

ANALISIS FAKTOR MAHASISWA MEMILIH UNIVERSITAS DENGAN K-MEANS DAN DIMENSI PSIKOGRAFIS

Abdul Rahman Alfaris Baraja¹, Erza Sofian²

^{1,2}Program Studi Sistem Informasi, Universitas Ary Ginanjar
e-mail : a.rahman.al.b@students.esqbs.ac.id

ABSTRACT

This research employs a quantitative approach with data mining techniques. Primary data was collected through questionnaires distributed to active students of UAG from the 2021–2024 cohorts. The dimensions used include campus quality, economic factors, and the influence of promotion. After preprocessing, the data was analyzed to form student clusters based on their preferences. The study identified three clusters of students in choosing Ary Ginanjar University (UAG). Cluster 1 (29%) exhibited highly positive perceptions toward campus quality, economic factors, and promotion, with advertising and promotional effects being the most dominant factor. Cluster 2 (18%) showed low agreement across all factors, particularly in economic and financial aspects, which may hinder their decision to choose UAG. Cluster 3 (53%) represented students with neutral perceptions, where campus quality and financial affordability were the main influencing factors. These findings provide a basis for the university to design more targeted marketing strategies, such as enhancing promotional campaigns for Cluster 1, offering scholarship information for Cluster 2, and highlighting the balance between quality and affordability for Cluster 3. This study recommends that UAG focus more on economically driven marketing strategies. Psychographic segmentation can help the university develop more personalized and effective approaches in facing market competition.

Keywords: market segmentation, k-means, psychographics, data mining

ABSTRAK

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik penambangan data. Data primer dikumpulkan melalui kuesioner yang dibagikan kepada mahasiswa aktif UAG dari angkatan 2021–2024. Dimensi yang digunakan meliputi kualitas kampus, faktor ekonomi, dan pengaruh promosi. Setelah pra-pemrosesan, data dianalisis untuk membentuk klaster mahasiswa berdasarkan preferensi mereka. Studi ini mengidentifikasi tiga klaster mahasiswa dalam memilih Universitas Ary Ginanjar (UAG). Klaster 1 (29%) menunjukkan persepsi yang sangat positif terhadap kualitas kampus, faktor ekonomi, dan promosi, dengan pengaruh iklan dan promosi sebagai faktor yang paling dominan. Klaster 2 (18%) menunjukkan kesepakatan yang rendah di semua faktor, khususnya dalam aspek ekonomi dan keuangan, yang dapat menghambat keputusan mereka untuk memilih UAG. Klaster 3 (53%) mewakili mahasiswa dengan persepsi netral, di mana kualitas kampus dan keterjangkauan finansial merupakan faktor utama yang memengaruhi. Temuan ini memberikan dasar bagi universitas untuk merancang strategi pemasaran yang lebih terarah, seperti meningkatkan kampanye promosi untuk Klaster 1, menawarkan informasi beasiswa untuk Klaster 2, dan menyoroti keseimbangan antara kualitas dan keterjangkauan untuk Klaster 3. Studi ini merekomendasikan agar UAG lebih fokus pada strategi pemasaran yang didorong oleh faktor ekonomi. Segmentasi psikografis dapat membantu universitas mengembangkan pendekatan yang lebih personal dan efektif dalam menghadapi persaingan pasar.

kata kunci : segmentasi pasar, k-means, psikografi, penambangan data

PENDAHULUAN

Arpay (2023) menyatakan bahwa persaingan di sektor perguruan tinggi memaksa lembaga perguruan tinggi untuk mengembangkan strategi pemasaran yang lebih inovatif untuk menghadapi persaingan dan tantangan dari sekolah lokal dan sekolah internasional di seluruh dunia. Hart dan

Rodgers (2023) juga menyatakan bahwa perguruan tinggi telah mengalami meningkatnya persaingan, yang mengarah pada minat ilmiah yang signifikan dalam konsep persaingan, daya saing, dan keunggulan kompetitif dalam lembaga perguruan tinggi.

Dalam persaingan global, perguruan tinggi tidak hanya harus bersaing di dalam negeri tetapi juga menghadapi tantangan dari organisasi internasional. Seperti yang dikemukakan Arpay (2023) dan Hart dan Rodgers (2023), meningkatnya persaingan mendorong lembaga pendidikan untuk mengadopsi strategi pemasaran yang inovatif. Selain itu, transformasi digital yang dipercepat oleh pandemi COVID-19 beberapa tahun yang lalu mengubah cara penyampaian pendidikan dan memaksa universitas beradaptasi dengan teknologi terkini untuk menarik dan mempertahankan mahasiswa. Seperti yang dinyatakan Chukwuere (2024) bahwa pandemi COVID-19 (new normal) mendisrupsi proses pendidikan dan memperjuangkan proses digitalisasi dan cara baru untuk menjalankan fungsi akademik pengajaran dan pembelajaran.

Menurut Ayuningtyas (2024), berdasarkan laporan Statista per Juli 2023, ada sekitar 3277 perguruan tinggi di Indonesia. Dengan jumlah perguruan tinggi yang sebanyak ini, Indonesia menduduki peringkat kedua sebagai negara dengan jumlah perguruan tinggi terbanyak di dunia. Jumlah mahasiswa baru di Indonesia juga terus meningkat dari tahun ke tahun. Menurut data yang diperoleh dari Badan Pusat Statistik (BPS), jumlah mahasiswa di Indonesia mencapai 9,32 juta orang pada tahun 2022, meningkat sebesar 4,02% dari tahun sebelumnya (Rizaty, 2023). Peningkatan ini terjadi meskipun terdapat berbagai tantangan, termasuk dampak pandemi COVID-19 yang mengubah cara pembelajaran menjadi daring (Putri, 2023). Selain itu, data dari Pangkalan Data Pendidikan Tinggi (PDDIKTI) juga mencatat peningkatan signifikan jumlah mahasiswa baru pada tahun ajaran 2019/2020 sebesar 20,13% dari tahun sebelumnya (Jayani, 2021). Peningkatan jumlah mahasiswa ini mencerminkan semangat dan keinginan masyarakat Indonesia untuk meningkatkan kualitas pendidikan dan sumber daya manusia. Menurut McDonald dan Dunbar (2013), Segmentasi pasar adalah proses pemisahan pelanggan, atau calon pelanggan, di pasar menjadi beberapa

kelompok atau segmen yang berbeda. Pentingnya segmentasi pasar juga disebutkan oleh Kara dan Kaynak (1997) di mana dinyatakan bahwa segmentasi pasar memberikan pedoman yang efektif bagi pengembangan strategi pemasaran dan alokasi sumber daya perusahaan di antara pasar produk mereka yang beragam. Menyesuaikan dengan konteks penelitian ini, yang akan mengisi peran pasar dan akan disegmentasikan nantinya adalah mahasiswa. Segmentasi menggunakan kuesioner yang ditujukan kepada mahasiswa sebagai teknik pengumpulan data dapat membantu kampus memahami kebutuhan dan preferensi mahasiswa, sehingga dapat mengembangkan strategi pemasaran yang lebih efektif. Goodrich et al. (2020) dalam penelitiannya menggunakan sekumpulan pertanyaan yang merepresentasikan kepuasan terhadap kehidupan akademis, perguruan tinggi, dan pribadi, serta berbagai karakteristik siswa yang diinginkan di sekolah selama tahap perencanaan perguruan tinggi mereka sebagai variabel psikografis untuk segmentasi pasar. Sementara itu, Khusnuliawati dan Putri (2021) dalam penelitiannya menggunakan beberapa variabel psikografis seperti kualitas kampus, ekonomi, dan efek iklan dan promosi, dan beberapa lainnya yang mendukung untuk meneliti segmentasi pasar mahasiswa Universitas Sahid Surakarta. Salah satu kampus yang mengalami persaingan ini adalah Universitas Ary Ginanjar atau biasa disingkat menjadi UAG. UAG yang tadinya merupakan sebuah sekolah tinggi mengganti statusnya dan membuka beberapa program studi baru. Terlihat di beberapa media sosial seperti Instagram dan website resminya, UAG masih aktif dalam mempromosikan penerimaan mahasiswa baru tahun ajaran 2025/2026.

Pernyataan bahwa UAG merupakan salah satu kampus yang mengalami persaingan antar perguruan tinggi ini juga dibuktikan dengan hasil wawancara penulis dengan salah satu tim Penerimaan Mahasiswa

Baru atau biasa disingkat dengan PMB UAG, Ali Zainal Abidin, yang dilaksanakan pada

Kamis, 28 November 2024. Beliau membenarkan dan menyetujui saat ditanyakan tentang isu kompetisi di antara perguruan tinggi ini. Beliau juga menyatakan bahwa ada sekitar beberapa saluran pemasaran yang dipakai oleh tim PMB UAG. Saluran ini antara lain: 1. Digital marketing, Instagram, Tiktok, dan Search Engine Optimization di internet browser. 2. Bekerja sama dengan sekolah-sekolah menengah jenjang akhir ataupun pertama yang masuk ke kriteria pemasaran UAG dan membuat program kemitraan sekolah atau biasa disebut *partnership school*, seperti *talkshow event*, *workshop*. 3. Metode *student get student*, di mana calon mahasiswa, mahasiswa aktif, ataupun alumni didorong dan dianjurkan untuk merujuk teman-temannya di luar kampus agar bisa berkuliah di UAG. 4. Adanya *shout-out* atau penyebutan dan pemutaran video marketing UAG pada acara *ESQ Training* yang rutin dilaksanakan. 5. Melalui event milik UAG, seperti *EduFair* yang merupakan program *visit UAG* ke sekolah-sekolah, *trial class* dan *Sehari Jadi Mahasiswa*, yang merupakan program simulasi perkuliahan yang ditujukan kepada calon mahasiswa baru UAG. 6. Bekerja sama dengan *influencer* untuk memberikan *awareness* terhadap UAG.

Maka dari itu, permasalahan ini dicoba untuk diteliti. Metode penelitian permasalahan ini akan menggunakan salah satu disiplin ilmu data sains, yakni penambangan data atau yang lebih dikenal dengan istilah *data mining*. Teknik atau metode penambangan data yang akan digunakan adalah teknik *clustering*, dan algoritma *clustering* yang dipakai adalah *K-Means Clustering*. Sementara itu, dimensi dari variabel berkuliah di UAG yang dipakai dalam penelitian ini adalah kualitas kampus, ekonomi, dan efek iklan dan promosi.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif untuk menganalisis data secara statistik dan mengidentifikasi pola yang relevan. Penelitian ini juga akan dilaksanakan dengan menggunakan pendekatan standar proses penambangan data

CRISP-DM. Seperti yang dijelaskan pada bab sebelumnya. CRISP-DM dimulai dari tahap *business understanding* dan diakhiri dengan tahap *deployment*. Berikut adalah pengaplikasian tahap-tahap CRISP-DM ke dalam penelitian ini: 1. *Business Understanding*: *literature review*, penentuan variabel dan dimensi, validasi masalah. 2. *Data Understanding*: Perencanaan pembuatan data kuesioner, pembuatan data kuesioner, pengumpulan data dengan kuesioner. 3. *Data Preparation*: Transformasi data, normalisasi data, transformasi data menjadi tabel olahan hasil kuesioner (akan dibahas pada subbab teknik analisis data). 4. *Modeling*: Kalkulasi dengan algoritma *K-Means Clustering*, penentuan *centroid* dan iterasinya (akan dibahas pada subbab teknik analisis data). 5. *Evaluation*: Penentuan model dengan menggunakan *Elbow Method* dan *DaviesBouldin Index* untuk menentukan kualitas *clustering* (akan dibahas pada subbab teknik analisis data). 6. *Deployment*: Pembahasan model, pengambilan *insight* terhadap model (akan dibahas pada subbab teknik analisis data). Objek pada penelitian ini adalah seluruh mahasiswa UAG angkatan 2021, 2022, 2023, dan 2024 yang berstatus aktif, tidak dalam status cuti, belum lulus, dan bukan kelas karyawan. Objek ini menjadi populasi penelitiannya. Sementara dari populasi, akan dilakukan *voluntary sampling* sebagai teknik pengambilan sampelnya. Menurut Murairwa (2015), *voluntary sampling* merupakan teknik pengambilan sampel non-probabilitas yang baru. Seperti namanya, sampel akhir dipilih dari responden potensial yang bersedia dan memenuhi syarat untuk berpartisipasi dalam survei. Jumlah sampelnya akan merujuk kepada nilai sampel dari jumlah populasi yang dibulatkan satu taraf di atas nilai populasi penelitian ini dengan menggunakan tabel sampel Isaac-Michael dengan persentase kesalahan 5%. Metode Isaac-Michael dipakai karena tidak memerlukan kalkulasi dalam menentukan jumlah sampel, seperti pada metode lainnya. Teknik pengambilan sampel menggunakan metode Isaac-Michael ini juga didukung dengan pernyataan Sukwika (2023) bahwa

metode Isaac-Michael bisa digunakan jika besaran atau jumlah populasi dapat diketahui. Menurut Sistem Akademik UAG yang diakses pada 26 Maret 2025, terdapat 353 mahasiswa angkatan 2021, 2022, 2023, dan 2024 dengan kriteria yang telah dijelaskan sebelumnya. Maka dari itu, jika dibulatkan ke angka yang lebih besar paling dekat pada tabel Isaac-Michael yang mana berjumlah 360, jumlah sampel yang perlu diperoleh dengan persenan kesalahan 5% adalah 177.

2.4. Objek dan Dimensi dari Variabel Penelitian. Objek penelitian ini adalah data kuesioner mahasiswa aktif UAG angkatan 2021, 2022, 2023, dan 2024 yang kriterianya telah dibahas pada subbab sebelumnya. Sementara itu, dimensi berdasarkan variabel berkuliah di UAG memiliki deskripsi sebagai berikut:

Tabel 1. Keterangan Dimensi Penelitian

Kode	Dimensi	Keterangan
KK	Kualitas Kampus	Kualitas kampus secara keseluruhan, prestasi yang dimiliki, reputasi kampus secara keseluruhan
EF	Ekonomi dan Finansial	Biaya kuliah yang terjangkau, sesuai dengan <i>budget</i> , tersedianya program beasiswa
IP	Efek Iklan dan Promosi	Hasil dari promosi kampus, rekomendasi dari teman atau keluarga

Dimensi ini dipilih berdasarkan teori segmentasi psikografis. Hal ini didasari oleh Lin (2002) yang menyatakan bahwa di dalam segmentasi psikografis, konsumen dibagi menjadi beberapa kelompok dengan gaya hidup dan kepribadian sebagai dasar penentuannya. Dijelaskan juga bahwa untuk segmentasi pasar perguruan tinggi swasta, faktor yang dipilih didasari oleh keterukuran, ketersediaan, keandalan, dan kemampuannya. Dalam konteks penelitian yang serupa, Khusnuliawati dan Putri (2021) juga menggunakan variabel-variabel seperti

faktor di atas untuk melakukan segmentasi pasar pada objek penelitiannya.

2.5 Teknik Analisis Data. Pada penelitian ini, akan dilakukan analisis clustering terhadap data kuesioner mahasiswa menggunakan algoritma K-Means Clustering. Proses kalkulasi pertama akan dilakukan secara manual dengan bantuan Microsoft Excel. Selanjutnya, akan ada pembuktian dengan Script Python. Proses kalkulasi dengan dua metode ini bertujuan untuk menjawab rumusan masalah yang juga menanyakan terkait bagaimana penerapan K-Means Clustering. Tahapan lengkapnya sendiri adalah sebagai berikut:

2.5.1 Persiapan dan Pengisian Kuesioner Pertama-tama, dibuat kuesioner yang akan disebarkan kepada mahasiswa. Kuesioner ini berisikan 30 pernyataan yang setiap faktor memiliki masing-masing 10 pernyataan. Faktor ini antara lain: 1. Kualitas kampus (F1) 2. Ekonomi dan Finansial (F2) 3. Efek Iklan dan Promosi (F3) Seperti yang sudah dibahas pada subbab sebelumnya, target responden kuesioner ini berjumlah 177 responden. Apabila sudah mencapai atau melebihi target responden, datanya sudah bisa masuk ke tahap pengolahan.

2.5.2 Normalisasi Data. Data dinormalisasi menggunakan Z-Score untuk meningkatkan kualitas clustering. Seperti yang dinyatakan oleh Mohamad dan Usman (2013), bahwa normalisasi data menstandarisasi data mentah dengan mengubahnya menjadi rentang tertentu menggunakan transformasi linier yang dapat menghasilkan klaster berkualitas baik dan meningkatkan akurasi algoritma clustering. Hal ini diterapkan saat proses clustering dengan Excel dan Python, agar setiap nilai dari kuesioner memiliki rata-rata 0 dan standar deviasi 1.

2.5.3 Proses K-Means Clustering dengan Excel dan Python. Proses clustering pertama dilakukan secara manual menggunakan Microsoft Excel untuk nilai K = 2, 3, dan 4. Langkah-langkah yang dilakukan mencakup: 1. Menentukan centroid awal dari data 2. Menghitung jarak

Euclidean setiap data ke masing-masing centroid 3. Menetapkan cluster berdasarkan jarak terdekat 4. Menghitung ulang centroid berdasarkan rata-rata anggota cluster 5. Mengulangi proses hingga tidak ada perubahan pengelompokan (konvergen) Kemudian setelah kalkulasi manual akan dijelaskan proses dan hasil dari kalkulasi menggunakan Python sebagai pembuktian.

2.5.4 Kalkulasi SSE untuk Tiap K dan Pembuatan Elbow Graph .Setelah hasil clustering konvergen, dilakukan penghitungan nilai SSE (Sum of Squared Errors) untuk masing-masing nilai K. SSE merupakan total kuadrat jarak antara setiap data dengan centroid-nya dan digunakan untuk mengukur seberapa kompak kluster yang terbentuk.

2.5.5 Menghitung DBI untuk K=2, 3, dan 4 Sebagai metrik evaluasi tambahan, dilakukan penghitungan nilai DaviesBouldin Index (DBI) secara manual untuk masing-masing nilai K. DBI digunakan untuk mengukur rasio kekompakan dan keterpisahan antar kluster. Semakin kecil nilai DBI, semakin baik kualitas clustering.

2.5.6 Clustering Otomatis dan Visualisasi dengan Python .Seluruh proses K-means juga dilakukan secara otomatis menggunakan bahasa pemrograman Python. Proses ini mencakup: 1. Normalisasi data 2. Clustering untuk K = 1 hingga K = 10 3. Perhitungan SSE dan DBI secara otomatis 4. Visualisasi hasil clustering dan elbow graph Tujuan dari langkah ini adalah untuk memvalidasi hasil manual serta memberikan gambaran lebih lengkap terhadap struktur data.

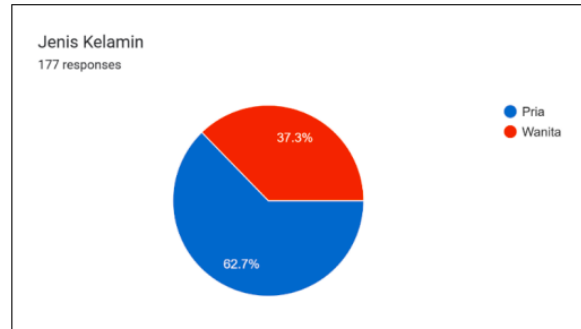
2.5.7 Penentuan Model Clustering Terbaik Model clustering terbaik ditentukan berdasarkan dua kriteria utama: 1. Nilai DBI terkecil (cluster paling terpisah dan kompak) 2. Posisi siku pada elbow graph (efisiensi jumlah kluster). Model terbaik ini digunakan sebagai acuan untuk tahap analisis clustering.

1. HASIL DAN PEMBAHASAN

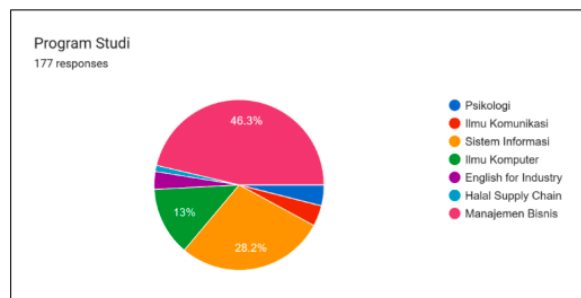
3.1 Pengumpulan data

Penyebaran kuesioner dimulai pada 1 Maret 2025, dan selesai pada 6 Juli 2025 karena sudah memenuhi target responden. Proses ini

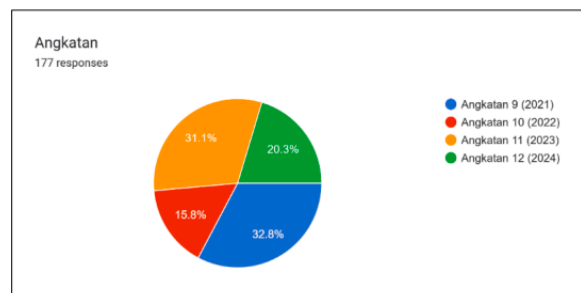
didasari tahap business understanding dan data understanding pada CRISP-DM, di mana variabel dan dimensi ditentukan saat perancangan penelitian dan kuesioner dibuat berdasarkan hasil yang sudah ditentukan. Demografi respondennya adalah sebagai berikut:



Gambar 1 Demografi Jenis Kelamin Responden



Gambar 2 Demografi Prodi Responden



Gambar 3 Demografi Angkatan Responden

Tabel 2 Demografi Jenis Kelamin Responden

Jenis Kelamin	Jumlah	Persen (%)
Pria	111	62.7%
Wanita	66	37.3%

Tabel 3 Demografi Program Studi Responden

Program Studi	Jumlah	Persen (%)
Psikologi	7	4%
Ilmu Komunikasi	7	4%
Sistem Informasi	50	28.2%
Ilmu Komputer	23	13%
English for Industry	6	3.4%
Halal Supply Chain	2	1.1%
Manajemen Bisnis	82	46.3%

Tabel 4 Demografi Angkatan Responden

Angkatan	Jumlah	Persen (%)
Angkatan 9 (2021)	58	32.8%
Angkatan 10 (2022)	28	15.8%
Angkatan 11 (2023)	55	31.1%
Angkatan 12 (2024)	36	20.3%

Dari tahap ini, diperoleh hasil kuesioner berupa tabel responden yang nantinya akan diolah pada tahap berikutnya.

3.2 Proses Data Preprocessing .Dari proses sebelumnya, diperoleh tabel responden yang sudah siap untuk diolah. Proses ini didasari tahap data preparation pada CRISP-DM, di mana data ditransformasi dan dinormalisasi agar bisa dikalkulasi. Proses pengolahannya sendiri adalah sebagai berikut:

1. Penghapusan kolom yang tidak diperlukan (timestamp, email address, jenis kelamin, program studi, angkatan)
2. Penambahan kolom ID di paling kiri. Nilai dari kolom ini merupakan nomor ID dengan data pertama memiliki ID 1, data kedua memiliki ID 2, dan seterusnya.
3. Penambahan kolom AVG KK, AVG EF, dan AVG IP setelah kolom pernyataan terakhir untuk setiap dimensi. Kolom ini berisikan rata-rata setiap dimensi. Untuk input nilai rata-ratanya, digunakan fungsi =AVERAGE() pada Excel. Fungsi ini diterapkan pada setiap kolom di atas dengan jarak fungsi menyesuaikan kolom nilai pernyataan setiap dimensi. Tahap ini didasari rumus untuk menghitung rata-rata dari nilai setiap pernyataan pada suatu dimensi. Rumus tersebut antara lain :

$$AVG_{X_n}^{(i)} = \frac{1}{p_n} \sum_{j=1}^{p_n} X_{ij} \quad (1)$$

Rumus Kalkulasi Rata-rata Nilai Dimensi
Dengan keterangan sebagai berikut:
 $AVGX_n(i)$ = rata-rata nilai dimensi n untuk responden ke-i
 x_{ij} = nilai responden ke-i pada pernyataan ke-j dari dimensi n
 p_n = jumlah pernyataan dalam dimensi n

4. Pembuatan lembar/sheet baru yang hanya berisikan kolom ID, AVG KK, AVG EF, dan AVG IP. Contoh tabelnya sebagai berikut:

Tabel 5 Contoh Tabel dengan Nilai Rata-rata Setiap Dimensi

ID	AVG KK	AVG EF	AVG IP
1	2.9	2	3.1
2	3.3	3.5	3.6
3	3	3.2	2.9
4	2.6	3.2	2.8
5	2.6	3.4	2.4
...	...	...	...
177	3.9	4	4

5. Kalkulasi rata-rata dari keseluruhan masing-masing nilai AVG KK, AVG EF, AVG IP, dan standar deviasi data dengan fungsi =AVERAGE() dan =STDEV.S() untuk mengetahui seberapa besar variasi atau konsistensi jawaban tiap responden terhadap pernyataan-pernyataan dalam satu dimensi. Fungsinya diterapkan pada nilai kolom AVG KK, AVG EF, dan AVG IP. Tahapan ini berisikan penghitungan rata-rata dimensi yang didasari rumus:

$$\mu_{X_n} = \frac{1}{N} \sum_{N=1}^N AVG_{X_n}^{(i)} \quad (2)$$

Rumus kalkulasi nilai rata rata keseluruhan dimensi
dengan keterangan sebagai berikut:
 μ_{X_n} = nilai rata-rata keseluruhan dimensi n
 $AVGX_n(i)$ = rata-rata nilai dimensi n untuk responden ke-i
Tahapan ini juga berisikan penghitungan nilai standar deviasi dimensi yang didasari rumus:

$$STD_{X_n}^{(i)} = \sqrt{\frac{1}{p_n} \sum_{j=1}^{p_n} (X_{ij} - x_i)^2} \quad (3)$$

Rumus Kalkulasi Standar Deviasi
 X_{ij} = Rata rata nilai X_{ij} dari semua pernyataan pada dimensi n untuk responden ke – 1
 $STD_{X_n}^{(i)}$ = Standar Deviasi dari dimensi n untuk reponden ke – i

Dari tahap ini, hasil kalkulasi rata-rata dan standar deviasinya adalah sebagai berikut:

Tabel 6 Hasil Kalkulasi Rata-rata dan Standar Deviasi Data

	KK	EF	IP
Mean	2.992655367	2.970621469	2.981355932
StdDev	0.593048998	0.621819157	0.605319575

6. Penambahan kolom Z_KK, Z_EF, dan Z_IP yang berisikan nilai yang dinormalisasi dengan metode Z-Score dari kolom AVG KK sampai AVG IP. Fungsi yang dipakai untuk mengkalkulasi nilainya didasari dari rumus normalisasi Z-Score.

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (4)$$

Rumus kalkulasi z- Score

Dengan keterangan sebagai berikut:

Z_{ij} = nilai hasil normalisasi (Z-Score) untuk data ke-i pada dimensi j

X_{ij} = nilai asli data ke-i pada dimensi j

μ_j = rata-rata semua nilai pada dimensi j

σ_j = standar deviasi semua nilai pada dimensi j
Dari tahap ini, contoh hasil tabelnya adalah sebagai berikut:

Tabel 7 Contoh Tabel dengan Kolom Z-Score untuk Setiap Dimensi.

ID	AVG KK	AVG EF	AVG IP	Z_KK	Z_EF	Z_IP
1	2.9	2	3.1	-0.1562	-1.673	0.1810
2	3.3	3.5	3.6	0.5182	0.8554	1.0241
3	3	3.2	2.9	0.0123	0.3496	-0.1564
4	2.6	3.2	2.8	-0.6621	0.3496	-0.3248
5	2.6	3.4	2.4	-0.6621	0.6868	-0.9993
...	...	...	...	...	...	...
177	3.9	4	4	1.5299	1.6985	1.6985

Perlu diketahui bahwa dalam tahap data preprocessing ini, tidak ada kegiatan menghapus missing values. Hal ini dikarenakan formulir kuesioner mewajibkan agar setiap kolom input diisi oleh responden. Tabel yang telah ditambahkan kolom Z-Score untuk masing-masing dimensi di atas merupakan tabel yang sudah siap untuk dikalkulasi dengan K-means.

3.3 Proses Clustering Dari proses data preprocessing, diperoleh tabel yang siap untuk dikalkulasi dengan K-means. Proses clustering ini didasari tahap modeling pada CRISP-DM, di mana data yang sudah bersih dikalkulasi dengan algoritma K-means sampai ditemukan beberapa model untuk

dievaluasi nantinya. Secara garis besar, tahapan pada proses ini akan mengacu pada Han et al. (2011) yang menyatakan bahwa rincian metode kalkulasi K-means adalah sebagai berikut: 1. Menentukan objek K secara bebas dari data sebagai centroid awal. 2. Menetapkan setiap objek ke kluster yang paling mirip dengan objek tersebut berdasarkan nilai rata-rata objek dalam kluster. 3. Memperbarui nilai rata-rata kluster dengan mengkalkulasi nilai rata-rata setiap objek dalam setiap kluster yang dimasuki. 4. Mengulang langkah 2 dan 3 sampai nilai rata-rata kluster tidak berubah. Kalkulasinya sendiri akan menggunakan 2 pendekatan, dengan Microsoft Excel dan Python.

3.3.1 Clustering Manual dengan Excel

Pertama, akan dilakukan kalkulasi manual dengan Excel dan dicontohkan untuk kalkulasi K=2 terlebih dahulu. Tahapan kalkulasinya sendiri adalah sebagai berikut: 1. Penentuan centroid atau nilai pusat awal dari data. Untuk data yang akan dijadikan centroid awalnya bisa dipilih secara acak dari data sumber. Misalkan pada contoh ini, diambil data dengan ID 3 dan 75. Kemudian, diambil nilai atribut Z_KK sampai Z_IP-nya menggunakan rumus =VLOOKUP(). Penentuan centroid awal dan nilainya dilakukan pada sheet baru. Berikut tabel centroid awalnya

Tabel 8 Centroid awal

Cluster	Data ke- (ID)	Z_KK	Z_EF	Z_IP
1	3	0.012385	0.349625	-0.15624
2	75	-0.6621	0.349625	-0.99934

2. Penghitungan jarak Euclidean setiap data ke masing-masing centroid. Dibuat sheet baru untuk memulai iterasi kalkulasi K-means. Struktur penempatan data pada sheet ini adalah seperti pada gambar di bawah

Cluster	Data ke-	KK	EF	IP	Kesimpulan Iterasi	Cluster	Anggota
1	3	0.012385	0.349625	-0.15624		1	126
2	75	-0.6621	0.349625	-0.99934		2	51

Dataset		Jarak ke Centroid		1	2	Terdekat	Cluster yang diikuti
Data ke-	KK	EF	IP				
1	-0.15624	-1.67382	0.181005	2.058271	2.396542	2.058271	1
2	0.518245	0.855485	1.024105	1.380215	2.396542	1.380215	1
3	0.012385	0.349625	-0.15624	0	1.079696	0	1
4	-0.6621	0.349625	-0.32486	0.695239	0.674481	0.674481	2
5	-0.6621	0.686865	-0.99934	1.131138	0.33724	0.33724	2
6	0.686865	0.855485	0.012385	0.859797	1.760446	0.859797	1
7	-2.01106	-2.3483	-2.51692	4.116542	3.376616	3.376616	2
8	-0.49348	1.361346	-1.5052	1.760446	1.143637	1.143637	2
9	-0.99934	-0.6621	0.012385	1.440691	1.469996	1.440691	1

Gambar 4.4 Contoh Struktur Tabel Iterasi Kalkulasi K-means.

Penjelasan dari struktur penempatan data ini antara lain: b. Ada dua tabel pada sheet ini. Tabel 1 merupakan tabel jumlah anggota per-kelompok (posisinya di sel A2:I4), dan tabel kedua yang merupakan tabel kalkulasi jarak Euclidean (posisinya di sel B6:I184). c. Pada tabel jumlah anggota per-kelompok, dari kolom "Data ke-" sampai "IP" diambil dari sheet centroid awal menggunakan rumus =VLOOKUP(). Kemudian ada kolom "Kesimpulan Iterasi" (disingkat K.I.) yang diisi sebagai penanda ketika hasil iterasi sudah konvergen. Kemudian ada kolom "Cluster" sebagai indeks kelompok untuk kolom selanjutnya, dan terakhir ada kolom "Anggota" yang berisikan jumlah anggota dari tiap kelompok. Nilai pada kolom "Anggota" diisi dengan fungsi =COUNTIF() yang menghitung jumlah anggota tiap kelompok berdasarkan hasil pengelompokan pada tabel kalkulasi jarak Euclidean. d. Pada tabel kalkulasi jarak Euclidean, kolom "Data ke-" sampai "IP" diisi dengan nilai dari data sumber. Setelahnya, ada kolom Jarak ke Centroid 1 dan 2 yang nilainya merupakan hasil kalkulasi jarak Euclidean. Rumusnya sendiri adalah sebagai berikut:

$$d = \sqrt{\sum_{i=1}^n (x_i - c)^2} \quad (6)$$

Rumus 4.5 Kalkulasi Jarak Euclidean
Dengan keterangan sebagai berikut:
 d = Nilai jarak Euclidean
 x_i = Nilai dimensi i
 c_i = Nilai centroid pada dimensi i

dikarenakan model yang sedang dikalkulasi adalah model K=2, rumusnya akan menjadi seperti ini:

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2} \quad (5)$$

Tabel 9 Jumlah Cluster Per-anggota Iterasi ke-2

Cluster	Data ke-	KK	EF	IP	K.I.	Cluster	Anggota
1	126	0.4259	0.2853	0.4339		1	111
2	51	-1.0522	-0.8340	-1.1382		2	66

Tabel 10 hasil clustering 2

Cluster	Data ke-	KK	EF	IP	K.I.	Cluster	Anggota
1	93	0.674173	0.64335	0.641537		1	93
2	84	-0.74641	-0.79057	-0.75042	Konv.	2	84

Tabel 11 Hasil clustering 3

Cluster	Data ke-	KK	EF	IP	K.I.	Cluster	Anggota
1	52	1.0370	1.0597	1.12787		1	52
2	31	-1.4072	-1.6575	-1.0646		2	31
3	94	-0.1096	-0.1096	-0.3087	Konv.	3	94

Tabel 12 Hasil Clustering 4

Cluster	Data ke-	KK	EF	IP	K.I.	Cluster	Anggota
1	39	-0.4632	-0.9431	-0.0524		1	39
2	68	-0.0099	0.1909	-0.4389		2	68
3	20	-1.6906	-1.8340	-1.4883		3	20
4	50	1.0510	1.0780	1.1657	Konv.	4	50

3.3.2 Kalkulasi dengan Python untuk Pembuktian. Perlu diketahui bahwa ada beberapa library yang digunakan dalam menulis script ini, seperti pandas, numpy, sklearn.cluster, sklearn.preprocessing, dan plotly. Hasil dari script yang dibuat akan menunjukkan jumlah anggota pada setiap cluster di terminal text editor, dan grafik plot chart di browser. Hasilnya sendiri adalah sebagai berikut:

```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
PS G:\My Drive\RISET SI (Abdul Rahman Alfari Bar...
rossoft/WindowsApps/python3.13.exe "g:/My Drive/RI...
eans Plot Chart K=2.py"
Cluster 0: 93 data point
Cluster 1: 84 data point
PS G:\My Drive\RISET SI (Abdul Rahman Alfari Bar...

```

Gbr Hasil Clustering K=2 pada Terminal Text Editor.

```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
PS G:\My Drive\RISET SI (Abdul Rahman Alfari Bar...
rossoft/WindowsApps/python3.13.exe "g:/My Drive/RI...
eans Plot Chart K=3.py"
Cluster 0: 94 data point
Cluster 1: 31 data point
Cluster 2: 52 data point

```

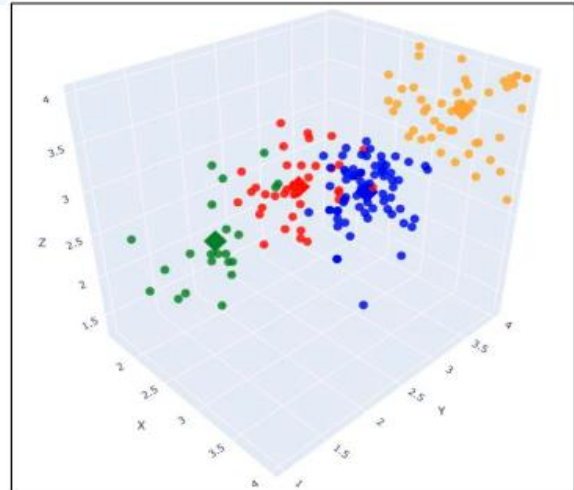
Gambar Hasil Clustering K=3 pada Terminal Text Editor.

```

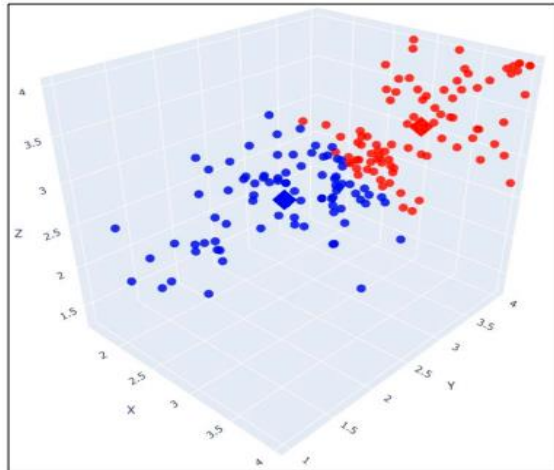
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS G:\My Drive\RISET SI (Abdul Rahman Alfari Bara
rosoft/WindowsApps/python3.13.exe "g:/My Drive/RI
eans Plot Chart K=4.py"
Cluster 0: 37 data point
Cluster 1: 68 data point
Cluster 2: 22 data point
Cluster 3: 50 data point
    
```

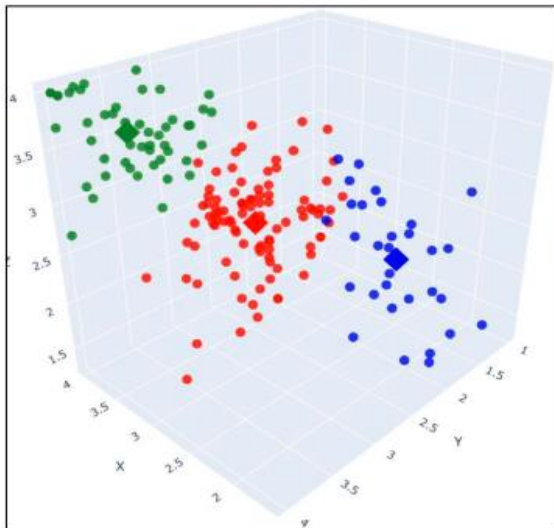
Gambar Hasil Clustering K=4 pada Terminal Text Editor



Gambar 4.10 Hasil Plot Chart Clustering K=4



Gambar Hasil Clustering K=2



Gambar 4.9 Hasil Plot Chart Clustering K=3

Ada beberapa hal yang perlu diperhatikan dari hasil clustering dan visualisasi menggunakan Python. Hal pertama adalah bahwa adanya perbedaan jumlah anggota klaster pada model K=4 pada hasil kalkulasi Excel dengan Python. Lebih jelasnya ada pada tabel di bawah:

Tabel 13 Perbedaan Hasil Clustering Kalkulasi Excel dan Python

Cluster	Jumlah anggota dalam cluster	
	Kalkulasi Excel	Kalkulasi Python
1	39	37
2	68	68
3	20	22
4	50	50

3.3.3 Elbow Method Manual dengan Excel
Dilakukan proses pembuatan elbow graph di Excel untuk menentukan jumlah klaster optimal. Untuk contohnya, model yang akan dikalkulasi adalah model K=2. Tahapannya adalah sebagai berikut:

1. Mengambil data tabel kalkulasi jarak Euclidean pada iterasi terakhir. Jadi, setelah iterasi terakhir pada tahap kalkulasi, diperoleh juga tabel kalkulasi jarak Euclidean yang mana setiap data sudah masuk ke klasternya masing-masing. Dari data ini, akan dikalkulasi nilai rata-rata per-dimensi dari anggota setiap klaster. Contohnya, berikut hasil tabel kalkulasi jarak Euclidean pada iterasi terakhir untuk model K=2:

Tabel 14 Contoh Tabel Kalkulasi Jarak Euclidean Iterasi Terakhir (K=2)

ID	KK	EF	IP	Cluster
1	2.9	2	3.1	2
2	3.3	3.5	3.6	1
3	3	3.2	2.9	1
4	2.6	3.2	2.8	2
5	2.6	3.4	2.4	2
...	...	...	...	...
177	3.9	4	4	1

2. Kalkulasi nilai centroid atau rata-rata anggota dimensi yang masuk dalam klaster tertentu. Setelah didapatkan tabel di atas, bisa dikalkulasi nilai rata-rata anggota dalam klaster tertentu. Rumus dasar untuk tahap ini adalah:

$$C_k^{(j)} = \frac{\sum_{i=1}^N \delta(c_i = k) \cdot x_i^{(j)}}{\sum_{i=1}^N \delta(c_i = k)} \quad (7)$$

dengan keterangan sebagai berikut:

$C_K^{(j)}$ = Nilai centroid pada cluster ke-k untuk dimensi j

$xx_j^{(i)}$ = Nilai data ke-i pada dimensi j

c_i = Cluster yang diikuti oleh data ke-i δ

$(c_i = k)$ = Fungsi indikator: 1 jika data ke-i termasuk ke cluster ke- k, 0 jika tidak

N = Jumlah total data

dengan dasar rumus di atas, dipakai rumus =AVERAGEIF() di excel yang fungsinya adalah mengembalikan nilai rata-rata dimensi setiap data yang disesuaikan dengan klasternya. Dari tahap ini, diperoleh tabel nilai rata-rata dimensi anggota per-klasternya. Contohnya adalah sebagai berikut:

Tabel 15 Tabel Nilai Rata-rata Dimensi Anggota Per-klaster

Cluster	KK	EF	IP
1	3.392473	3.374194	3.373118
2	2.55	2.52381	2.547619

2. Pembuatan kolom SSE per-data dan SUM SSE di akhir tabel. Pada tabel kalkulasi jarak Euclidean, ditambahkan kolom SSE di sebelah kolom cluster yang akan diisi dengan

nilai SSE per-data. Kalkulasi SSE-nya didasari rumus berikut:

$$SSE_{x(i)} = \sum_{j=1}^n (X_j^{(i)} - C_{k(i)j})^2 \quad (8)$$

Rumus Kalkulasi Nilai Sum of Squared Errors (SSE)

$SSE_{x(i)}$ = eror kuadrat (squared errors) dari data ke-i terhadap centroid clusternya.

$X_j^{(i)}$ = nilai dimensi j dari data ke- 1

$c(i)$ = nilai rata- rata dimensi dari klaster yang diikuti oleh data ke-i.

Rumus di atas diaplikasikan pada kolom SSE menggunakan fungsi =IF() untuk menentukan masuk ke klaster mana, kemudian dikalkulasi dengan tabel nilai rata-rata dimensi per-klaster. Contoh hasilnya adalah sebagai berikut:

Tabel 16 Contoh Tabel dengan Tambahan Kolom SSE

ID	KK	EF	IP	Cluster	SSE
1	2.9	2	3.1	2	0.702001
2	3.3	3.5	3.6	1	0.075854
3	3	3.2	2.9	1	0.408219
4	2.6	3.2	2.8	2	0.52343
5	2.6	3.4	2.4	2	0.792001
...	...	...	...	...	...
177	3.9	4	4	1	1.042198

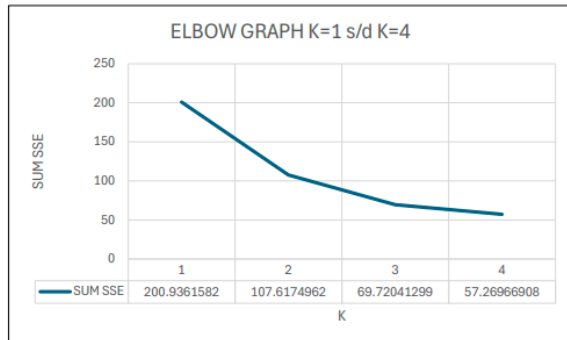
Setelah menambahkan kolom SSE, langkah selanjutnya adalah mengkalkulasi SUM SSE di akhir tabel. Untuk mengkalkulasi ini, digunakan fungsi =SUM() agar bisa mendapatkan total SSE dari setiap data. Didapati nilai SUM SSE untuk kasus K=2 senilai 107.6174962. Kalkulasi SSE dan SUM SSE juga dilakukan pada kasus K=3 dan K=4. Berikut tabel hasil SUM SSE dari keseluruhan kasus.

Tabel 17 Tabel SSE K=1 s/d K=4

K	SUM SSE
1	200.9362
2	107.6175
3	69.72041
4	57.26967

4. Pembuatan elbow graph. Elbow graph dibuat dengan membuat grafik garis/line chart dari nilai SSE di mana sumbu X diisi

dengan jumlah K dan sumbu Y diisi dengan nilai SUM SSE. Pembuatan elbow graph juga dilakukan di Excel dengan men-select tabel SUM SSE-nya, dan meng-klik opsi Quick Analysis. Kemudian, dipilih Line pada opsi Chart dan hasilnya akan terlihat seperti berikut



Gambar 4.11 Hasil Elbow Graph K=1 s/d K=4

3.3.4 Kalkulasi DBI dengan Excel

Setelah membuat elbow graph untuk menentukan jumlah kluster optimal, dilakukan kalkulasi nilai Davies-Bouldin Index atau DBI untuk mengukur kualitas pemisahan antar kluster dan kekompakan dalam kluster itu sendiri. Sebagai contoh, model yang akan dikalkulasi DBI-nya setelah ini adalah model K=3. Langkah-langkah kalkulasinya antara lain sebagai berikut:

1. Mengambil tabel kalkulasi jarak Euclidean K=3 pada iterasi terakhir. Langkah pertama ini sama persis dengan ketika ingin menghitung SSE untuk membuat elbow graph dengan Excel. Bisa diambil data yang sudah memiliki informasi terkait klasternya dan sudah tidak berubah-ubah lagi. Contohnya adalah sebagai berikut:

Tabel 18 Contoh Tabel Kalkulasi Jarak Euclidean Iterasi Terakhir (K=3)

ID	KK	EF	IP	Cluster
1	2.9	2	3.1	3
2	3.3	3.5	3.6	1
3	3	3.2	2.9	3
4	2.6	3.2	2.8	3
5	2.6	3.4	2.4	3
...	...	...	...	...
177	3.9	4	4	1

2. Kalkulasi nilai centroid atau rata-rata anggota dimensi yang masuk dalam kluster tertentu. Langkah ini juga masih sama seperti ketika ingin menghitung SSE. Pada tahap ini, setiap kluster akan memiliki nilai rata-rata dimensi dari setiap data yang masuk ke dalam kluster tersebut. Fungsi =AVERAGEIF() masih dipakai untuk memberikan parameter, agar hanya merata-ratakan nilai satu kluster tertentu saja. Hasil dari tahap ini adalah sebagai berikut:

Tabel 19 Hasil Kalkulasi Nilai Rata-rata Anggota Dimensi

Cluster	KK	EF	IP
1	3.607692	3.621154	3.661538
2	2.158065	2.009677	2.36129
3	2.92766	2.92766	2.809574

3. Membuat kolom baru untuk menghitung jarak Euclidean untuk setiap data. Hal ini hampir sama seperti saat proses clustering. Bedanya adalah sudah diketahui data tersebut masuk ke kluster mana, jadi hanya menghitung jarak ke kluster tersebut saja. Tahapan ini menggunakan fungsi =IF() untuk memberikan parameter, agar nilai jarak setiap data dihitung ke centroid yang dimasuki masing-masing data tersebut. Kemudian dipakai juga fungsi =SQRT() untuk mengkalkulasi nilai akar pangkat duanya. Contoh hasil dari tahap ini adalah sebagai berikut:

Tabel 20 Tambahan Kolom Jarak Euclidean ke Centroid Setiap Data

ID	KK	EF	IP	Cluster	Jarak ke Centroid
1	2.9	2	3.1	3	0.972452738
2	3.3	3.5	3.6	1	0.336362591
3	3	3.2	2.9	3	0.295937867
4	2.6	3.2	2.8	3	0.426171063
5	2.6	3.4	2.4	3	0.705845252
...	...	...	...	...	...
177	3.9	4	4	1	0.586109553

3. Kalkulasi nilai S atau rata-rata jarak antara anggota kluster dan centroid-nya. Dibuat kolom S di samping tabel kalkulasi rata-rata anggota dimensi dengan hasil dalam tabel dibawah ini :

Tabel 21 Nilai S dari Setiap Centroid

Centroid	KK	EF	IP	S
C1	3.607692	3.621154	3.661538	0.527655343
C2	2.158065	2.009677	2.36129	0.722902291
C3	2.92766	2.92766	2.809574	0.543696298

4. Kalkulasi nilai M atau jarak antar centroid dari dua klaster . dikalkulasi nilai M dari data. Karena ada 3 klaster, maka akan ada nilai. Nilai ini yakni nilai M_{12} , M_{23} , dan M_{13} . Digunakan fungsi $=SQRT()$ untuk memberikan nilai akar pangkat 2 dari hasil penjumlahan selisih nilai centroid kuadrat. Hasil dari tahap ini adalah sebagai berikut:

Tabel 22 Tabel Nilai M Antar Cluster

M_{ij}	Nilai
M12	2.527631744
M23	1.279033411
M13	1.291983541

5. Kalkulasi nilai R atau rasio kesamaan antar dua cluster . dikalkulasi nilai R dari data. Nilai R juga akan dibagi menjadi 3, yakni nilai R_{12} , R_{23} , dan R_{13} . Didapati hasil nilai R-nya adalah sebagai berikut:

Tabel 23 Tabel Nilai R Antar-cluster

R_{ij}	Nilai
R12	0.494755
R23	0.990278
R13	0.82923

6. Membuat tabel Max R. Tabel ini merupakan tabel yang menunjukkan nilai maksimum atau tertinggi dari setiap kandidat R_{ij} . Tabel ini memiliki 3 kolom, yakni kolom klaster, kandidat R, dan Max R. Contohnya adalah sebagai berikut:

Tabel 24 Tabel Max R

Cluster	Kandidat R	Max R
1	R_{12}, R_{13}	0.82923
2	R_{12}, R_{23}	0.990278
3	R_{13}, R_{23}	0.990278

Dari tabel di atas, bisa dilihat bahwa pada klaster 1, Max R-nya adalah 0.82923 karena nilai R_{13} lebih tinggi daripada R_{12} . Begitu

juga pada klaster selanjutnya, di mana nilai Max R diambil dari kandidat R yang bernilai lebih tinggi.

8. Penghitungan nilai DBI. Nilai DBI dikalkulasi menggunakan rumus sebagai berikut:

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} R_{ij} \quad (9)$$

Rumus Kalkulasi Nilai DBI

DBI = Davies-Bouldin Index.

K = Jumlah total klaster.

R_{ij} = Nilai rasio antara klaster i dan j

$\max R_{ij}$ = Nilai maksimum dari R_{ij} untuk semua $j \neq i$ (klaster lain).

dengan keterangan sebagai berikut: Dengan menggunakan rumus di atas, dikalkulasi DBI-nya dengan rumus $=AVERAGE()$ yang memberikan nilai rata-rata dari kolom Max R. Didapati nilai DBI untuk kasus $K=2$ adalah sebesar 0.936595. Setelah kalkulasi DBI untuk model $K=2$ selesai, dilakukan juga kalkulasi manual untuk model $K=3$ dan $K=4$. Hasilnya sendiri bisa dilihat pada tabel di bawah:

Tabel 25 Hasil DBI untuk $K=2$ s/d $K=4$

K	DBI
2	0.990995473
3	0.936595336
4	1.096798385

Berdasarkan hasil perhitungan Davies-Bouldin Index (DBI) pada tabel di atas, diketahui bahwa nilai DBI untuk jumlah cluster $K=2$ adalah sebesar 0,9909, untuk $K=3$ sebesar 0,9366, dan untuk $K=4$ meningkat menjadi 1,0968. Prinsip evaluasi DBI adalah semakin rendah nilainya, maka kualitas klaster semakin baik (Shalsadilla et al., 2023). Maka dari itu, jumlah klaster optimal terdapat pada model $K=3$ dengan nilai DBI terendah. Hal ini menunjukkan bahwa pada saat data dibagi menjadi 3 klaster, hasil clustering menunjukkan tingkat kompaksi dalam klaster yang tinggi serta pemisahan antar-klaster yang baik dibanding jumlah klaster lainnya.

4. KESIMPULAN

Hasil penelitian menunjukkan adanya tiga klaster mahasiswa dalam pemilihan Universitas Ary Ginanjar (UAG). Klaster 1 (29%) memiliki persepsi sangat positif terhadap kualitas kampus, faktor ekonomi, serta promosi, dengan dimensi efek iklan dan promosi sebagai faktor paling dominan. Klaster 2 (18%) menunjukkan tingkat kesetujuan rendah terhadap seluruh faktor, terutama aspek ekonomi dan finansial, yang berpotensi menjadi penghambat keputusan memilih UAG. Klaster 3 (53%) merepresentasikan mahasiswa dengan persepsi netral, di mana kualitas kampus dan kemudahan finansial menjadi faktor utama yang memengaruhi pilihan mereka. Hasil ini memberikan dasar bagi kampus untuk menyusun strategi pemasaran yang lebih terarah, seperti memperkuat kampanye promosi untuk Klaster 1, memberikan informasi beasiswa bagi Klaster 2, serta menonjolkan keseimbangan kualitas dan keterjangkauan biaya untuk Klaster 3. Penelitian ini dapat digunakan sebagai rekomendasi untuk UAG dalam strategi pemasaran pada segmentasi psikografis sehingga dapat membantu kampus lebih personal dan efektif dalam menghadapi persaingan

REFERENSI

- Ajeng Dwita Ayuningtyas. (2024, April 5). *Negara-Negara dengan Jumlah Perguruan Tinggi Terbanyak di Dunia*. <https://goodstats.id/article/negara-negara-ini-dipenuhi-perguruan-tinggi-285bU>
- Akrami, K., Akrami, M., Akrami, F., & Hakimi, M. (2024). Investigating the Integration of Big Data Technologies in Higher Education Settings. *Indonesian Journal of Multidisciplinary on Social and Technology*, 2(2), 1–12. <https://doi.org/10.31004/ijmst.v2i2.296>
- Arpay, M. N. (2023). Student Mining Using K-Means Clustering: A Basis For Improving Higher Education Marketing Strategies Psychology And Education: A Multidisciplinary Journal Student Mining Using K-Means Clustering: A Basis For Improving Higher Education Marketing Strategies. *Psych Educ*, 14, 2023–1233. <https://doi.org/10.5281/zenodo.8383341>
- Badhera, U., Verma, A., & Nahar, P. (2022). Applicability of K-medoids and K-means algorithms for segmenting students based on their scholastic performance. *Journal of Statistics and Management Systems*, 25(7), 1621–1632. <https://doi.org/10.1080/09720510.2022.2130566>
- Chukwuere, J. (2024). Rapid review of the COVID-19 pandemic's impact on the digitalization of higher education. *Research on Education and Media*, 16(1), 1–9. <https://doi.org/10.2478/rem-2024-0002>
- Dwi Hadya Jayani. (2021, March 27). *Mahasiswa Baru Indonesia Naik 20,13% pada 2019*. <https://databoks.katadata.co.id/pendidikan/statistik/83102dace0b3870/mahasiswa-baru-indonesia-naik-2013-pada-2019?form=MG0AV3>
- Gie, W., & Jollyta, D. (2020). Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index: Status Covid-19 Wilayah Riau. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2, 187–191.
- Goodrich, K., Swani, K., & Munch, J. (2020). How to connect with your best student prospects: Saying the right things, to the right students, in the right media. *Journal of Marketing Communications*, 26(4), 434–453. <https://doi.org/10.1080/13527266.2018.1514319>
- Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology (Singapore)*, 12(4), 1243–1257. <https://doi.org/10.1007/s41870-020-00427-7>

- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- Hardika Khusnuliawati, & Riskiana Putri, D. (2021). Identifikasi Segmen Pasar Mahasiswa Perguruan Tinggi Menggunakan Analisis Klaster Berdasarkan Variabel Psikografis. *Risenologi*, 6(1b), 44–49. <https://doi.org/10.47028/j.risenologi.2021.61b.243>
- Hart, P. F., & Rodgers, W. (2023). Competition, competitiveness, and competitive advantage in higher education institutions: a systematic literature review. *Studies in Higher Education*. <https://doi.org/10.1080/03075079.2023.2293926>
- Heryati, A., & Herdiansyah, M. I. (2020). The Application of Data Mining by using K Means Clustering Method in Determining New Students' Admission Promotion Strategy. *International Journal of Engineering and Advanced Technology*, 9(3), 824–833. <https://doi.org/10.35940/ijeat.C5414.029320>
- H, H. K., & Shaline, A. S. T. (2022). Data Mining Techniques and Applications. *International Journal for Research in Applied Science and Engineering Technology*, 10(12), 1139–1142. <https://doi.org/10.22214/ijraset.2022.48009>
- Kannengiesser, U., & Gero, J. S. (2023). Modelling The Design Of Models: An Example Using CRISP-DM. *Proceedings of the Design Society*, 3, 2705–2714. <https://doi.org/10.1017/pds.2023.271>
- Kara, A., & Kaynak, E. (1997). Markets of a single customer: exploiting conceptual developments in market segmentation. *European Journal of Marketing*, 31(11–12), 873–895. <https://doi.org/10.1108/03090569710190587>
- Lestari, W., Bina, S., & Kendari, B. (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari). In *SIMKOM* (Vol. 4, Issue 2). <http://ejurnal.stmikbinsa.ac.id/index.php/simkom35>
- Lin, C. F. (2002). Segmenting customer brand preference: Demographic or psychographic. *Journal of Product & Brand Management*, 11(4), 249–268. <https://doi.org/10.1108/10610420210435443>
- LU, M. (2023). Research on the application of university education management driven by big data. *Region - Educational Research and Reviews*, 5(4), 143. <https://doi.org/10.32629/rerr.v5i4.1410>
- McDonald, Malcolm., & Dunbar, Ian. (2013). *Market segmentation : how to do it, how to profit from it*. John Wiley & Sons.
- Misra, D. P., & Agarwal, V. (2020). GENERATING WORKING HYPOTHESES FOR ORIGINAL RESEARCH STUDIES. *Central Asian Journal of Medical Hypotheses and Ethics*, 1(1), 14–19. <https://doi.org/10.47316/cajmhe.2020.1.1.02>
- Monavia Ayu Rizaty. (2023, February 15). *Jumlah Mahasiswa Indonesia Sebanyak 9,32 Juta Orang pada 2022*. <https://dataindonesia.id/pendidikan/detail/jumlah-mahasiswa-indonesia-sebanyak-932-juta-orang-pada-2022?form=MG0AV3>
- Moubayed, A., Injadat, M., Shami, A., & Lutfiyya, H. (2020). Student Engagement Level in an e-Learning Environment: Clustering Using K-means. *American Journal of Distance Education*, 34(2), 137–156. <https://doi.org/10.1080/08923647.2020.1696140>
- Murairwa, S. (2015). VOLUNTARY SAMPLING DESIGN. *International Journal of Advanced Research in Management and Social Sciences*, 4(2).

- <https://www.researchgate.net/publication/340000298>
- Mussabayev, R., Mladenovic, N., Jarboui, B., & Mussabayev, R. (2022). *How to Use K-means for Big Data Clustering?* <https://doi.org/10.1016/j.patcog.2022.109269>
- North, Matthew. (2012). *Data mining for the masses*. Global Text Project.
- Pamungkas, L., Dewi, N. A., & Putri, N. A. (2024). Classification of Student Grade Data Using the K-Means Clustering Method. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 13(1), 86–91. <https://doi.org/10.32736/sisfokom.v13i1.1983>
- Partelow, S. (2023). What is a framework? Understanding their purpose, value, development and use. *Journal of Environmental Studies and Sciences*, 13(3), 510–519. <https://doi.org/10.1007/s13412-023-00833-w>
- Prastyabudi, W. A., Alifah, A. N., & Nurdin, A. (2024). Segmenting the Higher Education Market: An Analysis of Admissions Data Using K-Means Clustering. *Procedia Computer Science*, 234, 96–105. <https://doi.org/10.1016/j.procs.2024.02.156>
- Sarkum, S. (2020). Dissecting Changes in Consumer Characteristics through Psychographics. In *Promotion and Marketing Communications*. IntechOpen. <https://doi.org/10.5772/intechopen.89635>
- Sarli, A. (2011). *The Role of Psychographic For Distinguishing Main Categories of Consumers Based On Lifestyle, Personality And Value Variables*. www.ijeronline.com
- Setiyo, M., & Rochman, M. L. (2023). Literature review: An effective method for identifying science and technology updates. *Mechanical Engineering for Society and Industry*, 3(3 Special issue), 114–118. <https://doi.org/10.31603/mesi.10787>
- Shalsadilla, N., Martha, S., Perdana, H., Satyahadewi, N., Sulistianingsih, E., Program,), & Fakultas, S. S. (2023). *Penentuan Jumlah Cluster Optimum Menggunakan Davies Bouldin Index dalam Pengelompokan Wilayah Kemiskinan di Indonesia* (Vol. 23, Issue 1). <https://bps.go.id>
- Sultan Alalawi, S. J., Mohd Shaharane, I. N., & Mohd Jamil, J. (2023). CLUSTERING STUDENT PERFORMANCE DATA USING k-MEANS ALGORITHMS. *Journal of Computational Innovation and Analytics (JCIA)*, 2(1), 41–55. <https://doi.org/10.32890/jcia2023.2.1.3>
- Vazrina Putri. (2023, March 6). *Jurusan Kuliah dengan Pertumbuhan Mahasiswa Terbesar di Indonesia*. <https://journal.revou.co/pertumbuhan-program-studi-2023/?form=MG0AV3>
- Wicaksono, Y. (2019). Segmentasi Pelanggan Bisnis Dengan Multi Kriteria Menggunakan K-Means. *Indonesian Journal of Business Intelligence (IJUBI)*, 1(2), 45. <https://doi.org/10.21927/ijubi.v1i2.872>
- Wusu, O., & Jnr, B. (2021). K-Means, Clustering Algorithm For Student's Selection And Performance Prediction. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, 10(10). www.ijstr.org
- Zhang, T. (2022). *Asymptotics for The k-means*. <http://arxiv.org/abs/2211.10015>
- [12] Suwardika, I. G. I., Suariana, I. G. N., Bhiantara, I. B. P., & Arso, N. Y. (2019). Prediksi tingkat kelulusan mahasiswa tepat waktu menggunakan Naïve Bayes: Studi kasus Fakultas Ekonomi dan Bisnis Universitas Pendidikan Nasional. *Jurnal Ilmu Komputer Indonesia (JIKI)*, 4(2).