



Terbit *online* pada laman web jurnal :
<https://ejournal.sttp-yds.ac.id/index.php/js/index>

SAINSTEK

| ISSN (Print) 2337-6910 | ISSN (Online) 2460-1039 |



Optimasi Strategi Pemasaran Perguruan Tinggi Berbasis Segmentasi Data Menggunakan HDBSCAN dan GMM

Anwar Karnaidi^{a*}, Dhomas Hatta Fudholi^b

^{a,b}Magister Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia, Yogyakarta, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 22 April 2026

Revisi Akhir: 25 Juni 2026

Diterbitkan Online: 27 Juni 2026

KATA KUNCI

Segmentasi mahasiswa

HDBSCAN

GMM

STP

Pemasaran pendidikan tinggi

KORESPONDENSI

Telepon: -

E-mail: 23917011@students.uii.ac.id

ABSTRACT

Persaingan antar perguruan tinggi yang semakin ketat menuntut strategi pemasaran yang lebih presisi dan berbasis data. Namun, praktik segmentasi mahasiswa di banyak institusi masih didominasi oleh pendekatan konvensional yang bersifat subjektif dan kurang mampu menangkap heterogenitas karakteristik mahasiswa secara komprehensif. Penelitian ini bertujuan untuk mengoptimalkan strategi pemasaran perguruan tinggi melalui segmentasi mahasiswa berbasis data dengan pendekatan klusterisasi dua tahap (*two-stage clustering*), menggunakan data 731 mahasiswa baru Sekolah Tinggi Teknologi Pekanbaru (STTP) periode 2020–2023 dengan 97 atribut. Pada tahap awal, metode HDBSCAN digunakan untuk mengeksplorasi struktur global data mahasiswa, mengidentifikasi kluster mayoritas, serta menyaring *noise*. Selanjutnya, segmentasi lanjutan dilakukan menggunakan *Gaussian Mixture Model* (GMM) untuk memetakan heterogenitas internal pada kluster mayoritas secara probabilistik. Evaluasi jumlah kluster dilakukan menggunakan *Bayesian Information Criterion* (BIC) dan *Akaike Information Criterion* (AIC) guna memastikan keseimbangan antara kualitas model dan interpretabilitas. Hasil penelitian menunjukkan bahwa pendekatan *hybrid* ini mampu menghasilkan tiga segmen mahasiswa yang lebih homogen dan tajam dibandingkan penerapan GMM pada *dataset* utuh, yaitu: (1) *Mahasiswa Tradisional Berprestasi* — usia muda dengan pembiayaan eksternal dan tidak bekerja; (2) *Mahasiswa Pekerja Fleksibel* — seluruhnya bekerja dengan pembiayaan campuran; dan (3) *Mahasiswa Mandiri Adaptif* — hampir seluruhnya membiayai kuliah secara mandiri sambil bekerja dengan prestasi akademik tertinggi. Pembeda utama antar segmen terutama terletak pada pola pembiayaan pendidikan dan status keterlibatan kerja, sementara variabel usia dan sosial ekonomi berperan sebagai karakteristik deskriptif pendukung. Struktur segmentasi yang dihasilkan selanjutnya diterjemahkan ke dalam kerangka *Segmenting–Targeting–Positioning* (STP) untuk merumuskan rekomendasi strategi pemasaran yang relevan dan aplikatif. Temuan ini menegaskan bahwa segmentasi berbasis data dapat menjadi dasar yang lebih objektif dan efektif dalam perencanaan strategi pemasaran pendidikan tinggi.

1. PENDAHULUAN

Perguruan tinggi saat ini berada dalam lingkungan persaingan yang semakin kompetitif, baik antar institusi negeri maupun swasta. Beberapa studi Indonesia dan luar

negeri mengukur pengaruh bauran pemasaran (produk/prodi, harga, lokasi, promosi, orang, proses, bukti fisik) terhadap keputusan memilih perguruan tinggi [1,2]. Perubahan demografis, dinamika sosial ekonomi, serta meningkatnya pilihan jalur pendidikan menyebabkan karakteristik calon mahasiswa menjadi semakin beragam.

Kondisi ini menuntut perguruan tinggi untuk tidak lagi mengandalkan strategi pemasaran yang bersifat umum (*mass marketing*), melainkan mengarah pada pendekatan yang lebih tersegmentasi dan presisi.

Kondisi ini secara nyata dihadapi oleh Sekolah Tinggi Teknologi Pekanbaru (STTP), salah satu perguruan tinggi swasta yang beroperasi di tengah persaingan 44 perguruan tinggi di Kota Pekanbaru. Berdasarkan data penerimaan mahasiswa baru periode 2020–2023, tidak satu pun program studi STTP berhasil mencapai target ideal 100 mahasiswa per tahun ajaran, salah satunya program studi S1 Sistem Informasi yang hanya menerima 4 hingga 12 mahasiswa per tahun. Kondisi ini mengindikasikan bahwa strategi pemasaran yang selama ini diterapkan secara seragam belum mampu menjangkau segmen calon mahasiswa yang paling potensial secara tepat sasaran.

Dalam praktiknya, banyak strategi pemasaran perguruan tinggi masih disusun berdasarkan asumsi normatif atau pengalaman manajerial, seperti pengelompokan mahasiswa berdasarkan usia atau latar belakang pendidikan secara sederhana. Pendekatan tersebut cenderung mengabaikan kompleksitas dan tumpang tindih karakteristik mahasiswa, khususnya pada aspek pembiayaan pendidikan, status bekerja, serta kesiapan akademik. Akibatnya, strategi yang dirancang sering kali kurang tepat sasaran dan tidak sepenuhnya selaras dengan kebutuhan segmen mahasiswa yang dilayani. Studi di Mojokerto secara eksplisit mengembangkan strategi operasional dan pemasaran berbasis analisis statistik preferensi mahasiswa baru dan pemodelan sistem dinamis, bukan sekadar pengalaman pengelola [3].

Perkembangan *data-driven decision making* membuka peluang bagi perguruan tinggi untuk merumuskan strategi pemasaran yang lebih objektif melalui analisis data mahasiswa. Salah satu pendekatan yang relevan adalah segmentasi berbasis klusterisasi, yang memungkinkan pengelompokan mahasiswa berdasarkan pola kemiripan karakteristik secara empiris. Namun demikian, penggunaan satu metode klusterisasi sering kali menghadapi keterbatasan, terutama ketika data memiliki struktur kompleks, distribusi tidak seimbang, serta mengandung *noise* atau *outlier*. Tinjauan literatur mutakhir merekomendasikan segmentasi pasar, CRM, dan *digital/data-based marketing* (media sosial, analitik) sebagai kunci daya tarik perguruan tinggi [4].

Beberapa penelitian menunjukkan bahwa pendekatan klusterisasi bertahap (*two-stage clustering*) dapat meningkatkan kualitas segmentasi dengan memadukan keunggulan metode yang berbeda. Dalam konteks ini, metode berbasis kepadatan seperti HDBSCAN efektif

untuk mengidentifikasi struktur global data dan mendeteksi *noise*, tetapi kurang optimal dalam memetakan heterogenitas internal pada kluster mayoritas. Sebaliknya, metode probabilistik seperti *Gaussian Mixture Model* (GMM) mampu menangkap distribusi data yang saling tumpang tindih, sehingga lebih sesuai untuk eksplorasi sub-segmen mahasiswa yang memiliki karakteristik beririsan.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan optimasi strategi pemasaran perguruan tinggi melalui segmentasi mahasiswa berbasis data dengan pendekatan *hybrid* HDBSCAN–GMM. Segmentasi yang dihasilkan tidak hanya bertujuan untuk memetakan kelompok mahasiswa secara numerik, tetapi juga untuk menyediakan dasar empiris dalam penerapan kerangka pemasaran yaitu dengan *Segmenting–Targeting–Positioning* (STP). Dengan demikian, strategi pemasaran yang dirumuskan diharapkan tidak lagi bersifat subjektif, melainkan diturunkan secara sistematis dari karakteristik nyata masing-masing segmen mahasiswa. Pendekatan STP berbasis profil segmen empiris memungkinkan perumusan strategi yang lebih presisi dan dapat dipertanggungjawabkan secara metodologis [5].

2. TINJAUAN PUSTAKA

2.1. Segmentasi Mahasiswa dalam Pemasaran Perguruan Tinggi

Segmentasi pasar merupakan konsep fundamental dalam pemasaran strategis yang bertujuan untuk mengelompokkan pasar yang heterogen menjadi segmen-segmen yang relatif homogen agar strategi pemasaran dapat dirancang secara lebih efektif. Dalam konteks pendidikan tinggi, segmentasi mahasiswa menjadi semakin penting seiring dengan meningkatnya persaingan antar perguruan tinggi serta beragamnya karakteristik calon mahasiswa. Segmentasi yang tepat memungkinkan institusi pendidikan untuk menyesuaikan pesan komunikasi, penawaran program, serta layanan akademik sesuai dengan kebutuhan dan kemampuan masing-masing segmen.

Berbagai pendekatan segmentasi mahasiswa telah dikembangkan, mulai dari segmentasi demografis (usia, jenis kelamin), geografis, hingga sosio-ekonomi dan latar belakang pendidikan. Namun, pendekatan tersebut sering kali bersifat parsial dan tidak mampu menangkap kompleksitas karakteristik mahasiswa secara utuh. Selain itu, banyak praktik segmentasi masih bergantung pada asumsi atau klasifikasi manual, sehingga berpotensi menghasilkan segmentasi yang subjektif dan kurang replikatif. Riset Mojokerto menegaskan peran signifikan kualitas akademik, faktor internal psikologis, serta

teknologi dan informasi terhadap keputusan memilih kampus [3].

Dalam praktik pemasaran perguruan tinggi, keterbatasan segmentasi konvensional ini dapat berdampak pada rendahnya efektivitas strategi promosi, terutama ketika karakteristik mahasiswa saling tumpang tindih, misalnya antara mahasiswa pekerja dan non-pekerja, atau antara mahasiswa dengan sumber pembiayaan yang berbeda. Oleh karena itu, diperlukan pendekatan segmentasi yang mampu memanfaatkan data mahasiswa secara komprehensif dan objektif. Penggunaan *cluster analysis*, *fuzzy C-means*, dan *data mining* (K-Means, SVM) memungkinkan tiap mahasiswa memiliki kedekatan ke beberapa klaster dan menangkap pola tumpang tindih dengan lebih baik [6,7].

2.2. Segmentasi Berbasis Data dan Klusterisasi

Perkembangan data *analytics* mendorong pergeseran dari segmentasi berbasis asumsi menuju segmentasi berbasis data (*data-driven segmentation*). Pendekatan ini memanfaatkan teknik klusterisasi untuk mengelompokkan individu berdasarkan kemiripan pola karakteristik yang diobservasi secara empiris. Dalam konteks pendidikan tinggi, klusterisasi telah digunakan untuk mengidentifikasi profil mahasiswa berdasarkan performa akademik, kondisi sosial ekonomi, keterlibatan kerja, serta pola pembiayaan pendidikan. Banyak studi menggambarkan pergeseran dari segmentasi statis (demografis/geografis, *rule-based*) menuju segmentasi dinamis berbasis perilaku dan *machine learning*, karena metode lama sulit menangkap kompleksitas perilaku konsumen modern [8,9].

Metode klusterisasi yang umum digunakan, seperti k-means, memiliki keunggulan dalam kesederhanaan dan efisiensi komputasi. Namun demikian, metode tersebut mengasumsikan bentuk klaster yang homogen dan memerlukan penentuan jumlah klaster (k) di awal. Asumsi ini sering kali tidak sesuai dengan karakteristik data mahasiswa yang bersifat heterogen, tidak seimbang, dan mengandung tumpang tindih antar kelompok. Dalam berbagai industri (ritel, *e-commerce*, *hospitality*, asuransi, kesehatan), algoritma *clustering* (*K-means*, *hierarchical*, DBSCAN, *self-organizing maps*), PCA/SVD, dan teknik campuran digunakan untuk menemukan segmen yang tidak tampak dengan cara tradisional dan sering kali lebih halus serta lebih mudah ditindaklanjuti [9,10,11].

Keterbatasan metode klusterisasi tunggal menjadi semakin nyata ketika data memiliki struktur kompleks dan *noise*. Dalam kondisi tersebut, hasil segmentasi cenderung tidak stabil dan sulit diinterpretasikan secara praktis. Oleh karena itu, penelitian-penelitian terbaru mulai mengarah pada penggunaan pendekatan klusterisasi yang lebih

fleksibel dan adaptif terhadap struktur data. K-means mengasumsikan klaster cenderung bulat/konveks sehingga gagal pada klaster non-linier seperti lingkaran konsentris, sementara DBSCAN dapat menangani bentuk bebas tetapi sangat sensitif pada pemilihan parameter dan kesulitan membedakan klaster berdensitas berbeda [12]. *Hierarchical single linkage* bagus untuk bentuk memanjang, tetapi mudah terkena efek “*chaining*” di bawah *noise* [13].

2.3. HDBSCAN dan Gaussian Mixture Model dalam Segmentasi

HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) merupakan pengembangan dari metode *density-based clustering* yang dirancang untuk menangani data dengan kepadatan tidak seragam. Keunggulan utama HDBSCAN terletak pada kemampuannya dalam mengidentifikasi klaster dengan ukuran tidak seimbang, mendeteksi *noise* atau *outlier*, serta tidak memerlukan penentuan jumlah klaster secara a priori. Dalam konteks data mahasiswa, HDBSCAN efektif untuk mengeksplorasi struktur global data dan memisahkan kelompok minoritas yang tidak merepresentasikan pola utama populasi. HDBSCAN secara eksplisit membedakan titik klaster vs *noise* atau *outlier*, tidak memaksa semua titik masuk ke klaster [14].

Namun demikian, metode berbasis kepadatan seperti HDBSCAN memiliki keterbatasan dalam memetakan heterogenitas internal pada klaster mayoritas. Klaster besar yang terbentuk sering kali masih menyimpan variasi karakteristik yang signifikan, sehingga kurang informatif jika langsung digunakan sebagai dasar perumusan strategi pemasaran. Algoritma ini sensitif terhadap kelompok kecil yang lebih rapat dari latar belakang data, sehingga dapat menemukan kelompok minoritas yang signifikan sekaligus menandai titik tersebar sebagai *noise* [15]. Namun, pada data yang sebagian besar tidak membentuk klaster, HDBSCAN berpotensi menghasilkan banyak klaster kecil yang bersifat *false positive* dan memerlukan pasca-proses atau *threshold* tambahan untuk menghindari fragmentasi berlebihan [16].

Sebaliknya, *Gaussian Mixture Model* (GMM) merupakan metode klusterisasi probabilistik yang mengasumsikan bahwa data berasal dari campuran beberapa distribusi Gaussian. Pendekatan ini memungkinkan setiap observasi memiliki probabilitas keanggotaan pada lebih dari satu klaster, sehingga mampu menangkap karakteristik yang saling tumpang tindih. Dalam segmentasi mahasiswa, GMM relevan untuk mengidentifikasi sub-segmen yang tidak terpisah secara tegas, misalnya antara mahasiswa dengan pembiayaan campuran atau status kerja parsial. GMM memodelkan populasi sebagai campuran beberapa

distribusi Gaussian, di mana tiap komponen merepresentasikan subkelompok mahasiswa yang laten/tidak teramati langsung. Secara teoritis, campuran Gaussian dapat mendekati hampir semua distribusi berkelanjutan, sehingga fleksibel untuk pola perilaku dan pendanaan mahasiswa yang heterogen [17]. Pada data yang sangat tumpang tindih, solusi GMM dapat menghasilkan hasil yang tidak optimal apabila asumsi Gaussian sangat dilanggar, perlu perhatian pada pemilihan model dan penalti untuk menghindari klaster yang sulit diinterpretasi [18].

2.4. Pendekatan Two-Stage Clustering dalam Segmentasi Mahasiswa

Pendekatan *two-stage clustering* mengombinasikan keunggulan metode klasterisasi yang berbeda dalam satu kerangka analisis bertahap. Tahap awal digunakan untuk mengeksplorasi struktur makro data dan mengurangi kompleksitas, sementara tahap lanjutan difokuskan pada pemurnian dan pendalaman segmentasi. Pendekatan ini dinilai lebih *robust* dibandingkan penggunaan satu metode klasterisasi tunggal, terutama pada data berdimensi tinggi dan mengandung *noise*. Pendekatan dua tahap terbukti memberikan sejumlah keunggulan dibandingkan metode tunggal. Pertama, kinerja dan skalabilitas yang lebih baik karena dua tahap sering menurunkan kompleksitas komputasi sambil menjaga atau meningkatkan kualitas klaster [19]. Kedua, kualitas klaster yang meningkat, di mana indeks Silhouette, Calinski–Harabasz, Davies–Bouldin, dan metrik lain biasanya membaik dibandingkan satu metode tunggal [20,21]. Ketiga, fleksibilitas data yang lebih tinggi karena mampu menangani data besar, berdimensi tinggi, bentuk klaster arbitrer, dan campuran variabel numerik–kategorikal [21,22].

Dalam penelitian ini, HDBSCAN digunakan pada tahap awal untuk mengidentifikasi klaster mayoritas mahasiswa serta menyaring *noise* yang berpotensi mengaburkan struktur data. Selanjutnya, GMM diterapkan secara khusus pada klaster mayoritas untuk mengeksplorasi heterogenitas internal dan menghasilkan sub-segmen mahasiswa yang lebih homogen dan interpretatif. Kombinasi ini memungkinkan segmentasi yang tidak hanya valid secara statistik, tetapi juga relevan secara praktis dalam konteks pemasaran perguruan tinggi. Dengan demikian, tinjauan pustaka ini menegaskan bahwa segmentasi mahasiswa berdasarkan data dengan pendekatan *two-stage clustering* memiliki landasan teoritis dan metodologis yang kuat, serta berpotensi menghasilkan segmentasi yang lebih objektif dibandingkan pendekatan konvensional.

3. METODOLOGI

3.1. Data dan Variabel Penelitian

Penelitian ini menggunakan data mahasiswa aktif yang diperoleh dari basis data akademik perguruan tinggi. Data yang digunakan mencerminkan karakteristik demografis, akademik, dan sosial ekonomi mahasiswa yang relevan dalam konteks perumusan strategi pemasaran pendidikan tinggi. Seluruh data yang dianalisis bersifat anonim dan digunakan hanya untuk kepentingan penelitian.

Variabel yang digunakan dalam proses segmentasi meliputi usia awal masuk mahasiswa, latar belakang sosial ekonomi, status keterlibatan kerja, sumber pembiayaan pendidikan, serta indikator prestasi akademik. Variabel-variabel tersebut dipilih karena secara empiris sering dikaitkan dengan perilaku pemilihan perguruan tinggi dan respons mahasiswa terhadap strategi promosi. Dalam penelitian ini, variabel usia tidak diposisikan sebagai pembeda utama klaster, melainkan sebagai informasi deskriptif yang memperkaya profil segmen mahasiswa.

3.2. Pra-pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan kualitas data sebelum proses klasterisasi. Langkah-langkah yang dilakukan meliputi pembersihan data dari nilai hilang dan anomali, transformasi variabel kategorikal ke dalam bentuk numerik yang sesuai, serta standarisasi variabel numerik untuk menghindari dominasi skala tertentu dalam proses analisis.

Selain itu, dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) pada tahap awal analisis. Reduksi dimensi ini bertujuan untuk mereduksi kompleksitas data, mengurangi redundansi antar variabel, serta membantu proses eksplorasi struktur global data pada tahap klasterisasi awal. PCA tidak digunakan sebagai dasar interpretasi klaster, melainkan sebagai alat bantu untuk meningkatkan stabilitas dan efisiensi proses klasterisasi berbasis kepadatan.

3.3. Klasterisasi Tahap Pertama Menggunakan HDBSCAN

Pada tahap pertama, metode HDBSCAN digunakan untuk mengeksplorasi struktur makro data mahasiswa. HDBSCAN dipilih karena kemampuannya dalam mengidentifikasi klaster dengan ukuran tidak seimbang, mendeteksi *noise* atau *outlier*, serta tidak memerlukan penentuan jumlah klaster secara a priori. Karakteristik ini sesuai dengan kondisi data mahasiswa yang bersifat heterogen dan berpotensi mengandung kelompok minoritas.

Hasil klusterisasi HDBSCAN digunakan untuk mengidentifikasi kluster mayoritas yang merepresentasikan sebagian besar populasi mahasiswa. Kluster mayoritas ini selanjutnya menjadi fokus analisis lanjutan, sementara kluster minoritas dan *noise* tidak digunakan dalam tahap segmentasi berikutnya. Pendekatan ini bertujuan untuk menghindari distorsi hasil segmentasi akibat pengaruh data ekstrim yang tidak mencerminkan pola umum populasi mahasiswa.

Pemilihan konfigurasi parameter HDBSCAN dilakukan melalui eksplorasi sistematis terhadap kombinasi jumlah komponen PCA (rentang 30–80 komponen), nilai *min_cluster_size* (5–10), dan *min_samples* (1, 3, dan 5). Evaluasi setiap konfigurasi dilakukan berdasarkan empat kriteria: variansi data yang dipertahankan, jumlah kluster yang terbentuk, rasio *noise*, serta nilai *Silhouette Score* dan *Davies-Bouldin Index*. Konfigurasi optimal yang dipilih adalah 60 komponen PCA dengan *min_cluster_size*=5 dan *min_samples*=1, yang menghasilkan variansi sebesar 94,25%, tujuh kluster valid, dan rasio *noise* terendah sebesar 6,16%.

3.4. Klusterisasi Tahap Kedua Menggunakan Gaussian Mixture Model

Setelah kluster mayoritas diidentifikasi melalui HDBSCAN, dilakukan segmentasi lanjutan menggunakan GMM untuk mengeksplorasi heterogenitas internal kluster secara probabilistik. GMM memodelkan data sebagai campuran dari beberapa distribusi Gaussian, sehingga setiap observasi memiliki peluang keanggotaan pada lebih dari satu kluster. Secara matematis, distribusi campuran pada GMM dapat dinyatakan sebagai berikut:

$$p(x|\Theta) = \sum_{k=1}^K \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{1}$$

dengan:

- x menyatakan vektor fitur mahasiswa,
- K adalah jumlah kluster,
- π_k merupakan bobot kluster ke- k dengan $\sum_{k=1}^K \pi_k = 1$,
- μ_k dan Σ_k masing-masing adalah vektor mean dan matriks kovarians kluster ke- k ,
- Θ menyatakan himpunan parameter model GMM.

Pendekatan probabilistik ini memungkinkan GMM menangkap karakteristik mahasiswa yang saling tumpang tindih, khususnya pada variabel pembiayaan pendidikan, status kerja, dan kesiapan akademik. Proses estimasi parameter dilakukan menggunakan algoritma *Expectation–Maximization* (EM), namun detail algoritmik

tidak dibahas lebih lanjut karena tidak menjadi fokus utama penelitian ini. Penggunaan GMM pada penelitian ini didasarkan pada sifatnya sebagai *universal approximator*, campuran distribusi Gaussian dengan jumlah komponen yang memadai dapat mendekati distribusi probabilitas apapun, termasuk distribusi non-normal [23].

Oleh karena itu, GMM tidak mensyaratkan keseluruhan data berdistribusi normal secara ketat, melainkan hanya mensyaratkan bahwa setiap komponen kluster dapat dimodelkan sebagai distribusi Gaussian lokal. Kondisi ini terpenuhi pada data mahasiswa STTP meskipun hasil uji normalitas Shapiro-Wilk menunjukkan bahwa seluruh tujuh variabel input tidak berdistribusi normal secara global ($p < 0,05$). Selain itu, matriks kovarians per komponen (Σ_k) pada GMM memungkinkan setiap kluster memiliki bentuk dan orientasi distribusi yang berbeda, sehingga lebih adaptif dibandingkan K-Means yang mengasumsikan kluster berbentuk bulat seragam [24].

3.5. Penentuan Jumlah Kluster

Penentuan jumlah kluster optimal pada GMM dilakukan menggunakan *Bayesian Information Criterion* (BIC) dan *Akaike Information Criterion* (AIC). Kedua kriteria ini digunakan untuk menilai keseimbangan antara kecocokan model terhadap data dan kompleksitas model yang digunakan.

Nilai BIC dan AIC dirumuskan sebagai berikut:

$$\begin{aligned} \text{BIC} &= -2 \ln(L) + p \ln(n) \\ \text{AIC} &= -2 \ln(L) + 2p \end{aligned} \tag{2}$$

dengan:

- L menyatakan nilai likelihood maksimum model,
- p adalah jumlah parameter dalam model,
- n merupakan jumlah observasi.

Model dengan nilai BIC dan AIC yang lebih rendah menunjukkan keseimbangan yang lebih baik antara kompleksitas dan kecocokan model. Dalam penelitian ini, evaluasi dilakukan pada beberapa alternatif jumlah kluster. Pemilihan jumlah kluster akhir tidak hanya didasarkan pada nilai minimum BIC dan AIC, tetapi juga mempertimbangkan interpretabilitas hasil serta relevansinya dalam perumusan strategi pemasaran perguruan tinggi.

3.6. Kerangka Analisis Strategi Pemasaran

Sebagai tahap akhir analisis, hasil segmentasi mahasiswa diterjemahkan ke dalam kerangka *Segmenting–Targeting–Positioning* (STP). Segmentasi diperoleh secara objektif melalui hasil klusterisasi GMM, *targeting* ditentukan

berdasarkan ukuran dan relevansi masing-masing segmen terhadap kapasitas institusi, sedangkan *positioning* dirumuskan berdasarkan kebutuhan dan nilai dominan dari setiap segmen mahasiswa. Pendekatan STP dalam penelitian ini tidak digunakan sebagai alat perencanaan normatif, melainkan sebagai kerangka analitis untuk menginterpretasikan hasil segmentasi berbasis data. Dengan demikian, rekomendasi strategi pemasaran yang dihasilkan merupakan turunan langsung dari karakteristik empiris kluster mahasiswa.

4. HASIL DAN PEMBAHASAN

4.1. Eksplorasi Struktur Data Mahasiswa Menggunakan HDBSCAN

Tahap awal analisis dilakukan menggunakan metode HDBSCAN untuk mengeksplorasi struktur global data mahasiswa dan mengidentifikasi pola kepadatan utama. Data yang dianalisis terdiri atas 731 observasi mahasiswa dengan 97 variabel, yang mencakup karakteristik demografis, sosial ekonomi, akademik, dan perilaku mahasiswa. Tingginya dimensi data tersebut berpotensi memengaruhi stabilitas metode klusterisasi berbasis jarak maupun kepadatan.

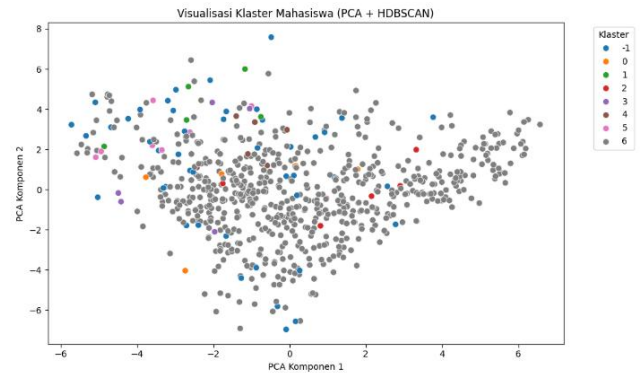
Untuk mengatasi kompleksitas dan redundansi antar variabel, dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) pada tahap awal analisis. PCA digunakan sebagai alat bantu untuk memproyeksikan data ke ruang berdimensi lebih rendah guna meningkatkan stabilitas proses klusterisasi, tanpa menghilangkan pola kepadatan utama dalam data. Reduksi dimensi ini tidak digunakan sebagai dasar interpretasi kluster, melainkan sebagai prasyarat teknis dalam eksplorasi struktur data menggunakan HDBSCAN. Pemilihan jumlah komponen PCA dilakukan melalui eksperimen sistematis pada rentang 30 hingga 80 komponen. Setiap konfigurasi dievaluasi berdasarkan variansi yang dipertahankan, jumlah kluster yang terbentuk, rasio *noise*, *Silhouette Score*, dan DBI. Hasil eksperimen pada konfigurasi terpilih disajikan pada Tabel 1.

Tabel 1. Eksperimen Pemilihan Jumlah Komponen PCA

Komponen PCA	Variansi (%)	Jumlah Kluster	Noise (%)	Nilai Silhouette	DBI
40	77,81	7	12,59	0,1755	1,3363
50	87,10	16	34,34	0,1048	1,5404
60	94,25	7	6,16	0,2114	1,3325
65	96,87	21	33,52	0,0977	1,5399
80	99,89	6	6,98	0,1893	1,3793

Konfigurasi 60 komponen PCA dipilih karena menghasilkan rasio noise terendah (6,16%), *Silhouette Score* tertinggi (0,2114), dan DBI terbaik (1,3325) di antara seluruh konfigurasi yang menghasilkan jumlah kluster yang interpretatif. Pada hasil klusterisasi

HDBSCAN menunjukkan keberadaan satu kluster mayoritas yang merepresentasikan sebagian besar populasi mahasiswa, serta sejumlah kecil observasi yang diklasifikasikan sebagai *noise*.



Gambar 1. Scatter Plot Kluster HDBSCAN

Gambar 1. Memperlihatkan visualisasi struktur kepadatan data mahasiswa menggunakan HDBSCAN pada ruang dua dimensi hasil reduksi PCA. Terlihat satu kluster mayoritas (abu-abu) yang merepresentasikan pola utama populasi mahasiswa serta sejumlah observasi yang terklasifikasi sebagai *noise* (biru), yang menunjukkan adanya karakteristik ekstrem di luar struktur utama data.

Hasil eksplorasi menggunakan HDBSCAN menunjukkan bahwa kluster mayoritas mencakup 651 dari 731 observasi mahasiswa, sedangkan sisanya terklasifikasi sebagai kluster minoritas dan *noise*. Temuan ini mengindikasikan bahwa meskipun data mahasiswa memiliki struktur utama yang relatif dominan, terdapat pula karakteristik ekstrem yang tidak sepenuhnya sejalan dengan pola umum populasi. Kualitas hasil klusterisasi HDBSCAN dievaluasi menggunakan *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Mengingat sebesar 89% data terkonsentrasi pada satu kluster mayoritas, evaluasi dilakukan secara global maupun per kelompok kluster untuk memberikan gambaran yang lebih representatif. Hasil evaluasi disajikan pada Tabel 2.

Tabel 2. Ringkasan Nilai Evaluasi Kluster HDBSCAN

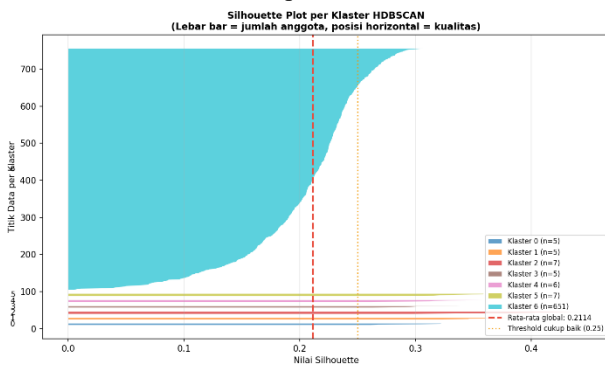
Metrik Evaluasi	Nilai	Interpretasi
Silhouette Score (Global)	0,2114	Di bawah 0,25, namun wajar pada distribusi kluster sangat tidak seimbang
Silhouette Score Kluster Minoritas (0–5)	0,29–0,39	Cukup baik — kluster minoritas terkohesi dan terpisah dengan memadai
Silhouette Score Kluster Mayoritas (6)	0,2049	Dipengaruhi dominasi 89% anggota populasi
5Davies-Bouldin Index (DBI)	1,3325	Acceptable — kluster relatif kompak dan terpisah
Noise Rate	6,16%	Baik — jauh di bawah batas toleransi 15–20%

Nilai *Silhouette* global yang berada di bawah 0,25 perlu diinterpretasikan dengan mempertimbangkan bahwa dominasi satu kluster mayoritas secara struktural menekan nilai rata-rata global. Kaufman dan Rousseeuw [25] secara eksplisit menyatakan bahwa *Silhouette Score* tidak

dirancang untuk data dengan kluster berukuran sangat timpang. Eksperimen penambahan fitur pemisah tidak menghasilkan peningkatan Silhouette yang berarti (selisih < 0,003), mengindikasikan bahwa nilai ini merupakan karakteristik intrinsik distribusi populasi mahasiswa STTP, bukan kekurangan metode.

Distribusi nilai *Silhouette* per kluster juga divisualisasikan pada Gambar 2, yang menunjukkan bahwa kluster-kluster minoritas (0–5) memiliki nilai *Silhouette* yang lebih tinggi (0,29–0,39) dibandingkan kluster mayoritas (6), mengonfirmasi bahwa kualitas pengelompokan pada segmen-segmen spesifik telah memadai meskipun nilai global tertekan oleh dominasi kluster 6.

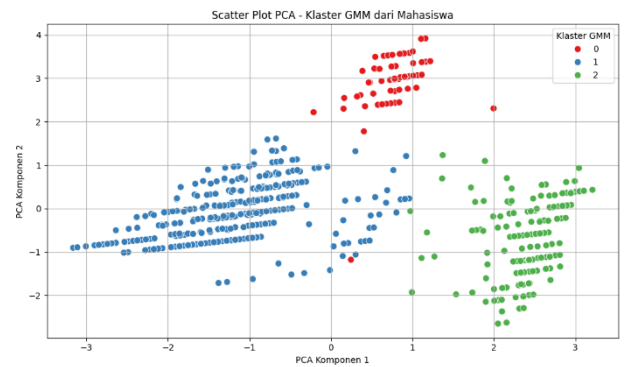
Gambar 2. Silhouette Plot per Kluster HDBSCAN



4.2. Segmentasi Mahasiswa Menggunakan GMM

Penerapan metode GMM menghasilkan tiga kelompok mahasiswa dengan jumlah anggota yang berbeda-beda. Sebelum menetapkan jumlah kluster, terlebih dahulu dilakukan evaluasi menggunakan dua kriteria statistik, yaitu AIC dan BIC. Evaluasi ini menggunakan dua kriteria yang umum dalam pemodelan statistik berbasis campuran, yaitu AIC dan BIC, untuk menilai kemampuan GMM dalam menangkap struktur segmentasi mahasiswa pada data yang masih mengandung heterogenitas tinggi. Pendekatan ini bertujuan untuk mengevaluasi kemampuan GMM dalam menangkap struktur segmentasi mahasiswa secara langsung pada data yang masih mengandung heterogenitas tinggi dan observasi ekstrem.

Namun demikian, hasil segmentasi pada tahap ini masih menunjukkan tingkat heterogenitas internal yang relatif tinggi pada beberapa kluster, khususnya pada variabel pembiayaan dan latar belakang non-tradisional.



Gambar 3. Scatter Plot Hasil Kluster GMM

Temuan ini mengindikasikan bahwa meskipun GMM mampu mengidentifikasi pola segmentasi pada dataset utuh, keberadaan observasi ekstrem dan variasi karakteristik yang tinggi berpotensi memengaruhi ketajaman dan stabilitas hasil klusterisasi.

4.3. Perbandingan Segmentasi GMM pada Dataset Utuh dan Dataset Terfilter HDBSCAN

Untuk mengevaluasi efektivitas pendekatan *two-stage clustering*, dilakukan perbandingan antara hasil segmentasi GMM yang diterapkan langsung pada dataset utuh dan GMM yang diterapkan pada dataset hasil penyaringan HDBSCAN. Hasil perbandingan tersebut dirangkum dalam Tabel 3.

Tabel 3. Perbandingan GMM dataset utuh vs GMM HDBSCAN 6 [3]

Aspek Utama	GMM Dataset Utuh	GMM HDBSCAN-6	Perbedaan Penting
Kluster Besar	±459 mahasiswa. Mayoritas mandiri (≈98%). Usia tersebar. Tampak homogen, padahal internal variatif.	±419 mahasiswa. Mayoritas mandiri tapi lebih jelas terbagi dari kluster lain. Usia tersebar serta relatif lebih dewasa.	Sama-sama dominan mandiri, tapi di HDBSCAN-6 variasi internal kluster besar “disaring keluar” sehingga lebih murni kelompok mandiri.
Kluster Sedang	±198 mahasiswa. Didukung ortu & beasiswa (parsial/penuh). Usia lebih muda (20–21 th). Ada yang bekerja.	±170 mahasiswa. Didukung ortu/beasiswa juga. Usia muda. Aksesibilitas tinggi.	Mirip segmen, tapi pada HDBSCAN-6 lebih jelas konsistensinya (muda + ortu/beasiswa), sedangkan pada GMM dataset utuh masih bercampur dengan pekerja.
Kluster Kecil	±74 mahasiswa. Profil pekerja dengan pembiayaan campuran/beasiswa. Rata-rata usia sedang dan cukup menyebar, namun distribusi usia heterogen dan tidak menunjukkan dominasi usia tertentu.	±62 mahasiswa. Profil pekerja dengan pembiayaan campuran/wali. Rata-rata usia sedang, namun sebaran usia luas dan tidak terkonsentrasi pada kelompok usia tua.	Filtering HDBSCAN memperjelas karakteristik pembiayaan dan status non-tradisional, sementara usia tetap bersifat deskriptif.

Berdasarkan Tabel 3, segmentasi GMM pada dataset terfilter HDBSCAN menghasilkan kluster dengan karakteristik yang lebih homogen dan pembeda yang lebih konsisten, terutama pada variabel pembiayaan dan status

kerja mahasiswa. Sebaliknya, segmentasi GMM pada dataset utuh menunjukkan tingkat tumpang tindih karakteristik yang lebih tinggi antar klaster, yang menyulitkan interpretasi dan pemanfaatan hasil segmentasi dalam konteks strategis. Temuan ini mengonfirmasi bahwa keberadaan *noise* dan kelompok minoritas dalam dataset utuh dapat memengaruhi estimasi parameter GMM. Dengan demikian, penggunaan HDBSCAN sebagai tahap awal terbukti berkontribusi dalam meningkatkan kualitas segmentasi lanjutan, baik dari sisi validitas statistik maupun kegunaan praktis.

4.4. Hasil Segmentasi Mahasiswa Menggunakan GMM pada Klaster Mayoritas HDBSCAN

Berdasarkan hasil penyaringan awal menggunakan HDBSCAN, analisis lanjutan difokuskan pada 651 observasi mahasiswa yang termasuk dalam klaster mayoritas. Subset data ini selanjutnya dianalisis menggunakan GMM untuk mengeksplorasi heterogenitas internal populasi mahasiswa dalam kerangka *two-stage clustering*. Segmentasi lanjutan dilakukan menggunakan GMM terhadap klaster mayoritas hasil HDBSCAN. Evaluasi jumlah klaster menggunakan *Bayesian Information Criterion* (BIC) dan *Akaike Information Criterion* (AIC) menunjukkan bahwa solusi dengan tiga klaster memberikan keseimbangan terbaik antara kecocokan model dan interpretabilitas. Evaluasi jumlah klaster menggunakan AIC dan BIC dilakukan pada rentang dua hingga enam klaster. Hasil evaluasi disajikan pada Tabel 4, dan laju penurunan nilai antar penambahan klaster pada Tabel 5.

Tabel 4. Nilai AIC dan BIC GMM pada Klaster Mayoritas HDBSCAN

Klaster GMM	BIC	AIC
2	-19,327	-20,137
3	-22,081	-23,299
4	-28,879	-30,505
5	-34,584	-36,617
6	-38,179	-40,619

Tabel 5. Laju Penurunan Nilai AIC dan BIC per Penambahan Klaster

Penambahan Klaster	Penurunan BIC	Penurunan AIC	Interpretasi
2 → 3	2,754	3,162	Substansial — peningkatan bermakna
3 → 4	6,798	7,206	Besar namun memerlukan justifikasi interpretasi
4 → 5	5,705	6,112	Penurunan mengecil
5 → 6	3,595	4,002	Penurunan semakin mengecil

Pemilihan tiga klaster didasarkan pada konvergensi tiga pertimbangan: (1) penurunan BIC dari 2 ke 3 klaster yang substansial, (2) ketiga klaster menghasilkan profil yang dapat dibedakan secara bermakna, dan (3) tiga segmen masih dapat dikelola secara operasional oleh tim PMB STTP.

Hasil segmentasi ini selanjutnya dianalisis sebagai segmentasi utama mahasiswa.

Tabel 6. Hasil Kluster GMM terhadap Kluster Mayoritas HDBSCAN

Klaster GMM	Jumlah Mahasiswa	Usia Masuk	Sosial Ekonomi	Prestasi Akademik	Bekerja	Pembiayaan Dominan	Label Segmen
GMM-0	170	19.82	1.02	69%	2%	Beasiswa + Ortu	Mahasiswa Tradisional Berprestasi
GMM-1	62	22.06	1.42	61%	98%	Campuran 100% (semua)	Mahasiswa Pekerja Fleksibel
GMM-2	419	29.08	0.65	78%	94%	Mandiri (98%)	Mahasiswa Mandiri Adaptif

Tabel 6. menyajikan ringkasan karakteristik masing-masing klaster hasil GMM terhadap klaster mayoritas HDBSCAN. Perbedaan antar klaster terutama terlihat pada variabel pembiayaan pendidikan, status keterlibatan kerja, dan indikator kesiapan akademik. Berdasarkan hasil analisis, tiga sub-segmen mahasiswa teridentifikasi dengan karakteristik pembeda yang kuat, terutama pada variabel pembiayaan pendidikan dan status keterlibatan kerja.

GMM-0 (Mahasiswa Tradisional Berprestasi) terdiri dari 170 mahasiswa yang seluruh pembiayaannya bersumber dari luar diri mahasiswa — 55,3% dari orang tua, 27,6% beasiswa penuh, dan 14,7% beasiswa tidak penuh. Hampir tidak ada yang bekerja (2%), dan usia masuk sangat terkonsentrasi pada kelompok muda dengan 74,1% berusia ≤20 tahun. Segmen ini juga menunjukkan proporsi prestasi non-akademik tertinggi di antara tiga segmen (86%). GMM-1 (Mahasiswa Pekerja Fleksibel) terdiri dari 62 mahasiswa dengan identitas yang paling eksklusif, dimana seluruhnya (100%) menggunakan pembiayaan campuran yang menggabungkan berbagai sumber sekaligus, dan hampir seluruhnya sudah bekerja (98%). Kombinasi dua karakteristik ini tidak dimiliki oleh segmen lain secara bersamaan, menjadikan GMM-1 sebagai segmen yang paling mudah diidentifikasi secara empiris.

GMM-2 (Mahasiswa Mandiri Adaptif) merupakan segmen terbesar dengan 419 mahasiswa, dengan pembiayaan hampir sepenuhnya mandiri (98,6%) dan mayoritas sudah bekerja (94%). Tidak ada satu pun mahasiswa dalam segmen ini yang menggunakan pembiayaan orang tua, beasiswa, campuran, maupun wali. Menariknya, segmen ini memiliki proporsi prestasi akademik tertinggi di antara tiga segmen (78%) meskipun hampir seluruhnya membiayai kuliah sendiri sambil bekerja, mengindikasikan adaptabilitas yang tinggi. Untuk memastikan bahwa perbedaan karakteristik antar klaster tidak terjadi secara acak, dilakukan validasi statistik menggunakan uji Kruskal-Wallis terhadap seluruh variabel input GMM.

Hasil uji menunjukkan bahwa variabel pembiayaan mandiri, pembiayaan campuran, pembiayaan orang tua,

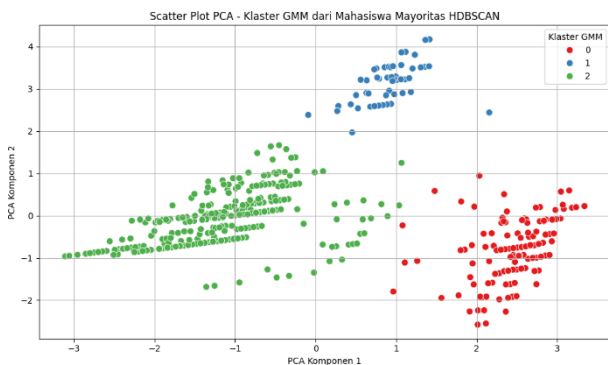
dan status bekerja memiliki perbedaan yang signifikan secara statistik antar kluster ($p < 0,05$) dengan *effect size* besar berdasarkan konvensi Eta-squared ($\eta^2 \geq 0,06$). Validasi ini mengonfirmasi bahwa ketiga segmen GMM yang dihasilkan bukan merupakan artefak algoritma, melainkan mencerminkan kelompok yang nyata dan dapat dibedakan dalam populasi mahasiswa STTP. Untuk memverifikasi bahwa perbedaan karakteristik antar kluster signifikan secara statistik, dilakukan uji Kruskal-Wallis terhadap seluruh variabel input GMM. Ringkasan hasil uji dan klasifikasi variabel disajikan pada Tabel 7 berikut.

Tabel 7. Klasifikasi Validitas Variabel Berdasarkan Uji Kruskal-Wallis

Variabel	p-value	Eta-Squared (η^2)	Klasifikasi
Pembiayaan Mandiri	< 0,001	> 0,14	Pembeda utama
Pembiayaan Campuran	< 0,001	> 0,14	Pembeda utama
Pembiayaan Orang Tua	< 0,001	> 0,14	Pembeda utama
Status Bekerja	< 0,001	> 0,14	Pembeda utama
Usia Awal Masuk	< 0,05	0,01–0,06	Pembeda pendukung (hanya pada GMM-0)
Status Sosial Ekonomi	> 0,05	< 0,01	Tidak valid sebagai pembeda
Prestasi Akademik	< 0,05	0,01–0,06	Pembeda pendukung

Hasil uji mengonfirmasi bahwa variabel pembiayaan dan status bekerja memiliki perbedaan yang signifikan secara statistik antar kluster ($p < 0,001$) dengan *effect size* besar ($\eta^2 > 0,14$ berdasarkan konvensi Cohen, 1988). Hal ini membuktikan bahwa ketiga segmen yang dihasilkan bukan artefak algoritma, melainkan mencerminkan kelompok yang nyata dalam populasi mahasiswa STTP.

Menjadi catatan penting bahwa variabel sosial ekonomi tidak dapat dijadikan pembeda utama antar kluster karena distribusinya tersebar atau bimodal di semua segmen. Demikian pula variabel usia hanya valid sebagai pembeda pada GMM-0 (std. deviasi $\pm 2,48$ tahun), namun tidak valid untuk GMM-2 (std. deviasi $\pm 8,46$ tahun). Variabel prestasi akademik dan non-akademik berperan sebagai variabel pendukung yang memperkaya interpretasi profil segmen.



Gambar 4. Scatter Plot Hasil Kluster GMM pada Kluster mayoritas HDBSCAN

Visualisasi hasil segmentasi GMM terhadap kluster mayoritas HDBSCAN ditunjukkan pada Gambar 4. Visualisasi ini memperlihatkan bahwa kluster hasil GMM memiliki pemisahan yang lebih jelas dibandingkan kluster awal HDBSCAN, khususnya dalam memetakan heterogenitas internal populasi mahasiswa. Temuan ini menunjukkan bahwa pendekatan *two-stage clustering* mampu menghasilkan segmentasi yang lebih informatif dibandingkan klusterisasi satu tahap.

4.5. Implikasi Segmentasi Berbasis Data terhadap Strategi STP

Hasil segmentasi mahasiswa berbasis GMM terhadap kluster mayoritas HDBSCAN selanjutnya diterjemahkan ke dalam kerangka *Segmenting–Targeting–Positioning* (STP). Proses segmenting dilakukan secara objektif berdasarkan hasil klusterisasi, sehingga setiap segmen memiliki dasar empiris yang jelas dan terukur.

Tabel 8. Penerapan Hasil Kluster ke STP

Kluster GMM	Segmenting (Karakteristik Utama Tervalidasi)	Targeting (Sasaran Kampus)	Positioning (Citra Kampus yang Dibentuk)	Dasar Positioning
GMM-0	<ul style="list-style-type: none"> Usia muda (~19.8 thn) Sosial-ekonomi menengah (1.02) Prestasi tinggi (0.69) Pembiayaan: Beasiswa & Orang Tua Tidak bekerja (2%) 	<ul style="list-style-type: none"> Lulusan SMA/SMK berprestasi Aktif online Peminat beasiswa 	Kampus unggulan berprestasi dan berorientasi akademik. Tempat terbaik melanjutkan studi setelah lulus sekolah.	Prestasi non-akademik tertinggi + pembiayaan beasiswa → responsif terhadap narasi prestasi dan kampus aktif
GMM-1	<ul style="list-style-type: none"> Usia dewasa (~22.06 thn) Sosial-ekonomi tinggi (1.42) Bekerja (98%) Pembiayaan campuran 100% Prestasi sedang 	<ul style="list-style-type: none"> Profesional muda & mahasiswa pekerja Alumni SMA gap year Mahasiswa sambil kerja 	Kampus fleksibel dan efisien untuk mahasiswa pembiayaan mandiri & bekerja. Tempat studi tanpa mengganggu pekerjaan	Bekerja 98% + campuran → kekhawatiran utama adalah konflik kuliah-pekerjaan
GMM-2	<ul style="list-style-type: none"> Usia tertinggi (~29.08 thn) Sosial-ekonomi (0.65) Pembiayaan mandiri (98%) Bekerja (94%) Prestasi baik 	<ul style="list-style-type: none"> Lulusan SMA non-tradisional Masyarakat ekonomi bawah Peminat beasiswa & kuliah terjangkau 	Kampus hemat, inklusif, dan profesional. Tempat pendidikan berkualitas untuk semua kalangan	Mandiri 98,6% + sosek rendah-menengah → keputusan kuliah berbasis kalkulasi biaya konkret dan keterjangkauan

Sebelum merumuskan strategi STP, dilakukan sintesis validitas variabel untuk memastikan bahwa rekomendasi yang dihasilkan hanya didasarkan pada variabel yang terbukti memiliki kekuatan diskriminatif antar segmen. Hasil sintesis menunjukkan bahwa variabel pembiayaan pendidikan dan status bekerja merupakan pembeda utama yang konsisten di seluruh kluster, dengan distribusi yang kontras dan stabil. Sebaliknya, variabel sosial ekonomi tidak menunjukkan konsistensi dalam membedakan kluster, distribusinya tersebar atau bimodal di semua segmen, sehingga tidak valid sebagai dasar pengambilan keputusan strategis. Variabel usia hanya valid sebagai

pembeda pada GMM-0 (74,1% berusia ≤ 20 tahun dengan std. deviasi $\pm 2,48$), namun tidak valid untuk GMM-1 dan GMM-2 yang memiliki sebaran usia luas. Variabel prestasi akademik dan non-akademik diposisikan sebagai variabel pendukung yang memperkaya profil segmen namun tidak menjadi dasar segmentasi utama.

Dengan demikian, seluruh elemen *targeting* dan *positioning* dalam kerangka STP berikut diturunkan hanya dari variabel yang telah tervalidasi secara empiris, yaitu pola pembiayaan dan status keterlibatan kerja. Tabel 3 menyajikan pemetaan dasar STP berdasarkan karakteristik utama masing-masing klaster mahasiswa. Klaster dengan dominasi pembiayaan orang tua dan kesiapan akademik tinggi diposisikan sebagai target utama untuk program unggulan dan skema beasiswa berbasis prestasi. Sementara itu, klaster mahasiswa pekerja dengan pembiayaan mandiri lebih sesuai ditargetkan melalui strategi promosi fleksibel, seperti program kelas malam atau pembelajaran campuran (*blended learning*).

Implementasi lebih lanjut dari pemetaan STP tersebut dirangkum dalam Tabel 9, yang menunjukkan bagaimana karakteristik klaster diterjemahkan menjadi rekomendasi strategi pemasaran yang spesifik.

Tabel 9. Penerapan Hasil Klaster ke STP

Klaster	Kanal Utama	Dasar Kanal	Pemilihan	Pesan Utama	Promosi	Penawaran Utama	Gaya Visual
GMM-0	TikTok, Instagram, Kunjungan SMA/SMK	Usia ≤ 20 (74,1%) pengguna dominan platform visual	\rightarrow thn	"Kampus juara, alumni berprestasi, kesempatan kerja luas."		Beasiswa unggulan, fast-track	Muda, energik, inspiratif, bergaya
GMM-1	LinkedIn, Email	Bekerja platform komunikasi profesional	98% jaringan dan	"Kuliah fleksibel, kerja tetap jalan."		Kelas malam, online, cicilan biaya, kelas fleksibel.	Profesional, transparansi, formal
GMM-2	Facebook, WA Group, Web	Usia tersebar luas dengan penetrasi lintas usia	\rightarrow platform	"Kuliah terjangkau, masa depan menjanjikan."		Beasiswa sosial, UKT ringan, potongan biaya.	Humoris, empatik, menggemakan.

Mengacu pada Kotler dan Keller (2016), *positioning* yang telah dirumuskan selanjutnya diterjemahkan ke dalam elemen promosi melalui bauran komunikasi pemasaran. Elemen promosi meliputi pemilihan saluran komunikasi, perancangan pesan utama, bentuk penawaran, serta gaya visual komunikasi. Setiap elemen tersebut dirancang agar selaras dengan karakteristik dan preferensi segmen sasaran. Pendekatan ini memastikan bahwa strategi pemasaran tidak lagi disusun berdasarkan asumsi normatif, melainkan diturunkan secara sistematis dari hasil segmentasi berdasarkan data.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa segmentasi mahasiswa berbasis data dapat menjadi pendekatan yang lebih objektif dan efektif dalam mendukung perumusan strategi pemasaran perguruan tinggi. Penerapan pendekatan *two-stage clustering* yang mengombinasikan HDBSCAN dan *Gaussian Mixture Model* (GMM) terbukti mampu meningkatkan kualitas segmentasi dibandingkan penerapan GMM secara langsung pada dataset utuh.

HDBSCAN berperan penting dalam mengeksplorasi struktur global data mahasiswa dan menyaring noise serta kelompok minoritas yang berpotensi mengaburkan pola utama populasi. Segmentasi lanjutan menggunakan GMM pada klaster mayoritas (651 dari 731 mahasiswa) menghasilkan tiga segmen mahasiswa yang lebih homogen dengan profil yang berbeda secara signifikan: (1) Mahasiswa Tradisional Berprestasi (GMM-0, n=170), usia muda terkonsentrasi ≤ 20 tahun, pembiayaan sepenuhnya eksternal, tidak bekerja, dan prestasi non-akademik tertinggi; (2) Mahasiswa Pekerja Fleksibel (GMM-1, n=62), seluruhnya bekerja dengan pembiayaan campuran secara eksklusif; dan (3) Mahasiswa Mandiri Adaptif (GMM-2, n=419), hampir seluruhnya membiayai kuliah secara mandiri sambil bekerja, dengan proporsi prestasi akademik tertinggi (78%).

Temuan penting dari penelitian ini adalah bahwa variabel pembiayaan pendidikan dan status keterlibatan kerja terbukti sebagai pembeda utama yang konsisten antar segmen, sementara variabel sosial ekonomi tidak valid sebagai dasar segmentasi karena distribusinya tersebar di semua klaster. Variabel usia hanya bersifat deskriptif dan valid sebagai pembeda hanya pada segmen GMM-0. Hasil segmentasi tersebut berhasil diterjemahkan ke dalam kerangka STP secara sistematis, di mana strategi pemasaran yang dirumuskan tidak lagi bersifat subjektif, melainkan diturunkan langsung dari karakteristik empiris masing-masing segmen mahasiswa.

Meskipun penelitian ini menghasilkan segmentasi mahasiswa yang informatif dan aplikatif, terdapat beberapa keterbatasan yang perlu diperhatikan. Data yang digunakan bersifat statis dan merepresentasikan kondisi pada periode tertentu, sehingga perubahan karakteristik mahasiswa dari waktu ke waktu belum sepenuhnya terakomodasi. Selain itu, penelitian ini belum mempertimbangkan faktor perilaku non-akademik secara lebih mendalam, seperti preferensi media promosi atau respons terhadap kampanye pemasaran. Rekomendasi strategi pemasaran yang dihasilkan juga belum melalui tahap validasi implementasi lapangan untuk mengukur efektivitas aktualnya, sehingga pengukuran perubahan

sebelum dan sesudah penerapan strategi perlu dilakukan pada penelitian lanjutan.

Penelitian selanjutnya disarankan untuk mengembangkan pendekatan segmentasi yang bersifat dinamis dengan memanfaatkan data longitudinal agar perubahan karakteristik mahasiswa dapat dianalisis secara temporal. Selain itu, integrasi variabel perilaku dan *digital engagement* berpotensi memperkaya profil segmen mahasiswa dan meningkatkan ketepatan strategi pemasaran. Pendekatan *two-stage clustering* juga dapat diuji pada konteks institusi pendidikan lain untuk menilai generalisasi dan konsistensi temuan penelitian ini.

DAFTAR PUSTAKA

- [1] Ayomi, D., Suharto, S., & Mazni, A. (2022). Pengaruh Bauran Pemasaran Jasa Pendidikan Terhadap Keputusan Mahasiswa Memilih Institut Agama Islam Negeri (Iain) Metro Lampung. *Simplex: Journal of Economic Management*. <https://doi.org/10.24127/simplex.v3i2.3204.J.E>.
- [2] Nafi'ah, A., & Ngadhimah, M. (2024). Strategi Pemasaran Jasa Pendidikan Dalam Meningkatkan Daya Saing Fakultas Ekonomi Dan Bisnis Islam Iain Ponorogo. *Muslim Heritage*. <https://doi.org/10.21154/muslimheritage.v9i1.8343>.
- [3] Budiyanto, F., Waluyo, S., Fitriyani, Z., Ridha, M., & Ashriana, A. (2025). Optimasi Strategi Operasional Dan Pemasaran Perguruan Tinggi Melalui Pemodelan Sistem Dinamis Dan Analisis Statistik Preferensi Mahasiswa Baru Di Wilayah Mojokerto. *Cendekia: Jurnal Ilmu Pengetahuan*. <https://doi.org/10.51878/cendekia.v5i3.5811>.
- [4] Akmaliah, Q., Kalsum, U., Ifendi, M., Artikel, I., Kunci—, K., Pemasaran, S., Tinggi, B., & Juliana, P. (2025). Strategi Pemasaran Jasa pada Perguruan Tinggi. *DIAJAR: Jurnal Pendidikan dan Pembelajaran*. <https://doi.org/10.54259/diajar.v4i1.3942>.
- [5] Stoenică, L. (2022). Development Of A Conceptual Model Of The Marketing Strategy In Higher Education. Case Study Romanian Higher Education System. *Inted2022 Proceedings*. <https://doi.org/10.21125/inted.2022.1480>.
- [6] Bakri, R., Sobirov, B., Astuti, N., Ahmar, A., & Singh, P. (2025). A New Framework for Dynamic Educational Marketing Segmentation in Student Recruitment: Optimizing Fuzzy C-Means with Metaheuristic Techniques. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*. <https://doi.org/10.29207/resti.v9i3.6515>.
- [7] Amanda, S., Aliano, K., Jahid, A., & Arifin, W. (2025). Analisis Komparatif Svm Dan K-Means Dalam Data Mining Untuk Promosi Perguruan Tinggi. *Jris: Jurnal Rekayasa Informasi Swadharma*. <https://doi.org/10.56486/jris.vol5no2.869>.
- [8] Ridwan, I. (2025). Transforming Customer Segmentation with Unsupervised Learning Models and Behavioral Data in Digital Commerce. *International Journal of Research Publication and Reviews*. <https://doi.org/10.55248/gengpi.6.0525.1652>.
- [9] Kuwar, K., Dhore, S., & Deshmukh, S. (2025). Big Data and its Impact on Consumer Segmentation Strategies. *International Journal Of Scientific Research In Engineering And Management*. <https://doi.org/10.55041/ijrsrem51523>.
- [10] Jodlbauer, H., Tripathi, S., Bachmann, N., & Brunner, M. (2024). Unlocking hidden market segments: A data-driven approach exemplified by the electric vehicle market. *Expert Syst. Appl.*, 254, 124331. <https://doi.org/10.1016/j.eswa.2024.124331>.
- [11] Hamed, O., & Olayinka, O. (2021). Data driven customer segmentation and personalization strategies in modern business intelligence frameworks. *World Journal of Advanced Research and Reviews*. <https://doi.org/10.30574/wjarr.2021.12.3.0658>.
- [12] Hindupur, J., D, N., H, H., & Sharma, S. (2025). Analyzing Clustering Algorithms for Non-Linear Data to Evaluate Robustness and Scalability. *Proceedings of the 3rd International Conference on Futuristic Technology*. <https://doi.org/10.5220/0013659600004664>.
- [13] Amhimmid, H., Alzahra, F., Aljehany, A., Mohamed, M., & Farag, K. (2025). Comparative Study of Four Methods in Hierarchical Cluster Analysis. *Libyan Journal of Medical and Applied Sciences*. <https://doi.org/10.64943/ljmas.v3i4.183>.
- [14] Stewart, G., & Al-Khassaweneh, M. (2022). An Implementation of the HDBSCAN* Clustering

Algorithm. Applied Sciences.
<https://doi.org/10.3390/app12052405>.

- [15] Peña-Asensio, E., & Ferrari, F. (2025). Meteoroid Stream Identification with HDBSCAN Unsupervised Clustering Algorithm. *The Astronomical Journal*, 170. <https://doi.org/10.3847/1538-3881/adc8c>.
- [16] Hunt, E., & Reffert, S. (2020). Improving the open cluster census. *Astronomy & Astrophysics*. <https://doi.org/10.1051/0004-6361/202039341>.
- [17] Scrucca, L., Saqr, M., L'opez-Pernas, S., & Murphy, K. (2023). An introduction and tutorial to model-based clustering in education via Gaussian mixture modelling.
- [18] Kasa, S., & Rajan, V. (2023). Avoiding inferior clusterings with misspecified Gaussian mixture models. *Scientific Reports*, 13. <https://doi.org/10.1038/s41598-023-44608-3>.
- [19] Kazakovtsev, V., & Markushin, E. (2025). Data segmentation through two-level clustering with greedy approach. *ITM Web of Conferences*. <https://doi.org/10.1051/itmconf/20257204007>.
- [20] Park, H., Shin, K., & Kim, J. (2025). Efficient Clustering Method for Graph Images Using Two-Stage Clustering Technique. *Electronics*. <https://doi.org/10.3390/electronics14061232>.
- [21] Giang, L., Lan, H., Van Manh, D., & Hoc, T. (2024). Applying a two-step cluster algorithm in traffic accident data analysis. *Transport and Communications Science Journal*. <https://doi.org/10.47869/tcsj.75.4.16>.
- [22] Akhter, M., & Mohanty, S. (2023). A fast $O(N \lg N)$ time hybrid clustering algorithm using the circumference proximity based merging technique for diversified datasets. *Eng. Appl. Artif. Intell.*, 125, 106737. <https://doi.org/10.1016/j.engappai.2023.106737>.
- [23] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- [24] McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons.
- [25] Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons. doi.org.