karakteristik data yang bersifat numerik dan telah melalui proses normalisasi. Algoritma *K-Means* bekerja sangat baik pada data numerik karena proses pengelompokan dilakukan berdasarkan perhitungan jarak terhadap *centroid*. Selain itu, metode ini juga memiliki proses komputasi yang lebih cepat sehingga mampu menangani jumlah data yang lebih besar dengan lebih efisien.

Sementara itu, *Hierarchical Clustering* menunjukkan performa yang relatif lebih rendah karena metode ini lebih sensitif terhadap jarak antar data dan tidak melakukan optimasi *centroid* secara global seperti *K-Means*. Akibatnya, beberapa *cluster* yang terbentuk masih saling tumpang tindih sehingga kualitas pemisahan antar kelompok menjadi kurang optimal.

Dari sisi implementasi, hasil *clustering* memberikan informasi yang sangat penting dalam proses seleksi calon penerima beasiswa. Pengelompokan yang dihasilkan memungkinkan pihak pengelola beasiswa untuk mengidentifikasi karakteristik mahasiswa berdasarkan kombinasi aspek akademik dan sosial ekonomi secara lebih sistematis. Melalui hasil *clustering*, mahasiswa yang memiliki karakteristik serupa akan berada pada kelompok yang sama sehingga proses evaluasi dapat dilakukan secara lebih objektif dibandingkan hanya berdasarkan satu atau dua kriteria tertentu.

Sebagai contoh, *cluster* yang memiliki rata-rata IPK dan nilai akademik tinggi, namun didukung oleh rata-rata penghasilan orang tua yang rendah serta jumlah tanggungan keluarga yang tinggi, dapat dikategorikan sebagai

kelompok prioritas utama calon penerima beasiswa. Kelompok ini menunjukkan mahasiswa yang memiliki prestasi akademik baik sekaligus membutuhkan dukungan finansial. Sebaliknya, cluster dengan kondisi ekonomi yang relatif baik dan capaian akademik yang lebih rendah dapat ditempatkan pada prioritas yang lebih rendah sesuai dengan tujuan program beasiswa yang berorientasi pada pemerataan akses pendidikan.

Selain membantu proses identifikasi calon penerima beasiswa, hasil clustering juga dapat digunakan sebagai dasar penyusunan kebijakan penyaluran bantuan pendidikan yang lebih tepat sasaran. Pihak perguruan tinggi tidak hanya dapat mengetahui siapa yang layak menerima beasiswa, tetapi juga memahami karakteristik setiap kelompok mahasiswa sehingga alokasi bantuan dapat dilakukan secara lebih transparan, konsisten, dan berbasis data. Dengan demikian, penggunaan metode clustering mampu mengurangi subjektivitas dalam proses seleksi sekaligus meningkatkan efektivitas pengambilan keputusan dalam program beasiswa.

Hasil penelitian ini menunjukkan bahwa penggunaan teknik *data mining*, khususnya *clustering*, mampu membantu proses pengelompokan calon penerima beasiswa secara lebih objektif dan berbasis data. Selain itu, penelitian ini juga membuktikan bahwa pemilihan algoritma *clustering* sangat memengaruhi kualitas hasil pengelompokan yang dihasilkan.

Dengan demikian, algoritma *K-Means* lebih direkomendasikan untuk digunakan dalam proses pengelompokan calon penerima beasiswa berdasarkan data akademik dan sosial ekonomi karena mampu menghasilkan kualitas *cluster* yang lebih baik dibandingkan *Hierarchical Clustering*. Penelitian selanjutnya diharapkan dapat menggunakan *dataset* yang lebih besar, menambahkan variabel lain yang relevan, serta membandingkan metode *clustering* lainnya agar diperoleh hasil yang lebih optimal.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *K-Means* dan *Hierarchical Clustering* mampu digunakan untuk melakukan pengelompokan calon penerima beasiswa berdasarkan data akademik dan sosial ekonomi mahasiswa. Proses *clustering* dilakukan menggunakan atribut IPK, penghasilan orang tua, jumlah tanggungan, semester, dan nilai akademik yang dianggap mampu merepresentasikan kondisi akademik serta latar belakang ekonomi mahasiswa. Sebelum proses *clustering* dilakukan, data terlebih dahulu melalui tahap *preprocessing* berupa seleksi fitur dan normalisasi menggunakan *Standard Scaler* agar seluruh atribut memiliki skala yang seragam dan tidak memengaruhi proses perhitungan jarak. Hasil pengujian menggunakan metode *Elbow* menunjukkan bahwa jumlah *cluster* optimal pada penelitian ini adalah sebanyak 4 cluster. Penerapan algoritma *K-Means* dan *Hierarchical Clustering* menghasilkan pengelompokan mahasiswa dengan karakteristik yang berbeda-beda, sehingga dapat memberikan gambaran mengenai kelompok mahasiswa berdasarkan tingkat prestasi akademik dan kondisi sosial ekonomi. Berdasarkan hasil visualisasi *clustering*, algoritma *K-Means* menunjukkan pemisahan cluster yang lebih jelas dibandingkan *Hierarchical Clustering*. Evaluasi menggunakan *Silhouette Score* menunjukkan bahwa algoritma *K-Means* memperoleh nilai sebesar 0,205, sedangkan *Hierarchical Clustering* memperoleh nilai sebesar 0,181. Hasil tersebut menunjukkan bahwa *K-Means* menghasilkan kualitas cluster yang relatif lebih baik dibandingkan *Hierarchical Clustering* pada dataset yang digunakan dalam penelitian ini. Meskipun demikian, nilai *Silhouette Score* yang diperoleh oleh kedua metode masih tergolong rendah, sehingga hasil clustering yang dihasilkan belum menunjukkan pemisahan cluster yang sangat kuat. Oleh karena itu, temuan penelitian ini menunjukkan bahwa *K-Means* lebih unggul secara relatif dibandingkan *Hierarchical Clustering* berdasarkan evaluasi internal yang digunakan, namun belum dapat disimpulkan sebagai metode yang paling optimal untuk seluruh kasus pengelompokan calon penerima beasiswa. Penelitian ini menunjukkan bahwa teknik clustering dapat dimanfaatkan sebagai alat bantu dalam mengidentifikasi kelompok mahasiswa berdasarkan karakteristik akademik dan sosial ekonomi secara lebih sistematis. Hasil pengelompokan yang diperoleh dapat digunakan sebagai informasi pendukung dalam proses pengambilan keputusan calon penerima beasiswa sehingga proses seleksi dapat dilakukan secara lebih objektif dan berbasis data. Namun demikian, keputusan akhir tetap perlu mempertimbangkan kebijakan dan kriteria beasiswa yang berlaku pada masing-masing institusi. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. *Dataset* yang digunakan masih terbatas pada jumlah data dan atribut tertentu sehingga hasil *clustering* belum sepenuhnya menggambarkan kondisi yang lebih kompleks. Selain itu, penelitian ini hanya membandingkan dua algoritma *clustering*, yaitu *K-Means* dan *Hierarchical Clustering*, serta menggunakan satu metode evaluasi berupa *Silhouette Score*. Oleh karena itu, penelitian selanjutnya diharapkan dapat menggunakan *dataset* yang lebih besar, menambahkan variabel lain yang lebih relevan, serta membandingkan metode *clustering* lainnya seperti *DBSCAN*, *Fuzzy C-Means*, atau *Gaussian Mixture Model* agar diperoleh hasil pengelompokan yang lebih optimal dan akurat.

REFERENCES

- [1] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm for Big Data," *Phys. Procedia*, vol. 78, no. December 2015, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.
- [2] D. Xu and Y. Tian, "A Comprehensive Survey of Clustering Algorithms," *Ann. Data Sci.*, vol. 2, no. 2, pp. 165–193,

- 2015, doi: 10.1007/s40745-015-0040-1.
- [3] S. Brina Winona Rizky, "Perbandingan Algoritma K-Means Dan Hierarchical Clustering Dalam Segmentasi Kabupaten/Kota Di Jawa Timur Berdasarkan Data Perjalanan Dan Pergerakan Wisatawan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 5, pp. 8499–8506, 2025, doi: 10.36040/jati.v9i5.15091.
- [4] T. Iklima and A. Pujiyanta, "Perbandingan Metode K-Means Clustering Dan Metode Ward Dalam Mengelompokkan Pelangan Mall," *J. Fasilkom*, vol. 13, no. 3, pp. 349–357, 2023, doi: 10.37859/jf.v13i3.6040.
- [5] M. A. Pakpahan, S. Sahid, M. M. F. Simanullang, and R. Putra, "Analisis Hierarchical Clustering untuk Segmentasi Pelanggan pada Dataset Mall Customers," vol. 4, no. 1, pp. 58–63, 2026.
- [6] N. Butarbutar, A. P. Windarto, D. Hartama, and S. Solikhun, "Komparasi Kinerja Algoritma Fuzzy C-Means Dan K-Means Dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai Akademik Siswa," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 1, no. 1, p. 46, 2017, doi: 10.30645/jurasik.v1i1.8.
- [7] J. Han, J. Pei, and H. Tong, *Data Mining Concepts and Techniques*, 4th ed. Katey Birtcher, 2022.
- [8] D. N. Cahya, "Impelentasi Algoritma K-Means Clustering untuk Menentukan Calon Penerima Beasiswa UKT di Universitas Sains Al-Qur'an," 2022.
- [9] Sindrawati, D. Syaripudin, and A. R. Raharja, "Penerapan Algoritma K-Means Clustering pada Data Nilai Siswa Untuk Menentukan Kelompok Penerima Beasiswa," *SisInfo; J. Sist. Inf. dan Inform.*, vol. 6, no. 2, pp. 47–55, 2024, [Online]. Available: <https://doi.org/10.37278/sisinfo.v6i2.900>
- [10] P. Fieguth, *An introduction to pattern recognition and machine learning*. 2022. doi: 10.1007/978-3-030-95995-1.
- [11] N. Asyifa and M. S. Novelan, "Performance Analysis of the K-Means Algorithm in Classifying Organic and Inorganic Waste Types," *Proc. Int. Conf. Islam. Community Stud.*, pp. 6681–6690, 2025, [Online]. Available: <https://proceeding.pancabudi.ac.id/index.php/ICIE/article/view/1284>
- [12] Y. Hasan, "Pengukuran Silhouette Score Dan Davies-Bouldin Index Pada Hasil Cluster K-Means Dan DbSCAN," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, 2024, doi: 10.23960/jitet.v12i3s1.5001.
- [13] S. Sutiono, R. F. Wijaya, and M. S. Novelan, "Analisis Metode K-Means Clustering Untuk Pengelompokan Jadwal Mengajar Guru Di SMKN 1 Stabat Dengan Visualisasi Data," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 8, no. 2, pp. 596–602, 2025, doi: 10.31539/intecomsv8i2.15154.
- [14] F. Murtagh and P. Contreras, "Methods of Hierarchical Clustering," in *Data Mining and Knowledge Discovery*, Hoboken, NJ, USA: Wiley-Interscience, 2011.
- [15] A. J. Sidauruk, J. S. Sirait, and S. Sipayung, "Penerapan Normalisasi Data pada Angkatan Kerja Indonesia Bulan Februari 2025 Berdasarkan Kelompok Umur," vol. 4, no. 1, 2026, doi: 10.54259/jdmis.v4i1.7023.
- [16] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.
- [17] B. Y. Putra, F. Y. Azzahra, and I. A. Erlanda, "Klasterisasi Pengunjung Mall Menggunakan Algoritma K-Means Berdasarkan Pendapatan Dan Pengeluaran," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, 2023, doi: 10.23960/jitet.v11i3s1.3392.
- [18] Z. Sitorus, I. Syahputra, C. Indra Angkat, and D. Sartika, "Implementation of K-Means Clustering for Inventory Projection," *Int. J. Sci. Technol. Manag.*, vol. 5, no. 3, pp. 673–678, 2024, doi: 10.46729/ijstm.v5i3.856.
- [19] I. P. Permatasari, A. F. Cobantoro, and D. Mustikasari, "Penggunaan Algoritma K-means untuk Menentukan Calon Penerima Beasiswa," *Forum Pendidik. Tinggi Tek. Elektro Indones. Reg. VII*, pp. 524–533, 2024.