Analisis Segmentasi Calon Mahasiswa Baru Berdasarkan Faktor Ekonomi Menggunakan Algoritma K-Medoids

Khamarudin Syarif[1)*], Ummu Wachidatul Latifah[2)], Yuli Wijayanti[3)], Siska Febriani[4)]
[1,2,3)] Sains Data, [4)] Informatika, Institut Teknologi Bisnis dan Kesehatan Bhakti Putra Bangsa Indonesia

Received: 10 March 2026 Accepted: 23 April 2026 Published: 1 June 2026

*komarudin.syarif@gmail.com

Kata Kunci: K-Medoids, Clustering, Data Mining, Faktor Ekonomi

Abstrak : Pengelompokan calon mahasiswa baru berdasarkan kondisi ekonomi merupakan langkah penting dalam mendukung pengambilan keputusan institusi pendidikan, khususnya dalam penentuan program bantuan atau beasiswa. Penelitian ini bertujuan untuk menerapkan algoritma K-Medoids dalam melakukan analisis clustering terhadap calon mahasiswa baru di IBISA berdasarkan faktor ekonomi. Data yang digunakan terdiri dari 20 data calon mahasiswa dengan indikator ekonomi meliputi pekerjaan orang tua, pendapatan per bulan, jumlah tanggungan keluarga, kepemilikan aset, dan pendidikan orang tua. Tahapan penelitian meliputi proses preprocessing data yang terdiri dari data cleaning, transformasi data, normalisasi, serta perhitungan jarak menggunakan metode Euclidean Distance. Selanjutnya dilakukan proses clustering menggunakan algoritma K-Medoids untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan karakteristik ekonomi. Hasil penelitian menunjukkan bahwa data calon mahasiswa dapat dikelompokkan menjadi tiga cluster utama yaitu kelompok ekonomi rendah, ekonomi menengah, dan ekonomi tinggi. Metode K-Medoids mampu menghasilkan pengelompokan yang cukup stabil karena menggunakan medoid sebagai pusat cluster yang lebih tahan terhadap outlier. Hasil clustering ini dapat dimanfaatkan oleh pihak perguruan tinggi sebagai dasar dalam pengambilan kebijakan terkait program bantuan pendidikan, penentuan beasiswa, serta perencanaan strategi akademik yang lebih tepat sasaran.

PENDAHULUAN

Pendidikan tinggi memiliki peran penting dalam meningkatkan kualitas sumber daya manusia dan daya saing bangsa. Perguruan tinggi tidak hanya dituntut menghasilkan lulusan yang memiliki kompetensi akademik, tetapi juga kemampuan analisis, pemecahan masalah, serta adaptasi terhadap perkembangan teknologi di era transformasi digital [1]. Oleh karena itu, pengelolaan data mahasiswa menjadi aspek penting dalam mendukung pengambilan keputusan strategis di lingkungan pendidikan tinggi [2]. Salah satu faktor yang memengaruhi keberhasilan studi mahasiswa adalah kondisi sosial ekonomi. Latar belakang ekonomi keluarga dapat memengaruhi kemampuan mahasiswa dalam memenuhi kebutuhan pendidikan, termasuk pembayaran biaya kuliah, akses terhadap sumber belajar, serta partisipasi dalam kegiatan akademik. Penelitian menunjukkan bahwa analisis data mahasiswa dapat memberikan informasi penting bagi perguruan tinggi dalam merancang kebijakan beasiswa, bantuan pendidikan, dan strategi penerimaan mahasiswa baru yang lebih tepat sasaran [3].

Seiring berkembangnya teknologi informasi, pendekatan data mining banyak digunakan untuk mengekstraksi pola dan pengetahuan dari data pendidikan dalam jumlah besar. Data mining dalam bidang pendidikan atau *Educational Data Mining* (EDM) memungkinkan institusi pendidikan memahami karakteristik mahasiswa serta mendukung pengambilan keputusan berbasis data [4]. Teknik analisis yang

sering digunakan adalah clustering, yaitu metode pengelompokan data berdasarkan tingkat kemiripan karakteristik tertentu sehingga objek dalam satu kelompok memiliki kesamaan yang tinggi dibandingkan dengan kelompok lainnya [5]. Teknik clustering telah banyak diterapkan dalam bidang pendidikan untuk mengidentifikasi pola perilaku belajar mahasiswa, menganalisis performa akademik, serta memetakan karakteristik mahasiswa berdasarkan berbagai atribut seperti nilai akademik, latar belakang sosial ekonomi, dan aktivitas pembelajaran [6] [7].

Salah satu algoritma clustering yaitu, algoritma K-Medoids, yaitu metode clustering berbasis partisi yang menggunakan objek data aktual sebagai pusat cluster (medoid). Algoritma ini memiliki keunggulan dalam menangani data yang mengandung outlier sehingga menghasilkan cluster yang lebih stabil dan representatif dibandingkan beberapa metode clustering lainnya . Berbeda dengan algoritma K-Means yang menggunakan nilai rata-rata sebagai pusat cluster, K-Medoids memiliki keunggulan dalam menangani data yang mengandung outlier atau nilai ekstrem sehingga menghasilkan cluster yang lebih stabil dan representatif. Selain itu, penggunaan medoid sebagai pusat cluster memudahkan interpretasi hasil analisis karena medoid merepresentasikan objek data nyata yang berada di dalam dataset [8]. Melalui penerapan metode clustering terhadap data calon mahasiswa berdasarkan faktor ekonomi, diharapkan dapat diperoleh segmentasi ekonomi calon mahasiswa yang dapat dimanfaatkan oleh pihak perguruan tinggi dalam mendukung kebijakan beasiswa dan perencanaan penerimaan mahasiswa baru secara lebih efektif dan berbasis data.

TINJAUAN LITERATUR

Algoritma K-Medoids

Algoritma K-Medoids merupakan salah satu metode clustering berbasis partisi yang digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan tingkat kemiripan karakteristik antar data. Berbeda dengan algoritma K-Means yang menggunakan nilai rata-rata sebagai pusat cluster (centroid), K-Medoids menggunakan objek data aktual sebagai pusat cluster (medoid). Medoid merupakan data yang memiliki total jarak paling kecil terhadap seluruh data lain dalam satu cluster. Penggunaan objek data nyata sebagai pusat cluster membuat algoritma ini lebih robust terhadap outlier atau nilai ekstrem dibandingkan K-Means. Metode K-Medoids sering digunakan dalam berbagai bidang seperti analisis pendidikan, kesehatan, bisnis, dan analisis data sosial karena kemampuannya dalam menemukan pola pengelompokan dari dataset yang tidak memiliki label kelas (unsupervised learning) [9]. Secara umum, algoritma K-Medoids menggunakan data asli sebagai pusat cluster sehingga lebih stabil terhadap data yang mengandung outlier atau noise [10], kemudian setiap data akan ditempatkan pada cluster yang memiliki jarak paling dekat dengan medoid tersebut. Selanjutnya dilakukan proses pertukaran antara medoid dengan data non-medoid untuk meminimalkan total jarak dalam cluster hingga tidak terdapat perubahan yang signifikan atau mencapai kondisi konvergen. Langkah-langkah algoritma K-Medoids secara umum adalah sebagai berikut:[11]

1. Menentukan jumlah cluster k.
2. Memilih secara acak k objek data sebagai medoid awal.
3. Menghitung jarak antara setiap data dengan masing-masing medoid.
4. Mengelompokkan data ke dalam cluster dengan jarak terdekat terhadap medoid.
5. Mengganti medoid dengan objek lain dalam cluster jika dapat menurunkan total jarak.
6. Mengulangi proses hingga tidak ada perubahan medoid atau nilai biaya (cost) minimum tercapai.

Fungsi objektif dari algoritma ini dapat dinyatakan sebagai berikut:

$$J = \sum_{i=1}^{k} \sum_{x_j \in C_i} d(x_j, m_i)$$

Keterangan:
J= total cost atau fungsi objektif
k = jumlah cluster

C_i = himpunan data pada cluster ke-i
x_j = objek data ke-j
m_i = medoid pada cluster ke-i
$d(x_j, m\_i)$ = jarak antara data x_j dengan medoid m_i
Perhitungan jarak yang umum digunakan adalah *Euclidean Distance*, yang dirumuskan sebagai:

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

Keterangan:
x = data pertama
y = data kedua
n = jumlah atribut
x_i, y_i = nilai atribut ke$-i$
Nilai jarak ini digunakan untuk menentukan kedekatan antara data dengan medoid sehingga data dapat ditempatkan pada cluster yang paling sesuai [12]

Evaluasi Cluster

Metode evaluasi clustering yang umum digunakan adalah *Silhouette Coefficient* dan *Davies Bouldin Index* (DBI).

1. Silhoutte Coefficient
   Silhouette Coefficient merupakan metode evaluasi clustering yang digunakan untuk mengukur kualitas pengelompokan berdasarkan jarak antara objek dalam cluster yang sama dan jarak terhadap cluster terdekat lainnya. Metode ini diperkenalkan oleh Rousseeuw dan hingga saat ini masih menjadi salah satu teknik evaluasi cluster yang paling populer dalam analisis data [13]
   Nilai silhouette untuk setiap data berada pada rentang −1 hingga 1.
   Interpretasi nilai Silhouette:
   a. Mendekati 1 objek sangat sesuai dengan clusternya
   b. Mendekati 0 objek berada di batas dua cluster
   c. Mendekati −1 objek kemungkinan berada pada cluster yang salah
   Silhouette untuk suatu data $i$dihitung menggunakan persamaan berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

   Keterangan:
   $s(i)$= nilai silhouette untuk data ke-i
   $a(i)$= rata-rata jarak antara data ke-i dengan semua data dalam cluster yang sama
   $b(i)$= rata-rata jarak minimum antara data ke-i dengan semua data pada cluster lain (cluster terdekat)
   Rata-rata nilai silhouette seluruh data digunakan untuk menilai kualitas keseluruhan cluster:

$$S = \frac{1}{n} \sum_{i=1}^{n} s(i)$$

   Keterangan:
   S= nilai rata-rata silhouette
   n= jumlah data
   Nilai S yang lebih besar menunjukkan kualitas clustering yang lebih baik, karena menunjukkan bahwa objek dalam cluster lebih dekat satu sama lain dibandingkan dengan objek pada cluster lain [14].
2. **Davies–Bouldin Index (DBI)**
   Davies–Bouldin Index (DBI) merupakan metode evaluasi clustering yang mengukur rata-rata kesamaan antara setiap cluster dengan cluster yang paling mirip dengannya. Metode ini mempertimbangkan dua komponen utama yaitu kompaksi cluster (cluster compactness) dan separasi antar cluster (cluster separation)[15]. Langkah pertama adalah menghitung scatter cluster:

$$S_i = \frac{1}{|C_i|} \sum_{x \in C_i} d(x, c_i)$$

Keterangan:
S_i = rata-rata jarak data pada cluster ke-i terhadap centroid/medoid cluster
C_i = cluster ke-i
c_i = pusat cluster
$d(x, c_i)$ = jarak antara data dengan pusat cluster
Selanjutnya dihitung rasio kemiripan antar cluster:

$$R_{ij} = \frac{S_i + S_j}{d(c_i, c_j)}$$

Keterangan:
R_{ij} = rasio kemiripan antara cluster i dan j
$d(c_i, c_j)$ = jarak antara pusat cluster i dan j
Kemudian nilai DBI dihitung dengan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} (R_{ij})$$

Keterangan:
k = jumlah cluster
R_{ij} = rasio kemiripan antar cluster
Nilai DBI yang kecil menunjukkan cluster yang lebih baik, karena menandakan cluster memiliki jarak yang jauh satu sama lain serta memiliki tingkat kepadatan yang baik [12]. Davies–Bouldin Index sering digunakan dalam penelitian clustering karena mampu mengevaluasi keseimbangan antara kedekatan data dalam cluster dan pemisahan antar cluster secara komprehensif [16].

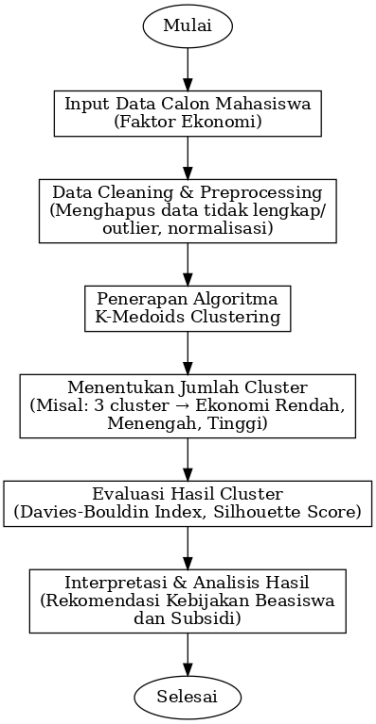
METODE PENELITIAN

Penelitian menggunakan pendekatan data mining dengan teknik clustering untuk mengelompokkan calon mahasiswa baru Institut Teknologi Bisnis dan Kesehatan Bhakti Putra Bangsa Indonesia (IBISA) berdasarkan faktor ekonomi. Tahapan penelitian dimulai dengan pengumpulan data calon mahasiswa baru yang diperoleh dari database penerimaan mahasiswa IBISA. Data yang digunakan meliputi beberapa indikator ekonomi, seperti pendapatan orang tua, pekerjaan orang tua, jumlah tanggungan keluarga, dan kondisi tempat tinggal. Selanjutnya dilakukan tahap preprocessing data yang meliputi pembersihan data (data cleaning), transformasi data, serta normalisasi untuk memastikan kualitas data sebelum proses analisis dilakukan [2]. Pengolahan data menggunakan software Python. Tahap data cleaning yang dilakukan mencakup tiga langkah utama: (1) pemeriksaan dan penanganan nilai kosong (missing values) pada dataset ini tidak ditemukan nilai kosong sehingga tidak diperlukan imputasi data; (2) identifikasi dan penanganan duplikasi data setelah dilakukan pengecekan, tidak terdapat entri data yang duplikat; serta (3) pemeriksaan konsistensi format data, khususnya pada kolom pendapatan yang distandarkan ke dalam satuan rupiah numerik dan kolom kepemilikan aset yang dikonversi ke dalam skala ordinal numerik. Setelah data cleaning, dilakukan transformasi data kategorikal (pekerjaan orang tua, kepemilikan aset, dan pendidikan orang tua) menjadi nilai numerik menggunakan label encoding, kemudian dinormalisasi dengan Min-Max Scaling agar seluruh fitur berada pada rentang nilai yang seragam (0–1) sebelum proses clustering dijalankan.

Setelah tahap preprocessing, dilakukan proses clustering menggunakan algoritma K-Medoids. Algoritma ini dipilih karena menggunakan objek data aktual sebagai pusat cluster (medoid) sehingga lebih stabil terhadap data yang mengandung outlier dibandingkan metode berbasis centroid seperti K-Means [3]. Proses clustering dimulai dengan menentukan jumlah cluster k, memilih medoid awal secara acak, menghitung jarak antara setiap data dengan medoid menggunakan *Euclidean distance*, kemudian

mengelompokkan data ke dalam cluster dengan jarak terdekat terhadap medoid. Proses iterasi dilakukan hingga tidak terjadi perubahan medoid atau nilai cost minimum tercapai [4].

Selanjutnya dilakukan evaluasi kualitas cluster menggunakan dua metode sekaligus, yaitu Silhouette Coefficient (SC) dan Davies-Bouldin Index (DBI). Untuk menentukan jumlah cluster optimal, dilakukan eksperimen dengan variasi nilai k dari k=2 hingga k=5 (sebanyak 4 kali percobaan). Pada k=2, nilai SC sebesar 0,42 dan DBI sebesar 1,21 menunjukkan pemisahan cluster yang kurang optimal. Pada k=3, diperoleh nilai SC tertinggi sebesar 0,61 dan nilai DBI terendah sebesar 0,74, yang mengindikasikan kualitas cluster terbaik. Pada k=4, nilai SC turun menjadi 0,38 dan DBI naik menjadi 1,05, sedangkan pada k=5, performa evaluasi semakin menurun. Berdasarkan hasil evaluasi tersebut, jumlah cluster k=3 dipilih sebagai konfigurasi optimal karena menghasilkan nilai SC tertinggi (mendekati 1) dan nilai DBI terendah (mendekati 0) secara bersamaan. Nilai SC yang mendekati 1 menunjukkan bahwa data dalam satu cluster lebih dekat satu sama lain dibandingkan dengan data di cluster lain, sedangkan nilai DBI yang kecil menunjukkan cluster yang kompak dan terpisah dengan baik [5]. Hasil clustering kemudian dianalisis untuk mengidentifikasi pola distribusi ekonomi calon mahasiswa yang dapat dimanfaatkan oleh pihak institusi dalam mendukung kebijakan penerimaan mahasiswa dan program bantuan pendidikan secara lebih tepat sasaran.
Berikut ini gambar diagram alir metode penelitian:



Gambar 1. Diagram Alir Metode Penelitian

HASIL PENELITIAN

Penelitian menggunakan data sebanyak 20 calon mahasiswa baru IBISA yang dianalisis berdasarkan beberapa indikator faktor ekonomi, yaitu pekerjaan orang tua, pendapatan per bulan, jumlah tanggungan keluarga, kepemilikan aset, dan tingkat pendidikan orang tua. Variabel-variabel tersebut digunakan untuk menggambarkan kondisi ekonomi keluarga calon mahasiswa secara lebih komprehensif. Berikut ini tabel yang diperoleh berdasarkan indikator faktor ekonomi:

Tabel 1. Data Mahasiswa Baru IBISA

No	Nama	Pekerjaan Orang Tua	Pendapatan/Bulan	Jumlah Tanggungan	Kepemilikan Aset	Pendidikan Orang Tua
1	AN	Buruh Harian	Rp1.500.000	5	Rumah kontrak	SMP
2	BR	Petani	Rp1.800.000	6	Rumah sederhana	SMA
3	CN	Pegawai Swasta	Rp4.000.000	4	Rumah, 1 motor	SMA

4	DN	Wiraswasta Kecil	Rp4.500.000	3	Rumah, 1 motor	SMA
5	ER	PNS	Rp9.000.000	2	Rumah, 2 motor, tanah	S1
6	FN	Buruh Bangunan	Rp1.700.000	5	Tidak ada, numpang keluarga	SMP
7	GN	Pedagang Kaki Lima	Rp3.200.000	4	Rumah, 1 motor	SMA
8	HN	Guru Honorer	Rp2.800.000	6	Rumah, 1 motor	D3
9	IN	Karyawan	Rp5.500.000	3	Rumah, 1 mobil	S1
10	JN	Petani	Rp2.000.000	5	Rumah sederhana	SMP
11	KN	PNS	Rp1.600.000	2	Rumah, 1 mobil, tanah	S1
12	LN	Wiraswasta (Toko)	Rp2.200.000	6	Rumah, 2 motor	S1
13	MN	Manajer	Rp7.000.000	2	Rumah kontrak, 1 motor	S1
14	UN	Buruh Pabrik	Rp2.600.000	5	Rumah sederhana	SMA
15	ON	Buruh Pabrik	Rp1.900.000	5	Rumah, 2 mobil, tanah	SD
16	PN	Pegawai Swasta	Rp3.600.000	4	Rumah, 1 motor	S1
17	QN	Wiraswasta	Rp2.400.000	6	Rumah, mobil, tanah	S1
18	RN	Dokter	Rp6.500.000	3	Rumah sederhana	S1
19	SN	Pedagang Kelontong	Rp8.200.000	2	Rumah, 1 motor	SMA
20	TN	PNS	Rp4.800.000	3	Rumah, mobil, tanah	S2

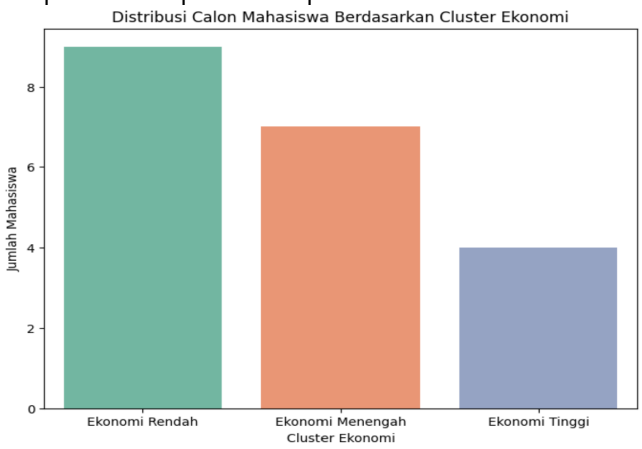
Dari tabel 1 terlihat adanya variasi yang cukup jelas pada tingkat pendapatan dan kondisi aset keluarga calon mahasiswa. Fokus penelitian ini pada faktor ekonomi, yang direpresentasikan oleh dua variabel utama Pendapatan Orang Tua (Rp/bulan) dan Jumlah Tanggungan Keluarga (orang).

Sebelum proses clustering dilakukan, data terlebih dahulu melalui tahap preprocessing yang meliputi pembersihan data dan transformasi data ke dalam bentuk numerik agar dapat diproses oleh algoritma clustering. Selanjutnya dilakukan proses pengelompokan menggunakan algoritma K-Medoids dengan jumlah cluster sebanyak 3 cluster, yang merepresentasikan kelompok ekonomi rendah, menengah, dan tinggi. Berdasarkan penerapan algoritma K-Medoids, diperoleh hasil clustering sebagai berikut:

Tabel 2. Hasil Clustering Data Mahasiswa

Cluster	Anggota Data	Karakteristik Umum
Cluster 1 (Ekonomi Rendah)	AN, BR, FN, JN, KN, LN, ON, QN, UN	Pendapatan < Rp3 juta, tanggungan ≥ 5, aset terbatas, pendidikan orang tua rendah (SD–SMA).
Cluster 2 (Ekonomi Menengah)	CN, DN, GN, HN, IN, PN, TN	Pendapatan Rp3–7 juta, aset sederhana (rumah + motor), pendidikan orang tua SMA–S1.
Cluster 3 (Ekonomi Tinggi)	ER, MN, RN, SN	Pendapatan umumnya > Rp6 juta (rata-rata Rp7.675.000), tanggungan rendah (rata-rata 2,2 orang), aset relatif lengkap, DAN/ATAU pendidikan orang tua tinggi (S1–S2). Karakteristik ini bersifat kombinasi (AND/OR): anggota cluster tidak harus memenuhi seluruh kriteria secara bersamaan, melainkan memiliki kedekatan jarak Euclidean terhadap medoid cluster dibandingkan cluster lainnya.

Distribusi mahasiswa pada tiap cluster dapat dilihat pada Gambar berikut:



Gambar 2. Grafik Distribusi Calon Mahasiswa

Berdasarkan Gambar 2, diperoleh pembagian cluster sebagai berikut:

a. Cluster 1 (Ekonomi Rendah): 9 mahasiswa
b. Cluster 2 (Ekonomi Menengah): 7 mahasiswa
c. Cluster 3 (Ekonomi Tinggi): 4 mahasiswa

Tabel 3. Hasil Clustering Mahasiswa Baru IBISA

Nama	Pendapatan (Rp)	Tanggungan	Cluster Ekonomi
AN	1.500.000	5	Rendah
BR	1.800.000	6	Rendah
CN	4.000.000	4	Menengah
DN	4.500.000	3	Menengah
ER	9.000.000	2	Tinggi
FN	1.700.000	5	Rendah
GN	3.200.000	4	Menengah
HN	2.800.000	6	Menengah
IN	5.500.000	3	Menengah
JN	2.000.000	5	Rendah
KN	1.600.000	4	Rendah
LN	2.200.000	6	Rendah
MN	7.000.000	2	Tinggi
ON	1.900.000	5	Rendah
PN	3.600.000	4	Menengah
QN	2.400.000	6	Rendah
RN	6.500.000	3	Tinggi
SN	8.200.000	2	Tinggi
TN	4.800.000	3	Menengah
UN	2.600.000	5	Rendah

Berikut jumlah anggota per kategori ekonomi:

Tabel 4. Jumlah Anggota per Kategori Ekonomi

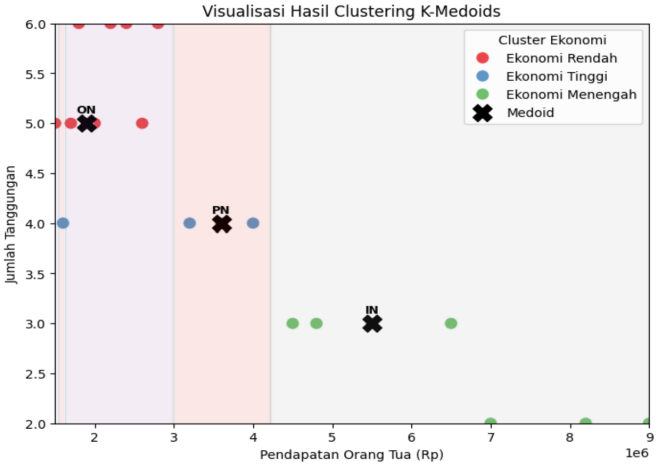
Cluster	Jumlah
Ekonomi Rendah	9 orang
Ekonomi Menengah	7 orang
Ekonomi Tinggi	4 orang

Komposisi menunjukkan bahwa calon mahasiswa dari ekonomi rendah merupakan kelompok terbesar, sehingga IBISA perlu memberi perhatian lebih terhadap potensi kebutuhan finansial mereka. Rata-rata pendapatan dan tanggungan dihitung untuk memahami karakteristik masing-masing cluster.

Tabel 5. Analisis Hasil Cluster

Cluster Ekonomi	Pendapatan Rata-rata	Tanggungan Rata-rata
Rendah	± Rp 1.887.000	5,1 orang
Menengah	± Rp 4.157.000	3,7 orang
Tinggi	± Rp 7.675.000	2,2 orang

Scatter plot yang dihasilkan dari kode Python sebelumnya menampilkan tiga kelompok (cluster) hasil algoritma K-Medoids, berdasarkan dua variabel:



Gambar 3. Visualisasi hasil Clustering K-Medoids

PEMBAHASAN

Penelitian dari permasalahan yang dihadapi oleh institusi pendidikan tinggi, khususnya IBISA, dalam mengelola keberagaman kondisi ekonomi calon mahasiswa baru secara sistematis dan berbasis data. Tanpa adanya pemetaan yang jelas terhadap latar belakang ekonomi mahasiswa, pengambilan keputusan terkait program beasiswa, bantuan biaya pendidikan, dan strategi penerimaan mahasiswa cenderung bersifat subjektif dan kurang tepat sasaran. Oleh karena itu, diperlukan pendekatan data mining berbasis clustering yang dapat menghasilkan segmentasi ekonomi secara objektif dan terukur.

Temuan utama menunjukkan bahwa algoritma K-Medoids berhasil mengelompokkan 20 calon mahasiswa baru IBISA ke dalam tiga cluster ekonomi yang bermakna: Cluster 1 (Ekonomi Rendah, n=9), Cluster 2 (Ekonomi Menengah, n=7), dan Cluster 3 (Ekonomi Tinggi, n=4). Cluster terbesar adalah kelompok ekonomi rendah (45% dari total mahasiswa) dengan rata-rata pendapatan orang tua sekitar Rp1.887.000 per bulan dan rata-rata tanggungan 5,1 orang. Kelompok ini menunjukkan bahwa hampir separuh calon mahasiswa IBISA berasal dari keluarga dengan kondisi ekonomi yang rentan secara finansial. Evaluasi kualitas cluster menggunakan Silhouette Coefficient (SC=0,61) dan Davies-Bouldin Index (DBI=0,74) pada k=3 mengkonfirmasi bahwa tiga cluster merupakan konfigurasi optimal, menghasilkan pemisahan antar kelompok yang cukup baik sekaligus kohesi yang kuat di dalam masing-masing kelompok.

Hasil sejalan dengan penelitian sebelumnya yang menerapkan teknik clustering pada data mahasiswa untuk mendukung pengambilan keputusan institusional. Diana et al. [5] menunjukkan bahwa segmentasi calon mahasiswa berbasis data mining dapat menghasilkan pemetaan yang lebih akurat dibandingkan pendekatan konvensional. Qomariyah dan Siregar [10] juga menemukan bahwa K-Medoids menghasilkan kualitas cluster yang baik pada data mahasiswa dengan nilai DBI yang rendah, sejalan dengan temuan penelitian ini. Penggunaan medoid sebagai pusat cluster terbukti memberikan keunggulan dalam menangani variasi data ekonomi yang beragam, karena medoid merupakan objek data nyata sehingga lebih representatif dan robust terhadap outlier dibandingkan centroid pada K-Means. Pola yang ditemukan dimana kelompok ekonomi rendah mendominasi juga konsisten dengan karakteristik demografis mahasiswa di institusi pendidikan tinggi swasta di daerah yang memiliki basis masyarakat dengan pendapatan menengah ke bawah.

Dari sisi implikasi praktis, hasil segmentasi ini memiliki nilai strategis yang signifikan bagi IBISA. Pertama, hasil clustering dapat dijadikan dasar objektif dalam merancang program beasiswa dan bantuan UKT (Uang Kuliah Tunggal) yang lebih tepat sasaran, khususnya bagi 45% mahasiswa yang masuk dalam kelompok ekonomi rendah. Kedua, profil tiap cluster dapat membantu bagian kemahasiswaan dalam menyusun program pendampingan finansial, seperti program kerja paruh waktu atau skema cicilan biaya kuliah yang disesuaikan dengan kemampuan ekonomi masing-masing kelompok. Ketiga, dari perspektif kebijakan penerimaan mahasiswa, pimpinan institusi dapat menggunakan informasi distribusi ekonomi ini untuk menentukan kuota penerimaan melalui jalur beasiswa versus jalur reguler secara lebih berimbang. Keempat, dalam konteks pendidikan yang lebih luas, temuan ini mendukung prinsip pemerataan akses pendidikan tinggi sebagaimana diamanatkan dalam kebijakan nasional, dengan memberikan landasan empiris bagi institusi dalam memformulasikan strategi inklusi finansial.

Beberapa keterbatasan yang perlu diakui. Pertama, ukuran sampel yang digunakan hanya sebanyak 20 data calon mahasiswa, sehingga generalisasi temuan ke populasi yang lebih besar perlu dilakukan dengan hati-hati. Jumlah data yang kecil juga berpotensi mempengaruhi stabilitas cluster dan keandalan nilai evaluasi SC maupun DBI. Kedua, variabel yang digunakan terbatas pada lima indikator ekonomi dari formulir pendaftaran yang tersedia, sehingga dimensi ekonomi lainnya seperti pengeluaran keluarga, hutang, atau akses terhadap infrastruktur tidak tercakup. Ketiga, algoritma K-Medoids yang digunakan dalam penelitian ini memerlukan penentuan jumlah cluster k secara a priori, yang membutuhkan eksperimen berulang dan pertimbangan kontekstual. Keempat, data yang digunakan bersifat cross-sectional (satu periode penerimaan), sehingga perubahan kondisi ekonomi mahasiswa dari waktu ke waktu tidak dapat ditangkap.

Berdasarkan keterbatasan tersebut, beberapa arah penelitian di masa mendatang dapat disarankan. Pertama, replikasi penelitian dengan dataset yang lebih besar dan mencakup beberapa angkatan penerimaan mahasiswa akan meningkatkan validitas dan reliabilitas temuan. Kedua, pengembangan model clustering hybrid yang menggabungkan K-Medoids dengan metode lain seperti DBSCAN atau hierarchical clustering dapat dieksplorasi untuk membandingkan kualitas pengelompokan pada data dengan karakteristik yang lebih beragam. Ketiga, integrasi variabel non-ekonomi seperti indeks prestasi, jarak tempuh, atau akses terhadap teknologi ke dalam model clustering dapat memberikan gambaran profil mahasiswa yang lebih holistik. Keempat, pengembangan sistem rekomendasi otomatis berbasis hasil clustering untuk membantu pengambilan keputusan beasiswa secara real-time merupakan topik penelitian lanjutan yang berpotensi memberikan dampak praktis yang lebih besar bagi institusi pendidikan.

KESIMPULAN

Berdasarkan hasil penelitian mengenai penerapan algoritma K-Medoids dalam pengelompokan calon mahasiswa baru berdasarkan faktor ekonomi, diperoleh beberapa kesimpulan. Pertama, data penelitian menggunakan variabel ekonomi yang terdiri dari pekerjaan orang tua, pendapatan per bulan, jumlah tanggungan keluarga, kepemilikan aset, dan pendidikan orang tua. Data tersebut melalui tahapan preprocessing yang meliputi data cleaning, transformasi data, normalisasi, serta perhitungan jarak menggunakan metode Euclidean Distance.

Kedua, penerapan algoritma K-Medoids mampu mengelompokkan data calon mahasiswa baru menjadi tiga cluster berdasarkan kemiripan karakteristik ekonomi. Cluster pertama merepresentasikan kelompok mahasiswa dengan kondisi ekonomi rendah yang ditandai oleh pendapatan orang tua rendah dan jumlah tanggungan yang relatif tinggi. Cluster kedua menunjukkan kelompok mahasiswa dengan kondisi ekonomi menengah dengan pendapatan yang relatif stabil dan tanggungan keluarga moderat. Sementara itu, cluster ketiga menggambarkan mahasiswa dengan kondisi ekonomi tinggi yang ditandai dengan pendapatan orang tua lebih besar serta jumlah tanggungan yang relatif sedikit.

Ketiga, hasil pengelompokan menunjukkan bahwa algoritma K-Medoids mampu menghasilkan cluster yang cukup jelas dan stabil. Penggunaan medoid sebagai pusat cluster membuat metode ini lebih tahan terhadap pengaruh outlier dibandingkan metode berbasis centroid seperti K-Means, sehingga hasil pengelompokan menjadi lebih representatif terhadap kondisi data

REFERENCES

[1] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, p. 11, Dec. 2022, doi: 10.1186/s40561-022-00192-z.

[2] W. Xiao, P. Ji, and J. Hu, "A survey on educational data mining methods used for predicting students' performance," *Engineering Reports*, vol. 4, no. 5, May 2022, doi: 10.1002/eng2.12482.

[3] Y. Lu, S. Yeom, J. Maktoubian, M. M. Rahman, and S.-H. Kim, "Improve Student Risk Prediction with Clustering Techniques: A Systematic Review in Education Data Mining," *Educ. Sci. (Basel).*, vol. 15, no. 12, p. 1695, Dec. 2025, doi: 10.3390/educsci15121695.

[4] D. J. Lemay, C. Baek, and T. Doleck, "Comparison of learning analytics and educational data mining: A topic modeling approach," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100016, 2021, doi: 10.1016/j.caeai.2021.100016.

[5] A. Diana, A. Ariesta, A. Wibowo, and D. A. B. Risaychi, "NEW STUDENT CLUSTERIZATION BASED ON NEW STUDENT ADMISSION USING DATA MINING METHOD," *Jurnal Pilar Nusa Mandiri*, vol. 19, no. 1, pp. 1–10, Mar. 2023, doi: 10.33480/pilar.v19i1.4089.

[6] I. Imron Rosyadi, F. Nurhadits, and C. Juliane, "Application of Data Mining to Measure the Effectiveness of the Islamic Boarding School's Independent Curriculum based on Learning Achievement using the Clustering Method," *Journal of World Science*, vol. 3, no. 5, pp. 514–519, May 2024, doi: 10.58344/jws.v3i5.595.

[7] C. Agusniar, S. Mulya Ulfa, and H. Rhomadhona, "Utilizing K-Means Clustering for Grouping Student Achievement Data to Evaluate Learning Activeness," *Journal of Advanced Computer Knowledge and Algorithms*, vol. 2, no. 3, pp. 56–59, Jul. 2025, doi: 10.29103/jacka.v2i3.22591.

[8] J. Heidari, N. Daneshpour, and A. Zangeneh, "A novel K-means and K-medoids algorithms for clustering non-spherical-shape clusters non-sensitive to outliers," *Pattern Recognit.*, vol. 155, p. 110639, Nov. 2024, doi: 10.1016/j.patcog.2024.110639.

[9] M. D. Ananda, K. N. Malik, A. F. N. Masruriyah, and M. Mardiah, "Studi Komparatif Algoritma K-Means dan K-Medoids untuk Segmentasi Informasi Kesehatan," *Computer Science (CO-SCIENCE)*, vol. 5, no. 2, pp. 103–112, Jul. 2025, doi: 10.31294/coscience.v5i2.9207.

[10] Qomariyah and M. U. Siregar, "Comparative Study of K-Means Clustering Algorithm and K-Medoids Clustering in Student Data Clustering," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 7, no. 2, pp. 91–99, May 2022, doi: 10.14421/jiska.2022.7.2.91-99.

[11] G. Guojun, M. Chaoqun, and W. Jianhong, "Center-Based Clustering Algorithms," in *Data Clustering: Theory, Algorithms, and Applications*, Society for Industrial and Applied Mathematics, 2007, pp. 161–182. doi: 10.1137/1.9780898718348.ch9.

[12] M. Bargavi, S. Krishnamoorthy, A. Chakravarty, K. Jadhav, and R. Gupta, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Predicting Student Performance in Higher Education Using A Cluster-Based Distributed Architecture (CDA) 1," 2023. [Online]. Available: www.ijisae.org

[13] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.

[14] M. H. Bin Roslan and C. J. Chen, "Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015-2021)," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 05, pp. 147–179, Mar. 2022, doi: 10.3991/ijet.v17i05.27685.

[15] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.

[16] A. F. Mohamed Nafuri, N. S. Sani, N. F. A. Zainudin, A. H. A. Rahman, and M. Aliff, "Clustering Analysis for Classifying Student Academic Performance in Higher Education," *Applied Sciences*, vol. 12, no. 19, p. 9467, Sep. 2022, doi: 10.3390/app12199467.