

Clustering Prestasi Akademik Lulusan Menggunakan Metode K-Means

Rezqiwati Ishak
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
rezqi.uig@gmail.com

Amiruddin*
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
amier.76@gmail.com*

Diterima : September 2023
Disetujui : Januari 2024
Dipublikasi : Januari 2024

Abstrak—Prestasi akademik merupakan salah satu indikator penting untuk mengukur keberhasilan seorang mahasiswa dalam menyelesaikan studinya di perguruan tinggi. Prestasi ini dapat dilihat dari berbagai aspek, seperti lama studi dan Indeks Prestasi Kumulatif (IPK). Analisis ini digunakan untuk meningkatkan kualitas Pendidikan pada Perguruan Tinggi itu sendiri, serta untuk membantu Mahasiswa dalam mencapai prestasi yang optimal. Penelitian ini bertujuan untuk melakukan *clustering* prestasi akademik lulusan pada Universitas Ichsan Gorontalo untuk Tahun Akademik 2023/2024 semester Ganjil dengan menerapkan metode *K-Means*. Jumlah dataset lulusan yang digunakan sebanyak 240 data. Analisis *clustering* dilakukan berdasarkan atribut lama studi, umur, dan Indeks Prestasi Kumulatif (IPK). Hasil penelitian ini menunjukkan adanya 3 (tiga) *cluster* utama. *Cluster 1* (satu) merupakan kelompok lulusan dengan prestasi akademik cukup baik, terdiri dari 56 lulusan. *Cluster 2* (dua) menggambarkan kelompok lulusan dengan prestasi akademik sangat baik, terdiri dari 138 lulusan. Sementara itu, *Cluster 3* (tiga) menunjukkan kelompok lulusan dengan prestasi akademik kurang baik jika dilihat dari lama studi, terdiri dari 45 lulusan. Pemilihan jumlah *cluster* sebanyak 3 didasarkan pada hasil perhitungan teknik *Elbow* dan evaluasi *Davies-Bouldin Index* yang memberikan nilai terkecil yakni 0,79 sehingga hasil *clustering* masuk kategori baik karena nilai DBInya di bawah 1.

Kata Kunci—*Clustering; K-Means; Elbow; Prestasi Akademik Lulusan; Davies-Bouldin Index.*

Abstract — *Academic achievement is one of the important indicators to measure a student's success in completing their studies at the university. This achievement can be observed from various aspects, such as the duration of study and the Cumulative Grade Point Average (GPA). This analysis is used to improve the quality of education at the university itself and to assist students in achieving optimal performance. This research aims to cluster the academic achievements of graduates at Ichsan Gorontalo University for the Academic Year 2023/2024 Odd Semester using the K-Means method. The number of graduate datasets used is 240. The clustering analysis is based on attributes such as the duration of study, age, and Cumulative Grade Point Average (GPA). The results of this study indicate the existence of 3 main clusters. Cluster 1 represents graduates with fairly good academic achievements, consisting of 56 graduates. Cluster 2 describes a group of graduates with excellent academic achievements, totaling 138 graduates. Meanwhile, Cluster 3 indicates a group of*

graduates with less satisfactory academic achievements when considering the duration of study, consisting of 45 graduates. The selection of 3 clusters is based on the results of the Elbow technique calculation and the evaluation of the Davies-Bouldin Index, which gives the smallest value of 0.79. Therefore, the clustering results are considered good because the DBI value is below 1.

Keywords— *Clustering; K-Means; Elbow; Graduate Academic Performance; Davies-Bouldin Index.*

I. PENDAHULUAN

Prestasi akademik lulusan perguruan tinggi merupakan parameter yang sangat penting dalam mengukur kualitas pendidikan dan kontribusi suatu lembaga terhadap pembentukan sumber daya manusia berkualitas. Setiap perguruan tinggi dituntut untuk bisa memenuhi Indikator Kinerja Utama (IKU) Perguruan Tinggi yang sudah ditetapkan oleh Pemerintah khususnya IKU 1 (satu) yaitu kesiapan kerja lulusan[1].

Proses analisis dalam memetakan prestasi akademik lulusan pada suatu perguruan tinggi dapat menggunakan teknik data mining yaitu metode *clustering*. Salah satu metode *clustering* yang populer adalah metode K-Means. Dimana metode K-Means bekerja dengan membagi data menjadi k kelompok, dimana setiap kelompok memiliki karakteristik yang sama. Kelompok-kelompok ini dibentuk berdasarkan jarak antar data. Data yang memiliki jarak yang dekat akan dikelompokkan dalam satu kelompok yang sama[2],[3].

Penggunaan metode K-Means untuk melakukan *clustering* prestasi akademik, ini sudah digunakan oleh beberapa peneliti sebelumnya diantaranya [4],[5],[6]. Penelitian [4] menggunakan 3 atribut yaitu jalur masuk perguruan tinggi, pendapatan orang tua dan indeks prestasi kumulatif dengan hasil *clustering* sebanyak 3 *cluster*. Penelitian [5] proses *clustering* menggunakan 8 atribut yaitu jurusan, daerah asal, umur, jenis kelamin, Indeks Prestasi Kumulatif (IPK), Satuan Kredit Semester (SKS), tahun masuk dan lama studi dengan menghasilkan 4 *cluster*. Penelitian [6] menggunakan 6 atribut dalam proses *clustering* yaitu IPS1 sampai dengan IPS6 dengan hasil *clustering* sebanyak 2 *cluster*.

Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada jenis dan jumlah atribut yang digunakan serta tujuan analisa *clustering* yang ingin dicapai. Penelitian [4] tujuan analisisnya yaitu klasterisasi kinerja akademik Mahasiswa. Penelitian [5] tujuan analisisnya adalah pengelompokan pola masa studi Mahasiswa, dan penelitian [6] tujuan analisisnya adalah pengelompokan Indeks Prestasi Akademik Mahasiswa sedangkan pada penelitian ini tujuan analisisnya adalah *clustering* prestasi akademik lulusan.

Metode K-Means juga memiliki kelemahan diantaranya penentuan jumlah *cluster* yang optimum tidak bisa ditentukan secara otomatis, sehingga diperlukan metode lain dalam penentuannya yaitu menggunakan teknik *Elbow*. Beberapa penelitian sebelumnya yang sudah menerapkan metode *K-Means* dengan *Elbow* dalam peningkatan kinerja metode tersebut [7],[8],[9],[10], ini membuktikan bahwa teknik *Elbow* sangat diperlukan dalam metode K-Means agar jumlah kluster bisa langsung ditentukan.

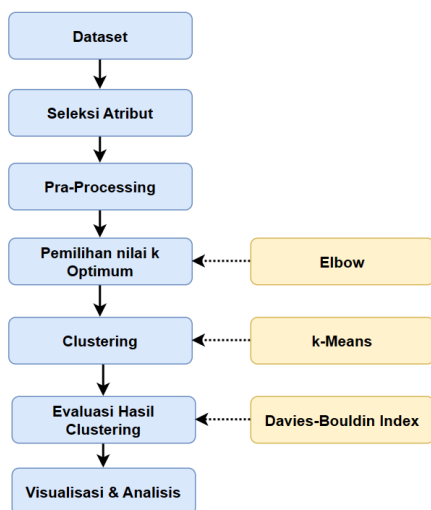
Dalam penelitian ini untuk mendapatkan jumlah *cluster* yang paling ideal digunakan teknik evaluasi *Davies-Bouldin Index (DBI)*. Teknik ini juga sudah dibuktikan oleh beberapa peneliti sebelumnya [11],[12],[13],[14], sehingga dengan demikian penerapan teknik *Elbow* dan evaluasi *Davies-Bouldin Index (DBI)* menjadikan metode K-Means dalam proses *clustering* lebih baik, efisien dan efektif.

Tujuan dari penelitian ini adalah mengelompokkan prestasi akademik lulusan pada Universitas Ichsan Gorontalo yang lulus pada periode Tahun Akademik 2023/2024 semester Ganjil dengan menggunakan metode *K-Means* berdasarkan atribut lama studi, umur, dan IPK. Atribut tersebut dipilih untuk memudahkan dalam analisa hasil *clustering* yang ditinjau dari kerkaitan ketiga atribut tersebut. Hasil penelitian ini diharapkan dapat memberikan informasi tentang pola prestasi akademik lulusan demi peningkatan pelayanan dan kualitas pendidikan serta untuk membantu mahasiswa dalam mencapai prestasi akademik yang optimal.

II. METODE

A. Tahapan Penelitian

Tahapan penelitian yang digunakan ditunjukkan pada Gambar 1:



Gambar 1. Tahapan Penelitian

1. Dataset : pada tahapan ini menentukan dataset yang akan digunakan yaitu data lulusan pada Universitas Ichsan

Gorontalo Tahun Akademik 2023/2024 Semester Ganjil sebanyak 240 Data

2. Seleksi Atribut : pada tahapan ini dilakukan pemilihan atribut yang akan digunakan.
3. *Pra-Processing* : pada tahapan normalisasi nilai atribut perlu dilakukan agar memiliki nilai *range* yang sama.
4. Pemilihan nilai K optimum : pada tahap ini digunakan teknik *Elbow* untuk menentukan nilai K atau jumlah *cluster* optimum.
5. *Clustering* : pada tahap ini dilakukan proses *Clustering* menggunakan metode *K-Means*
6. Evaluasi Hasil *Clustering* : pada tahap digunakan algoritma *Davies-Bouldin Index* untuk Evaluasi hasil *clustering* yang ideal.
7. Visualisasi dan Analisis: Pada tahap ini, hasil kelompok divisualisasikan dan dianalisis untuk membuat kesimpulan tentang hasil kelompok masing-masing.

B. Metode *Elbow*

Metode *elbow* merupakan teknik dalam analisis kluster untuk menentukan jumlah kluster yang optimal dalam sebuah kumpulan data. Nama "*elbow*" (siku) berasal dari bentuk grafik hasil analisis, yang mirip dengan siku pada lengan. Tujuan utama metode *elbow* adalah menemukan titik di mana penambahan kluster tambahan memberikan penurunan yang berkurang secara signifikan dalam jumlah variasi atau inersia dalam data.[15],[16],[10].

Berikut tahapan dalam metode *Elbow*:

1. Pilih Rentang Jumlah Kluster (k): Tentukan rentang nilai k yang mungkin. misalnya memulai dari 1 hingga 10.
2. Terapkan K-Means: Gunakan algoritma K-Means untuk setiap nilai k dalam rentang yang telah ditentukan. Ini melibatkan pengelompokan data menjadi k kluster dan menghitung inersia untuk setiap pengelompokan.
3. Hitung Inersia: Inersia diukur sebagai jumlah kuadrat jarak *Euclidean* antara setiap titik data dan pusat klusternya. Hitung inersia untuk setiap nilai k.
4. Plot *Elbow Curve*: Buat grafik dengan sumbu x adalah nilai k dan sumbu y adalah inersia yang dihitung.
5. Analisis *Elbow Curve*: pada grafik. Cari titik di mana penurunan inersia mulai melambat dan grafik menyerupai siku. Titik ini adalah "*elbow*" dan mencerminkan jumlah kluster optimal.
6. Pilih Jumlah Kluster yang Optimal: Berdasarkan analisis *elbow curve*, pilih jumlah kluster yang optimal untuk diterapkan pada kasus yang akan di kluster.
7. Terapkan K-Means dengan Jumlah Kluster Optimal: Gunakan jumlah kluster yang optimal yang telah dipilih untuk melatih model k-means pada seluruh data. Ini akan menjadi jumlah kluster yang memberikan keseimbangan terbaik antara pemisahan data dan kompleksitas model.

C. *K-Means Clustering*

K-Means adalah algoritma pengelompokan (*clustering*) yang digunakan dalam analisis data untuk memisahkan titik-titik data ke dalam kelompok-kelompok atau kluster berdasarkan kemiripan karakteristik. Algoritma ini mencoba meminimalkan varian dalam setiap kluster dan memaksimalkan varian antar kluster.[3],[17].

Secara umum, langkah-langkah dalam algoritma K-Means dapat dirinci sebagai berikut:

1. Pemilihan Jumlah Kluster (k): tentukan jumlah kluster yang diinginkan berdasarkan hasil jumlah kluster yang optimal pada metode *Elbow*.
2. Inisialisasi Pusat Kluster: pilih secara acak k titik dari data sebagai pusat awal kluster.
3. Pengelompokan (*Assignment*): Hitung jarak antara setiap titik data dengan pusat kluster. Tentukan kluster untuk setiap titik berdasarkan pusat kluster terdekat menggunakan metrik jarak tertentu (misalnya, jarak *Euclidean* atau jarak *Manhattan*).

$$d_{ik} = \sqrt{\sum_{i=1}^n (C_{ij} - C_{kj})^2} \quad (1)$$

dimana:

C_{ij} : pusat *cluster*

C_{kj} : data

4. Pembaruan Pusat Kluster: hitung ulang pusat kluster sebagai rata-rata dari semua titik data dalam kluster.
5. Iterasi: ulangi langkah-langkah 3 dan 4 hingga tidak ada atau hanya sedikit perubahan dalam pengelompokan dan pusat kluster.

D. Davies-Bouldin Index

Davies-Bouldin Indeks (DBI) adalah metrik evaluasi yang digunakan dalam analisis kluster untuk mengukur seberapa baik kluster telah dipisahkan satu sama lain. Tujuan utama DBI adalah mengukur keberagaman antara kluster dan memastikan bahwa setiap kluster memiliki pusat yang berbeda satu sama lain. Nilai DBI yang lebih rendah menunjukkan hasil yang lebih baik[15].

Langkah-langkah penghitungan DBI sebagai berikut:

1. Hitung *Similitude* (Kesamaan) Antar Kluster: hitung nilai kesamaan antara setiap pasangan kluster. Kesamaan diukur dengan rasio antara jarak antara pusat kluster (*centroid*) dan ukuran kluster. Semakin kecil nilai *similitude*, semakin baik kluster tersebut terpisah.
2. Hitung DBI untuk setiap kluster: untuk setiap kluster, pilih kluster lain yang memiliki *similitude* terbesar. Misalnya, untuk kluster i, cari kluster j sehingga nilai kesamaan antara i dan j adalah maksimum.

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (2)$$

dimana:

- R_{ij} : nilai R untuk *cluster* i dan j.
 - d_{ij} : jarak antara *centroid cluster* i dan j.
 - S_i dan S_j : jarak rata-rata antara setiap titik cluster i dan *centroid* dari *cluster* j
3. Hitung nilai akhir DBI, dimana nilai DBI tersebut merupakan rata-rata dari nilai DBI untuk setiap *cluster*. Rumusnya adalah:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max R_{ij, i \neq j} \quad (3)$$

dimana :

- R_{ij} : nilai R untuk *cluster* i dan j.
- k : jumlah *cluster*

III. HASIL DAN PEMBAHASAN

Implementasi metode K-Means dalam proses *clustering* prestasi akademik lulusan dilakukan berdasarkan tahapan

penelitian yang telah diuraikan sebelumnya, dimulai dari pembacaan dataset, seleksi atribut, pra-processing hingga visualisasi hasil *clustering*. Proses ini dilakukan dengan menggunakan *Google Colaboratory* sebagai platform dan menggunakan bahasa pemrograman *Python*. Pada proses ini, digunakan pula beberapa library seperti *Numpy*, *Pandas*, *Seaborn*, *Matplotlib*, dan *Scikit Learn* untuk proses analisis dan implementasi algoritma *clustering* yakni K-Means.

A. Pengumpulan Dataset

Dataset yang digunakan bersumber dari data lulusan pada Universitas Ichsan Gorontalo Tahun Akademik 2023/2024 Semester Ganjil sebanyak 240 data[18]. Sampel dataset yang digunakan ditunjukkan pada Tabel 1.

Tabel 1. Sampel Dataset

No	NIM	Lama Studi	Umur	IPK
1	E1117116	6.3	21	3,84
2	E1117139	6.4	24	3,3
3	E1117145	5.0	29	3,03
4	E1118010	5.3	23	3,29
5	E1118037	5.3	24	3,52
6	E1118046	5.3	23	3,54
7	E1118052	5.3	22	3,36
8	E1119007	4.3	26	3,63
9	E1119010	4.2	22	3,43
10	E1119012	4.2	21	3,42
11	E1119013	4.2	22	3,8
12	E1119017	4.2	23	3,68
13	E1119018	4.2	23	3,86
14	E1119020	4.2	23	3,69
15	E1119021	4.2	23	3,68

B. Seleksi Atribut

Berdasarkan dataset pada Tabel 1, dilakukan seleksi atribut yang akan digunakan dalam proses *clustering* yaitu lama studi, umur dan IPK sedangkan atribut NIM tidak digunakan karena tidak berpengaruh ke hasil *clustering*.

C. Pra-Processing

Pra-processing yang dilakukan pada tahapan ini adalah melakukan normalisasi data dari ketiga nilai atribut yaitu lama studi, umur dan IPK agar didapatkan nilai *ranges* yang sama sehingga hasil *clustering* bisa lebih optimal. Berikut potongan kode program untuk melakukan normalisasi data yang ditunjukkan pada Gambar 2 :

```
# Normalization
from sklearn import preprocessing
min_max_scaler = preprocessing.MinMaxScaler()
X2 = min_max_scaler.fit_transform(x1)
X2
```

Gambar 2. Kode Program Normalisasi Data

Setelah dijalankan kode program pada Gambar 2, maka akan didapatkan hasil normalisasi data yang ditunjukkan pada Gambar 3:

```
array([[0.95833333, 0.35294118, 0.
[1.
[0.91666667, 0.17647059, 0.47747748],
[0.95833333, 0.17647059, 0.28828829],
[0.95833333, 0.17647059, 0.20720721],
[0.91666667, 0.
[0.95833333, 0.17647059, 0.44144144],
[0.375
[0.5
[0.5
[0.5
[0.08333333, 0.29411765, 0.73873874],
[0.04166667, 0.05882353, 0.55855856],
[0.04166667, 0.
, 0.54954955],
```

Gambar 3. Sampel Hasil Normalisasi Data

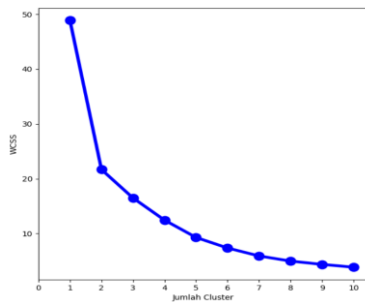
D. Pemilihan Nilai k Optimum

Penentuan nilai k yang optimum pada pemodelan K-Means dapat ditentukan dengan menggunakan salah satu metode yaitu metode *Elbow*. Berikut potongan kode program untuk penggunaan metode *Elbow* :

```
# Menentukan jumlah cluster yang optimum dengan teknik Elbow
x=X2
WCSS = []
for i in range(1,11):
    model = KMeans(n_clusters = i, init = "k-means++",
max_iter = 300, n_init = 10, random_state = 0)
    model.fit(x)
    WCSS.append(model.inertia_)
fig = plt.figure(figsize = (7,7))
plt.plot(range(1,11),WCSS, linewidth=4, markersize=12,marker='o',color = 'blue')
plt.xticks(np.arange(11))
plt.xlabel("Jumlah Cluster")
plt.ylabel("WCSS")
plt.show()
```

Gambar 4. Kode Program *Elbow*

Hasil dari kode program pada Gambar 4, hasilnya ditunjukkan pada Gambar 5:



Gambar 5. Visualisasi Hasil Metode *Elbow*

Berdasarkan Gambar 5, terlihat bahwa posisi jumlah cluster yang optimal terletak pada nilai 3 (tiga) cluster. Hal ini dapat dilihat dari titik pada grafik di mana penurunan inersia tidak lagi signifikan setelah titik jumlah cluster 3 (tiga). Oleh karena itu, nilai $k = 3$ dijadikan sebagai nilai optimal dalam pemodelan K-Means pada *clustering* prestasi akademik lulusan di penelitian ini.

E. Proses Clustering dengan K-Means

Setelah didapatkan nilai k yang optimum menggunakan metode *Elbow*. Selanjutnya dilakukan pemodelan menggunakan algoritma K-Means. Dimana Proses pemodelan ini dilakukan dengan menggunakan bahasa pemrograman Python dan memanfaatkan *library* K-Means dari Sklearn. Gambar 6, merupakan potongan kode program untuk melakukan *clustering* dengan metode K-Means.

```
# Pemodelan KMeans dengan jumlah cluster yang optimum
model = KMeans(n_clusters = 3, init = "k-means++",
max_iter = 300, n_init = 10, random_state = 0)
y_clusters = model.fit_predict(x)
```

Gambar 6. Kode Program Clustering dengan K-Means

Adapun hasil pemodelan dengan menggunakan 3 (tiga) *cluster* sesuai dengan kode program pada Gambar 6, hasilnya direkap pada Tabel 4:

Tabel 4. Hasil *Clustering*

Cluster	Titik Centroids			Jumlah Data
	Lama Studi	Umur	IPK	
1	5,2	24,5	3,31	56
2	4,2	23,8	3,53	138
3	6,3	24,7	3,27	45

F. Evaluasi Hasil Clustering

Hasil *clustering* menggunakan algoritma *K-Means*, dapat diukur atau dinilai berdasarkan jumlah cluster yang dibuat. Sebuah indeks yang mengukur tingkat perbedaan antar cluster dikenal sebagai *Davies-Bouldin Index*, DBI ini digunakan untuk melakukan evaluasi, dimana nilai indeks yang lebih rendah, maka hasil *clustering*nya dianggap lebih baik. Hasil Eksprimen pada *clustering* prestasi akademik lulusan ditunjukkan pada Tabel 5:

Tabel 5. Hasil Pengujian

Cluster	DBI
2	0,92
3	0,79
4	0,98

Berdasarkan Tabel 5, dapat dilihat bahwa nilai index terkecil yaitu 0,79 pada jumlah *cluster* 3 (tiga), sehingga pada *clustering* prestasi akademik lulusan pada Universitas Ichsan Gorontalo untuk Tahun Akademik 2023/2024 Ganjil dipilih 3 (tiga) *cluster* sebagai jumlah *cluster* yang paling ideal.

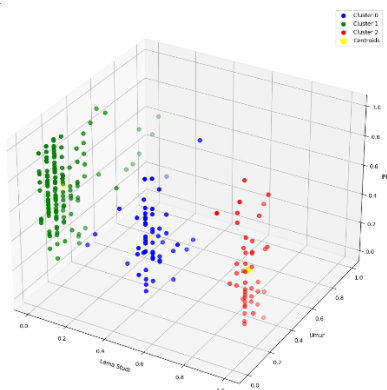
G. Visualisasi dan Analisis Hasil Clustering

Penggunaan algoritma K-Means, hasil *clustering* dapat diukur atau dinilai berdasarkan jumlah cluster yang dibuat. Sebuah indeks yang mengukur tingkat perbedaan antar cluster dikenal sebagai *Davies-Bouldin Index* digunakan untuk melakukan evaluasi. Dengan nilai indeks yang lebih rendah, hasil *clustering* dianggap lebih baik. Kode program visualisasi hasil *clustering* ditunjukkan pada Gambar 7:

```
# Visualisasi hasil cluster dengan 3d scatterplot menggunakan matplotlib
fig = plt.figure(figsize = (13,13))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(x[y_clusters == 0,0],x[y_clusters == 0,1],x[y_clusters == 0,2],
s = 50 , color = 'blue', label = "Cluster 0")
ax.scatter(x[y_clusters == 1,0],x[y_clusters == 1,1],x[y_clusters == 1,2],
s = 50 , color = 'red', label = "Cluster 1")
ax.scatter(x[y_clusters == 2,0],x[y_clusters == 2,1],x[y_clusters == 2,2],
s = 50 , color = 'green', label = "Cluster 2")
ax.scatter(model.cluster_centers[:,0],model.cluster_centers[:,1],
model.cluster_centers[:,2], s = 100, c = "yellow",
label = "Centroids")
ax.set_xlabel('Lama Studi')
ax.set_ylabel('Umur')
ax.set_zlabel('IPK')
ax.legend()
plt.show()
```

Gambar 7. Kode Program Visualisasi Hasil *Clustering*

Hasil visualisasi *clustering* prestasi akademik lulusan sesuai hasil kode program pada Gambar 7, hasilnya ditunjukkan pada Gambar 8:



Gambar 8. Visualisasi Hasil *Clustering*

Berdasarkan hasil visualisasi yang ditunjukkan pada Gambar 8, dengan jelas setiap kelompok/*cluster* antara *cluster* 1 (satu) dengan *cluster* lainnya dikelompokkan sesuai kemiripan karakteristik data lulusan yang diberikan warna berbeda untuk setiap *cluster*.

Berdasarkan Tabel 4 dan Gambar 8 dapat dilakukan analisis hasil *clustering* yaitu (a). *Cluster* 1 yang diberikan warna biru: merupakan kelompok lulusan yang memiliki prestasi akademik cukup baik, dimana lama studi rata-rata 5,2 tahun, umur 24,5 tahun dan IPK rata-rata 3,31, (b). *Cluster* 2 yang diberikan warna hijau: merupakan kelompok lulusan yang memiliki prestasi akademik sangat baik, dimana lama studi rata-rata 4,2 tahun, umur 23,8 tahun dan IPK 3,53 sehingga lulusan ini masih masuk kategori lulusan dengan predikat Pujian, (c). *Cluster* 3 yang diberikan warna merah: merupakan kelompok lulusan yang memiliki prestasi akademik kurang baik jika ditinjau dari sisi lama studi tetapi dari sisi umur dan IPK masih masuk kategori baik, dimana lama studi rata-rata 6,3 tahun, umur 24,7 tahun, dan IPK 3,27, sehingga *cluster* 3 (tiga) menjadi perhatian yang harus ditingkatkan prestasinya mengingat masa studi yang sudah melewati 6 tahun atau 12 semester. Hasil *clustering* secara detail ditunjukkan pada Tabel 6 :

Tabel 6. Sampel Hasil *Clustering*

No	NIM	Lama Studi	Umur	IPK	Cluster
1	E1117116	6,3	21	3,84	C3
2	E1117139	6,4	24	3,3	C3
3	E1117145	5	29	3,03	C1
4	E1118010	5,3	23	3,29	C1
5	E1118037	5,3	24	3,52	C1
6	E1118046	5,3	23	3,54	C1
7	E1118052	5,3	22	3,36	C1
8	E1119007	4,3	26	3,63	C2
9	E1119010	4,2	22	3,43	C2
10	E1119012	4,2	21	3,42	C2
11	E1119013	4,2	22	3,8	C2
12	E1119017	4,2	23	3,68	C2
13	E1119018	4,2	23	3,86	C2
14	E1119020	4,2	23	3,69	C2

No	NIM	Lama Studi	Umur	IPK	Cluster
15	E1119021	4,2	23	3,68	C2

IV. KESIMPULAN

Berdasarkan pembahasan yang sudah diuraikan di atas, maka dapat disimpulkan bahwa *clustering* prestasi akademik lulusan pada Universitas Ichsan Gorontalo untuk Tahun Akademik 2023/2024 semester Ganjil dengan menggunakan atribut terdiri dari lama studi, umur dan IPK didapatkan 3 (tiga) *cluster* yaitu *cluster* 1 (satu) merupakan kelompok lulusan yang memiliki prestasi akademik cukup baik sebanyak 56 lulusan, *cluster* 2 (dua) merupakan kelompok lulusan dengan prestasi akademik sangat baik sebanyak 138 lulusan dan *cluster* 3 (tiga) merupakan kelompok lulusan yang memiliki prestasi akademik kurang baik dari sisi lama studi sebanyak 45 lulusan. Jumlah *cluster* sebanyak 3 kelompok dijadikan sebagai jumlah *cluster* paling ideal sesuai hasil perhitungan teknik *Elbow* dan nilai evaluasi pengujian *Davies-Bouldin Index* yang paling kecil yaitu 0,79 hal ini menunjukkan bahwa pusat-pusat *cluster* berada jauh satu sama lain sehingga hasil *clustering* masuk kategori baik karena nilai DBI dibawah 1.

REFERENSI

- [1] Mendikbudristek, *Keputusan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia Nomor 210/M/2023 Tentang Indikator Kinerja Utama Perguruan Tinggi dan LL-DIKTI di Kemendikbudristek*. 2023.
- [2] E. Prasetyo, "Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab," *Penerbit Andi*, 2014.
- [3] R. Primartha, *Algoritma Machine Learning*. Bandung: Informatika, 2021.
- [4] F. N. R. F. J. Aziz, B. D. Setiawan, and I. Arwani, "Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 6, pp. 2243–2251, 2018.
- [5] A. Hardianti and D. Agushinta R, "Universitas Darma Persada Menggunakan Metode Clustering K-Means Pattern Analysis of the Student Study Period in the Faculty of Engineering At Darma Persada University Using K-Means," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 4, pp. 861–868, 2020, doi: 10.25126/jtiik.202071001.
- [6] S. Suraya, M. Sholeh, and D. Andayati, "Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa," *Skanika*, vol. 6, no. 1, pp. 51–60, 2023, doi: 10.36080/skanika.v6i1.2982.
- [7] A. Almohammed and M. Abido, "RES on Power Operation: K-Means Clustering Over Elbow Approach," *WSEAS Trans. POWER Syst.*, vol. 15, pp. 214–221, Dec. 2020, doi: 10.37394/232016.2020.15.25.
- [8] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- [9] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and

- B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, p. 012017, Apr. 2018, doi: 10.1088/1757-899X/336/1/012017.
- [10] R. Ishak and Amiruddin, "Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika Dengan Metode K-Means," *Jambura J. Electr. Electron. Eng.*, vol. 4, pp. 65–69, 2022, doi: 10.37905/jjee.v4i1.11997.
- [11] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 1, p. 56, 2023, doi: 10.30998/string.v8i1.16388.
- [12] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, p. 96, 2021, doi: 10.31294/evolusi.v9i1.10428.
- [13] A. K. Singh, S. Mittal, P. Malhotra, and Y. V. Srivastava, "Clustering Evaluation by Davies-Bouldin Index(DBI) in Cereal data using K-Means," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, Mar. 2020, pp. 306–310, doi: 10.1109/ICCMC48092.2020.ICCMC-00057.
- [14] A. Bengnga and R. Ishak, "Penerapan XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 2, pp. 192–196, 2023, doi: 10.37905/jjee.v5i2.20253.
- [15] "Clustering Performance Evaluation," *scikit-learn developers*. <https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation> (accessed May 20, 2023).
- [16] A. B. H. Kiat, Y. Azhar, and V. Rahmayanti, "Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency & Monetary)," *Repositor*, vol. 2, no. 7, pp. 945–952, 2020.
- [17] "Clustering," *scikit-learn developers*. <https://scikit-learn.org/stable/modules/clustering.html#> (accessed May 20, 2023).
- [18] BAAK, "Data Lulusan," Gorontalo, 2023. [Online]. Available: <https://siakun.unisan.ac.id>.