JDMSI, Vol. 7, No. 1, 2026, 20-29, ISSN: 2748-8485

Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Program Studi Berdasarkan Daya Tampung UTBK 2019

Khoirun Nisa[1)], Yoga Alaudin Zuansyah[2)], Elkin Rilvani[3)], Indra Permana[4)]

[1234]Program Studi Bisnis Digital, Universitas Pelita Bangsa
[1234]Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Kabupaten Bekasi, Jawa Barat
Email: [1]grunzz.21@gmail.com, [2]yogaalaudinzuansyah@gmail.com,
[3]elkin.rilvani@pelitabangsa.ac.id, [4]indrapermana@pelitabangsa.ac.id

Abstract

The rapid growth of educational data requires appropriate analytical methods to extract useful information and support decision-making processes. Data mining is one of the approaches used to discover hidden patterns from large datasets. Clustering is a data mining technique that groups data based on similarities in their characteristics. This study aims to apply the K-Means Clustering algorithm to classify study programs based on admission capacity using the 2019 UTBK dataset. The dataset consists of 3,167 study programs from various universities in Indonesia. The research was conducted using RapidMiner software by utilizing the K-Means algorithm with K = 3 clusters. The clustering process resulted in three groups consisting of 2,507 data in Cluster 0, 616 data in Cluster 2, and 44 data in Cluster 1. The centroid values obtained were 37.914, 92.440, and 229.864 respectively. The results indicate that most study programs belong to the low-capacity category, while only a small number of study programs have very high admission capacities.

Keyword: *Data Mining, Clustering, K-Means, UTBK, Admission Capacity*

Abstrak

Perkembangan data pendidikan yang semakin pesat memerlukan metode analisis yang tepat untuk menghasilkan informasi yang bermanfaat dan mendukung pengambilan keputusan. Salah satu pendekatan yang dapat digunakan adalah *data mining*, yaitu proses menemukan pola atau informasi tersembunyi dari sekumpulan data berukuran besar. *Clustering* merupakan salah satu teknik *data mining* yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik yang dimiliki. Penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* dalam mengelompokkan program studi berdasarkan daya tampung menggunakan *dataset* UTBK tahun 2019. Dataset yang digunakan terdiri dari 3.167 program studi dari berbagai perguruan tinggi di Indonesia. Penelitian dilakukan menggunakan aplikasi *RapidMiner* dengan jumlah *cluster* sebanyak tiga kelompok (K=3). Hasil *clustering* menunjukkan bahwa *cluster* 0 terdiri dari 2.507 data, *cluster* 2 terdiri dari 616 data, dan *cluster* 1 terdiri dari 44 data. Nilai *centroid* yang diperoleh masing-masing sebesar 37,914; 92,440; dan 229,864. Hasil penelitian menunjukkan bahwa sebagian besar program studi memiliki daya tampung rendah, sedangkan hanya sebagian kecil program studi yang memiliki daya tampung sangat tinggi.

Kata Kunci: *Data Mining, K-Means, Clustering,* UTBK, Daya Tampung

1. PENDAHULUAN

Perkembangan teknologi informasi telah menghasilkan data dalam jumlah yang sangat besar pada berbagai bidang, termasuk bidang pendidikan. Data yang tersimpan dalam jumlah besar memiliki potensi untuk menghasilkan informasi yang bermanfaat apabila diolah menggunakan metode yang tepat. Salah satu pendekatan yang digunakan untuk mengolah data tersebut adalah *data mining*. *Data mining* adalah metode pengolahan data yang menemukan pola, hubungan, dan pengetahuan tersembunyi dalam sekumpulan data untuk digunakan dalam proses pengambilan keputusan [1]. Dalam bidang pendidikan, *data mining* telah banyak diterapkan untuk melakukan analisis data akademik, evaluasi pendidikan, serta pengelompokan data berdasarkan karakteristik tertentu [2], [3].

Salah satu teknik yang paling sering digunakan dalam *data mining* adalah *clustering*. *Clustering* merupakan proses pengelompokan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik yang dimiliki. Data yang berada dalam satu kelompok memiliki tingkat kemiripan yang lebih tinggi dibandingkan dengan data pada kelompok lainnya [7]. Karena prosesnya yang sederhana dan mudah digunakan, algoritma *K-Means* menjadi salah satu metode *clustering* yang paling populer. Selain itu, algoritma ini memiliki kemampuan untuk mengolah sejumlah besar data dengan waktu komputasi yang cukup cepat [10]. Selain itu, beberapa penelitian juga membandingkan performa *K-Means* dengan metode *clustering* lainnya seperti *Fuzzy C-Means* dan menunjukkan bahwa *K-Means* tetap memiliki performa yang baik dalam proses pengelompokan data pendidikan [4], [5].

Studi sebelumnya menunjukkan bahwa algoritma *K-Means* dapat digunakan dalam pendidikan. Dengan menggunakan algoritma *K-Means*, Wibowo dan Habanabakize mengelompokkan standar kualifikasi pendidikan dan pengalaman kerja guru SMK di Indonesia, menghasilkan beberapa kelompok dengan karakteristik yang berbeda [2]. Hasil penelitian tersebut menunjukkan bahwa algoritma *K-Means* mampu membantu proses analisis data pendidikan secara efektif dan menghasilkan informasi yang berguna dalam mendukung pengambilan keputusan.

Salah satu data pendidikan yang menarik untuk dianalisis adalah data daya tampung program studi pada Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) atau UTBK. Daya tampung program studi mencerminkan kapasitas penerimaan mahasiswa baru pada suatu program studi dan dapat digunakan untuk melihat distribusi kapasitas pendidikan pada perguruan tinggi di Indonesia. Dengan jumlah program studi yang sangat banyak, diperlukan metode yang mampu mengelompokkan data secara otomatis berdasarkan karakteristik daya tampung yang dimiliki.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* dalam mengelompokkan program studi berdasarkan daya tampung menggunakan dataset UTBK tahun 2019. Penelitian dilakukan menggunakan aplikasi *RapidMiner* dengan jumlah *cluster* sebanyak tiga kelompok. Hasil penelitian diharapkan dapat memberikan gambaran tentang distribusi daya tampung program studi di Indonesia dan menjadi referensi untuk menggunakan teknik *data mining* dalam pendidikan.

2. METODE PENELITIAN

2.1 Dataset Penelitian

Penelitian ini menggunakan dataset daya tampung program studi UTBK tahun 2019 yang diperoleh dari data penerimaan mahasiswa baru perguruan tinggi di Indonesia. Dataset terdiri dari 3.167 data program studi yang berasal dari berbagai perguruan tinggi negeri. Data tersebut memiliki beberapa atribut yaitu *id_university*, *id_major*, *major_name*, *type*, dan *capacity*. Dalam penelitian ini, atribut yang digunakan sebagai dasar pengelompokan adalah atribut *capacity* atau daya tampung program studi. Pemilihan atribut tersebut dilakukan karena tujuan penelitian adalah mengelompokkan program studi berdasarkan kapasitas penerimaan mahasiswa baru yang dimiliki. Atribut lainnya hanya digunakan sebagai identitas data dan tidak dilibatkan dalam proses *clustering*.

Tabel 1. Atribut dataset UTBK

Atribut	Keterangan
id_university	Identitas Perguruan Tinggi
id_major	Identitas Program Studi
major_name	Nama Program Studi
type	Jenis Program Studi
capacity	Daya Tampung Program Studi

Dataset yang digunakan memiliki karakteristik numerik pada atribut *capacity* sehingga sesuai untuk diterapkan pada algoritma *K-Means* yang membutuhkan atribut numerik dalam proses perhitungan jarak antar data [7].

2.2 Tahapan Penelitian

Penelitian ini dilakukan dalam beberapa langkah, mulai dari pengumpulan data hingga analisis hasil *clustering*. Berikut adalah langkah-langkah yang diambil dalam penelitian ini.:

1. Pengumpulan *Dataset*
2. *Preprocessing Data*
3. Seleksi Atribut
4. Proses *Clustering* Menggunakan *K-Means*
5. Analisis Hasil *Clustering*
6. Penarikan Kesimpulan

Tahapan tersebut dilakukan secara sistematis agar proses *clustering* menghasilkan kelompok data yang sesuai dengan karakteristik daya tampung program studi [1].

2.3 Preprocessing Data

Tahap *preprocessing* dilakukan untuk memastikan kualitas data yang digunakan dalam penelitian. Data yang berkualitas akan menghasilkan proses *clustering* yang lebih baik dan akurat [9]. Pada tahap ini dilakukan pemeriksaan terhadap:

a. Kelengkapan data (*missing value*)

b. Konsistensi data
c. Tipe data atribut
d. Data duplikat

Setelah proses pemeriksaan selesai dilakukan, atribut yang digunakan dalam penelitian dipilih menggunakan operator *Select Attributes* pada aplikasi *RapidMiner*. Atribut *capacity* dipilih sebagai atribut utama karena menjadi fokus penelitian, sedangkan atribut lainnya tidak digunakan dalam proses perhitungan *clustering*. Tahapan preprocessing merupakan langkah penting dalam *data mining* karena berpengaruh terhadap kualitas hasil analisis yang diperoleh [10].

2.4 Algoritma K-Means

K-Means merupakan algoritma *clustering* yang termasuk ke dalam metode *unsupervised learning*. Algoritma ini bekerja dengan mengelompokkan data berdasarkan tingkat kemiripan karakteristik yang dimilikinya. [7]. Memaksimalkan perbedaan antar *cluster* dan jarak antar data dalam satu *cluster* adalah tujuan utama algoritma *K-Means* [1]. Data dibagi ke dalam sejumlah *cluster* menggunakan metode *partitional clustering K-Means* berdasarkan jarak terdekat terhadap *centroid*. Keunggulan algoritma ini adalah proses komputasinya yang relatif cepat dan mudah diimplementasikan pada berbagai jenis data numerik. Namun demikian, hasil *clustering* sangat dipengaruhi oleh jumlah *cluster* yang digunakan serta pemilihan *centroid* awal [7], [10]. Karena prosesnya sederhana dan mudah digunakan, metode ini banyak digunakan. Metode ini juga mampu mengolah data berukuran besar dengan efisien [4], [5].

Langkah-langkah algoritma K-Means adalah sebagai berikut:

1. Menentukan jumlah *cluster* (K).
2. Menentukan *centroid* awal secara acak.
3. Menghitung jarak setiap data terhadap *centroid.*
4. Menempatkan data pada *cluster* dengan jarak terdekat.
5. Menghitung ulang nilai *centroid* berdasarkan rata-rata anggota *cluster.*
6. Mengulangi proses hingga *centroid* stabil dan tidak terjadi perubahan *cluster.*

Perhitungan jarak dilakukan menggunakan metode *Euclidean Distance* sebagai berikut:

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

Keterangan:

a. d(x,y) = jarak antara dua data
b. xi = nilai data ke-i
c. yi = nilai *centroid* ke-i
d. n = jumlah atribut

Dalam penelitian ini digunakan nilai K = 3 untuk menghasilkan tiga kelompok program studi berdasarkan tingkat daya tampung yaitu rendah, sedang, dan tinggi.

2.5 Implementasi Menggunakan RapidMiner

Proses *clustering* pada penelitian ini dilakukan menggunakan aplikasi *RapidMiner*. *RapidMiner* merupakan salah satu perangkat lunak *data mining* yang banyak digunakan untuk proses analisis data karena menyediakan berbagai operator yang mendukung

implementasi algoritma *machine learning* dan *data mining* [3]. Tahapan implementasi pada *RapidMiner* dilakukan dengan menggunakan beberapa operator sebagai berikut:

1. *Retrieve*
   Digunakan untuk mengambil *dataset* yang telah diimpor ke dalam *repository RapidMiner*.
2. *Set Role*
   Digunakan untuk menentukan atribut identitas sehingga tidak memengaruhi *proses clustering*.
3. *Select Attributes*
   Digunakan untuk memilih atribut *capacity* sebagai atribut utama penelitian.
4. *K-Means*
   Digunakan untuk melakukan proses pengelompokan data berdasarkan nilai daya tampung.

Nilai parameter K pada operator *K-Means* ditentukan sebanyak tiga *cluster* sesuai tujuan penelitian. Setelah proses dijalankan, *RapidMiner* menghasilkan informasi berupa jumlah anggota setiap *cluster*, nilai *centroid*, dan distribusi data pada masing-masing cluster [2], [6]. Hasil *clustering* yang diperoleh kemudian dianalisis untuk mengetahui karakteristik setiap kelompok program studi berdasarkan kapasitas daya tampung yang dimiliki.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Clustering

Proses *clustering* dilakukan menggunakan algoritma *K-Means* dengan jumlah *cluster* sebanyak tiga kelompok (K=3). Pengelompokan dilakukan berdasarkan atribut *capacity* yang menunjukkan daya tampung program studi pada *dataset* UTBK tahun 2019. Setelah proses *clustering* dijalankan menggunakan *RapidMiner*, diperoleh tiga *cluster* dengan jumlah anggota yang berbeda.

Tabel 2. Hasil Clustering Program Studi

Cluster	Jumlah Data
Cluster 0	2507
Cluster 1	44
Cluster 2	616

Berdasarkan Tabel 2 dapat diketahui bahwa sebagian besar program studi berada pada *cluster 0* dengan jumlah anggota sebanyak 2.507 data atau sekitar 79,16% dari keseluruhan dataset. *Cluster 2* memiliki 616 data atau sekitar 19,45%, sedangkan *cluster 1* hanya terdiri dari 44 data atau sekitar 1,39%. Perbedaan jumlah anggota pada setiap *cluster* menunjukkan bahwa distribusi daya tampung program studi di Indonesia tidak merata. Sebagian besar program studi memiliki daya tampung yang relatif rendah, sedangkan program studi dengan daya tampung sangat tinggi jumlahnya relatif sedikit.

3.2 Analisis Nilai Centroid

Nilai *centroid* digunakan untuk mengetahui karakteristik masing-masing *cluster*

berdasarkan rata-rata daya tampung program studi yang menjadi anggota *cluster* tersebut.

Tabel 3. Nilai Centroid Setiap Cluster

Cluster	Jumlah Data
Cluster 0	37,914
Cluster 1	229,864
Cluster 2	92,440

Berdasarkan Tabel 3 terlihat bahwa *Cluster 1* memiliki nilai *centroid* tertinggi yaitu 229,864. Hal tersebut menunjukkan bahwa anggota *cluster* tersebut memiliki kapasitas penerimaan mahasiswa yang jauh lebih besar dibandingkan *cluster* lainnya. Sebaliknya, *Cluster 0* memiliki nilai *centroid* terendah yaitu 37,914 sehingga dapat dikategorikan sebagai kelompok program studi dengan daya tampung rendah.

3.3 Karakteristik Setiap Cluster

Berdasarkan hasil *clustering* yang diperoleh, masing-masing *cluster* memiliki karakteristik yang berbeda berdasarkan nilai *centroid* dan jumlah anggota *cluster*. Nilai *centroid* digunakan sebagai representasi rata-rata daya tampung program studi pada setiap kelompok. Semakin tinggi nilai *centroid*, maka semakin besar kapasitas penerimaan mahasiswa yang dimiliki oleh program studi dalam *cluster* tersebut. Oleh karena itu, hasil *clustering* dapat diinterpretasikan menjadi tiga kategori utama, yaitu daya tampung rendah, daya tampung sedang, dan daya tampung tinggi. Analisis karakteristik masing-masing *cluster* disajikan sebagai berikut.

3.3.1 Cluster 0 (Daya Tampung Rendah)

Cluster 0 memiliki jumlah anggota sebanyak 2.507 program studi dengan nilai *centroid* sebesar 37,914. *Cluster* ini merupakan kelompok terbesar dalam penelitian. Nilai *centroid* yang rendah menunjukkan bahwa program studi pada *cluster* ini memiliki kapasitas penerimaan mahasiswa yang relatif kecil. Dominasi jumlah anggota pada *cluster* ini menunjukkan bahwa sebagian besar program studi di Indonesia memiliki daya tampung yang terbatas. Kondisi tersebut dapat dipengaruhi oleh berbagai faktor seperti keterbatasan ruang kuliah, jumlah tenaga pengajar, kapasitas laboratorium, maupun kebijakan akademik masing-masing perguruan tinggi. Hasil ini menunjukkan bahwa mayoritas program studi masih menerapkan kapasitas penerimaan mahasiswa dalam jumlah yang relatif sedikit dibandingkan program studi lainnya.

3.3.2 Cluster 2 (Daya Tampung Sedang)

Cluster 2 terdiri dari 616 program studi dengan nilai *centroid* sebesar 92,440. Program studi yang termasuk dalam *cluster* ini memiliki kapasitas penerimaan mahasiswa yang lebih besar dibandingkan *cluster 0* namun belum mencapai kategori daya tampung tinggi. Kelompok ini dapat dikategorikan sebagai program studi dengan daya tampung sedang. Jumlah anggota *cluster* yang cukup banyak menunjukkan bahwa terdapat sejumlah program studi yang memiliki kemampuan menerima mahasiswa dalam jumlah lebih besar dibandingkan rata-rata program studi lainnya. Program studi pada kelompok ini

umumnya memiliki sumber daya yang lebih memadai sehingga mampu menampung mahasiswa dalam jumlah yang lebih besar.

3.3.3 Cluster 1 (Daya Tampung Tinggi)

Cluster 1 merupakan kelompok dengan jumlah anggota paling sedikit yaitu sebanyak 44 program studi. Namun demikian, *cluster* ini memiliki nilai *centroid* tertinggi sebesar 229,864. *Nilai centroid* yang tinggi menunjukkan bahwa program studi pada *cluster* ini memiliki kapasitas penerimaan mahasiswa yang sangat besar. Jumlah anggota yang sedikit mengindikasikan bahwa hanya sebagian kecil program studi yang memiliki daya tampung sangat tinggi dibandingkan program studi lainnya.Program studi yang berada pada *cluster* ini umumnya berasal dari perguruan tinggi dengan fasilitas yang lebih lengkap, jumlah tenaga pengajar yang memadai, serta tingkat peminat yang tinggi sehingga mampu menyediakan kapasitas penerimaan mahasiswa yang lebih besar.

3.4 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *K-Means* berhasil mengelompokkan program studi berdasarkan karakteristik daya tampung yang dimiliki. Perbedaan nilai *centroid* yang cukup signifikan antar *cluster* menunjukkan bahwa proses *clustering* mampu memisahkan data sesuai dengan tingkat kapasitas penerimaan mahasiswa. Studi ini sejalan dengan penelitian Liu [1] yang menemukan bahwa algoritma *K-Means* dapat menghasilkan pengelompokan data pendidikan yang efektif berdasarkan atribut tertentu. Melalui proses *clustering*, data yang sebelumnya sulit dianalisis dapat dikelompokkan menjadi beberapa kategori yang lebih mudah dipahami dan diinterpretasikan. Hasil penelitian ini juga mendukung penelitian Wibowo dan Habanabakize [2] yang berhasil mengelompokkan standar kualifikasi pendidikan dan pengalaman kerja guru SMK menggunakan algoritma *K-Means*. Pada kedua penelitian tersebut, *K-Means* mampu membentuk kelompok data yang memiliki karakteristik berbeda sehingga mempermudah proses analisis dan pengambilan keputusan.

Penelitian terkait analisis data akademik mahasiswa yang dilakukan pada lingkungan perguruan tinggi juga menunjukkan bahwa algoritma *K-Means* efektif digunakan untuk mengelompokkan data pendidikan berdasarkan kesamaan karakteristik [3]. Hal ini menunjukkan bahwa metode *K-Means* memiliki fleksibilitas yang tinggi dan dapat diterapkan pada berbagai jenis data pendidikan. Selain itu, penelitian Ardan [4] dan Apriyanto [5] yang membandingkan algoritma *K-Means* dengan *Fuzzy C-Means* menunjukkan bahwa kedua metode mampu menghasilkan pengelompokan data yang baik, namun *K-Means* memiliki keunggulan dari sisi kesederhanaan proses dan kemudahan implementasi. Oleh karena itu, penggunaan *K-Means* pada penelitian ini dinilai sesuai untuk mengelompokkan program studi berdasarkan daya tampung. Penelitian Basalamah dan Setyadi [6] mengenai tingkat penyelesaian pendidikan juga menghasilkan beberapa kelompok data yang berbeda berdasarkan karakteristik yang dimiliki. Hasil tersebut memperkuat bahwa teknik *clustering* dapat dimanfaatkan untuk mendukung analisis data pendidikan dalam berbagai aspek.

Proses *clustering* yang dilakukan pada penelitian ini menggunakan prinsip dasar algoritma *K-Means* sebagaimana dijelaskan dalam berbagai penelitian sebelumnya [7], [9], [10]. Metode algoritma mencapai *centroid* yang stabil dengan menghitung jarak antara data dan *centroid*, kemudian mengelompokkan data ke dalam *cluster* yang terletak lebih dekat satu sama lain. Penelitian mengenai gaya belajar mahasiswa yang dilakukan oleh

Risnasari dkk. [8] juga menunjukkan kemampuan algoritma *K-Means* untuk menghasilkan kelompok data dengan berbagai karakteristik, yang dapat digunakan sebagai dasar untuk membangun strategi pembelajaran. Semakin banyak bukti bahwa algoritma *K-Means* dapat membantu pengambilan keputusan berbasis data di bidang pendidikan. Berdasarkan hasil penelitian dan perbandingan dengan penelitian terdahulu, dapat disimpulkan bahwa algoritma *K-Means* mampu mengelompokkan program studi berdasarkan daya tampung secara efektif. Hasil *clustering* memberikan gambaran yang lebih jelas mengenai distribusi kapasitas penerimaan mahasiswa pada program studi di Indonesia sehingga dapat dimanfaatkan sebagai informasi pendukung dalam proses evaluasi dan perencanaan pendidikan.

3.4 Implikasi Hasil Penelitian

Hasil *clustering* yang diperoleh pada penelitian ini memberikan gambaran mengenai distribusi daya tampung program studi di Indonesia berdasarkan data UTBK tahun 2019. Melalui proses pengelompokan menggunakan algoritma *K-Means*, program studi dapat diklasifikasikan ke dalam kategori daya tampung rendah, sedang, dan tinggi. Informasi tersebut dapat membantu dalam memahami karakteristik kapasitas penerimaan mahasiswa pada berbagai program studi. Bagi perguruan tinggi, hasil pengelompokan ini dapat digunakan sebagai bahan evaluasi dalam perencanaan kapasitas penerimaan mahasiswa baru. Program studi yang berada pada kelompok daya tampung rendah dapat melakukan evaluasi terhadap ketersediaan sumber daya yang dimiliki, seperti jumlah dosen, ruang kuliah, maupun fasilitas pendukung pembelajaran. Sementara itu, program studi dengan daya tampung tinggi dapat menjadi acuan dalam pengelolaan jumlah mahasiswa agar kualitas pembelajaran tetap terjaga.

Selain itu, hasil penelitian ini juga dapat memberikan informasi tambahan bagi calon mahasiswa yang ingin melanjutkan pendidikan ke perguruan tinggi. Daya tampung program studi dapat menjadi salah satu pertimbangan dalam menentukan pilihan program studi yang sesuai dengan minat dan peluang penerimaan yang tersedia. Penelitian ini menunjukkan dari perspektif akademik bahwa algoritma *K-Means* dapat diterapkan pada data pendidikan untuk menghasilkan informasi yang lebih terstruktur dan lebih mudah dipahami. Oleh karena itu, metode *clustering* dapat menjadi salah satu alternatif dalam mendukung proses analisis data pendidikan serta membantu pengambilan keputusan berbasis data pada lingkungan perguruan tinggi.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *K-Means Clustering* untuk mengelompokkan program studi berdasarkan daya tampung menggunakan dataset UTBK tahun 2019. *Dataset* yang digunakan terdiri dari 3.167 program studi dari berbagai perguruan tinggi di Indonesia dengan atribut utama berupa *capacity* atau daya tampung program studi. Proses pengolahan data dilakukan menggunakan aplikasi *RapidMiner* dengan jumlah *cluster* yang ditentukan sebanyak tiga kelompok (K=3). Hasil *clustering* menunjukkan kemampuan algoritma *K-Means* untuk mengelompokkan data ke dalam tiga *cluster* yang masing-masing memiliki atribut unik. *Cluster 0* terdiri dari 2.507 program studi dengan nilai *centroid* sebesar 37,914 yang menunjukkan kelompok program studi dengan daya tampung rendah. *Cluster 2* terdiri dari 616 program studi dengan nilai *centroid* sebesar 92,440 yang merepresentasikan kelompok program studi dengan daya tampung

sedang. Sementara itu, *Cluster 1* terdiri dari 44 program studi dengan nilai *centroid* sebesar 229,864 yang menunjukkan kelompok program studi dengan daya tampung tinggi.

Berdasarkan hasil penelitian dapat diketahui bahwa sebagian besar program studi di Indonesia memiliki kapasitas penerimaan mahasiswa yang relatif rendah. Hanya sebagian kecil program studi yang memiliki kapasitas penerimaan sangat tinggi. Perbedaan nilai *centroid* yang cukup signifikan antar *cluster* menunjukkan bahwa algoritma *K-Means* mampu memisahkan data berdasarkan karakteristik daya tampung yang dimiliki secara efektif. Hasil penelitian ini juga menunjukkan bahwa teknik *clustering* dapat digunakan sebagai salah satu pendekatan dalam analisis data pendidikan. Pengelompokan program studi berdasarkan daya tampung dapat memberikan gambaran mengenai distribusi kapasitas penerimaan mahasiswa pada perguruan tinggi di Indonesia. Informasi tersebut dapat dimanfaatkan sebagai bahan pertimbangan dalam proses evaluasi maupun perencanaan pendidikan.

Dengan demikian, algoritma *K-Means* dapat menjadi salah satu metode yang efektif untuk mengolah dan menganalisis data pendidikan dalam jumlah besar. Untuk penelitian selanjutnya, disarankan menggunakan atribut tambahan seperti jumlah peminat, tingkat keketatan, akreditasi program studi, maupun jenis perguruan tinggi sehingga hasil c*lustering* yang diperoleh dapat memberikan informasi yang lebih komprehensif.

UCAPAN TERIMAKASIH

Puji syukkur Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan sukses. Penulis berterima kasih kepada dosen pengampu mata kuliah *Data Mining* yang telah membantu, membimbing, dan memberi masukan selama proses penelitian. Selain itu, kami mengucapkan terima kasih kepada semua orang yang telah membantu dalam proses pengumpulan data, proses pengolahan data menggunakan aplikasi *RapidMiner*, dan juga orang-orang yang telah membantu menyusun artikel ilmiah ini. Kami berharap temuan penelitian ini akan berkontribusi pada kemajuan ilmu pengetahuan, khususnya di bidang penggalian data dan analisis data pendidikan.

REFERENCES

[1] R. Liu, "Data Analysis of Educational Evaluation Using K-Means Clustering Method," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–10, 2022.

[2] A. E. Wibowo and T. Habanabakize, "K-Means Clustering untuk Klasifikasi Standar Kualifikasi Pendidikan dan Pengalaman Kerja Guru SMK di Indonesia," *Jurnal Dinamika Vokasional Teknik Mesin*, vol. 7, no. 2, pp. 124–132, 2022.

[3] A. M. Fauzi, "Penerapan Data Mining Menggunakan Algoritma K-Means pada Data Akademik Mahasiswa," Repository UIN K.H. Abdurrahman Wahid Pekalongan, 2022.

[4] M. Ardan, "Perbandingan Algoritma K-Means dan Fuzzy C-Means dalam Klasterisasi Data Pendidikan di Wilayah Jakarta," *Jurnal Komputer dan Informatika*, vol. 13, no. 1, pp. 1–10, 2024.

[5] J. Apriyanto, "Penerapan Clustering pada Rata-rata Lama Sekolah di Provinsi Indonesia Menggunakan Algoritma K-Means dan Fuzzy C-Means," *Jurnal Komputer dan Informatika*, vol. 13, no. 1, pp. 11–20, 2024.

[6] A. Basalamah and R. Setyadi, "Penerapan Algoritma K-Means Clustering pada Tingkat Penyelesaian Pendidikan di Provinsi Indonesia," *JICOM: Journal of Informatics and Computer Science*, vol. 4, no. 2, pp. 95–103, 2023.

[7] A. A. Alhassan, M. A. Alenezi, and A. A. Alreshidi, "A Comprehensive Study of K-Means Clustering Algorithm and Its Applications," *Procedia Computer Science*, vol. 220, pp. 1048–1055, 2022.

[8] M. Risnasari, N. Aulia, and L. Cahyani, "Clustering of Student Learning Styles in the Industry 4.0 Using K-Means Algorithm," *Jurnal Teknologi Pendidikan*, vol. 24, no. 2, pp. 120–129, 2022.

[9] R. Rahmadani, A. Fitri, and M. Syahputra, "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means," *Jurnal Sains dan Teknologi (SAINTEK)*, vol. 4, no. 2, pp. 57–65, 2020.

[10] D. Pratama and A. Nugroho, "Implementasi Algoritma K-Means Clustering untuk Pengelompokan Data," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 88–97, 2023.