

Evaluating K-Means and K-Medoids Using Silhouette Score for Eysenck Personality-Based Clustering of Prospective Students

Muhammad Arifin ^{1*}, Ifnu Wisma Dwi Prastya ², Jauhara Rana Budiani ³

^{1,2}Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri

³Statistika, Universitas Nahdlatul Ulama Sunan Giri

muhammadarifin3054@gmail.com¹, ifnuprastya@unugiri.ac.id², jbudiani@unugiri.ac.id³

Article Info

Article history:

Received 2026-01-22

Revised 2026-03-03

Accepted 2026-04-10

Keyword:

Clustering,
K-Means,
K-Medoids,
Personality,
Study Program,
Silhouette Score.

ABSTRACT

The selection of an academic major aligned with students' personality characteristics plays a crucial role in enhancing academic performance and long-term career development. This study compares the performance of the K-Means and K-Medoids algorithms in clustering 118 students from the Bojonegoro region based on introversion–extraversion tendencies, operationalized using Eysenck's personality framework. Data were collected through a 30-item personality questionnaire measured on a five-point Likert scale and validated through expert-based content validity assessment. Preprocessing involved transforming responses into numerical values to enable distance-based clustering analysis. Four clustering scenarios were evaluated by combining two algorithms (K-Means and K-Medoids) with two distance metrics (Euclidean and Manhattan). Cluster quality was assessed using the Silhouette Score as an internal validation measure. The results show that K-Means with Euclidean Distance achieved the highest Silhouette Score of 0.44013, indicating a moderate cluster structure and outperforming the other configurations. External validation using expert-defined ground truth revealed 100% agreement for the introverted cluster, while the extraverted cluster demonstrated several mismatches, reflecting the heterogeneous and continuous nature of extraversion traits. These findings highlight the importance of algorithm and distance metric selection in personality-based clustering and demonstrate the potential of integrating psychological theory with data mining techniques to support objective and evidence-based academic major recommendation systems.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Pemilihan program studi merupakan keputusan strategis yang tidak hanya menentukan jalur akademik, tetapi juga berdampak jangka panjang terhadap perkembangan karier mahasiswa [1],[2]. Keputusan yang selaras dengan minat, kemampuan kognitif, serta karakteristik kepribadian individu diketahui dapat meningkatkan motivasi intrinsik, persistensi akademik, dan capaian hasil belajar [3]. Salah satu dimensi kepribadian yang banyak dikaji dalam konteks pendidikan adalah kecenderungan introvert dan ekstrovert, sebagaimana dirumuskan dalam model kepribadian Eysenck, yang berperan dalam membentuk preferensi terhadap lingkungan belajar, pola interaksi sosial, dan strategi kognitif individu [4]. Ketidaksesuaian antara orientasi kepribadian dan tuntutan

akademik suatu program studi berpotensi memicu kesulitan adaptasi, penurunan keterlibatan belajar, serta rendahnya kepuasan studi [5]. Oleh karena itu, integrasi asesmen kepribadian dalam proses orientasi akademik menjadi penting untuk mendukung pengambilan keputusan studi yang lebih adaptif dan berkelanjutan [6].

Perkembangan teknologi informasi telah membuka peluang penerapan teknik data mining dalam mendukung pengambilan keputusan akademik yang lebih objektif dan berbasis data, termasuk dalam proses pemilihan program studi [7], [8]. Salah satu pendekatan yang banyak digunakan adalah clustering, yaitu metode unsupervised learning yang mengelompokkan data berdasarkan tingkat kemiripan karakteristik intrinsik. Dalam konteks penelitian ini, karakteristik kepribadian calon mahasiswa digunakan sebagai

variabel input untuk membentuk kluster yang merepresentasikan kecenderungan introvert dan ekstrovert. Dalam implementasinya, algoritma *K-Means* dan *K-Medoids* merupakan dua metode klusterisasi yang paling umum digunakan. Algoritma *K-Means* memiliki keunggulan dari sisi efisiensi komputasi karena menentukan pusat kluster berdasarkan nilai rata-rata anggota kluster [7], namun bersifat sensitif terhadap keberadaan outlier. Sebaliknya, algoritma *K-Medoids* lebih robust karena menetapkan pusat kluster berdasarkan objek aktual (medoid) dalam dataset, sehingga relatif lebih stabil ketika data mengandung noise atau nilai ekstrem [9]. Oleh karena itu, pemilihan algoritma yang tepat menjadi aspek krusial dalam memastikan kualitas, validitas, dan interpretabilitas hasil klusterisasi sebagai dasar pendukung rekomendasi akademik.

Sejumlah penelitian terdahulu telah mengaplikasikan algoritma *K-Means* dan *K-Medoids* dalam berbagai konteks pendidikan, seperti segmentasi mahasiswa, analisis profil kepribadian, serta pengembangan sistem rekomendasi akademik [10]. Namun, efektivitas kedua algoritma tersebut sangat dipengaruhi oleh karakteristik dataset dan tujuan analisis yang digunakan. Algoritma *K-Means* umumnya menunjukkan performa optimal pada dataset berukuran besar dengan distribusi relatif homogen dan minim outlier [11], [12], sedangkan *K-Medoids* cenderung lebih stabil ketika diterapkan pada data yang heterogen atau mengandung nilai ekstrem karena tidak bergantung pada nilai rata-rata aritmetika [13]. Meskipun demikian, penelitian yang secara eksplisit membandingkan kinerja kedua algoritma dalam konteks pemilihan program studi berbasis dimensi kepribadian, khususnya kecenderungan introvert dan ekstrovert, masih terbatas [14]. Selain itu, sebagian besar studi sebelumnya belum melakukan evaluasi kualitas kluster secara komprehensif menggunakan metrik validasi internal yang objektif, seperti *Silhouette Score*, yang mampu menilai koherensi intra-kluster dan pemisahan antar-kluster secara simultan [15]. Kesenjangan ini menunjukkan perlunya penelitian yang tidak hanya membandingkan performa algoritmik, tetapi juga mengaitkannya dengan relevansi psikologis dan akademik dalam mendukung proses orientasi studi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membandingkan kinerja algoritma *K-Means* dan *K-Medoids* dalam mengelompokkan calon mahasiswa berdasarkan dimensi kepribadian introversi dan ekstrovert. Proses klusterisasi dilakukan menggunakan data karakteristik psikologis responden, dengan kualitas kluster dievaluasi secara objektif melalui metrik *Silhouette Score* yang mengukur koherensi intra-kluster dan keterpisahan antar-kluster. Selain evaluasi kuantitatif, hasil klusterisasi juga dianalisis secara kualitatif untuk menilai kesesuaian pola pengelompokan dengan konstruk kepribadian teoretis. Penelitian ini diharapkan memberikan kontribusi secara teoretis melalui pengayaan kajian penerapan metode clustering dalam konteks psikometri pendidikan, serta kontribusi praktis dalam pengembangan kerangka kerja

sistem rekomendasi program studi yang lebih adaptif, objektif, dan berbasis bukti. Dengan demikian, temuan penelitian ini dapat menjadi dasar dalam perancangan proses orientasi akademik yang lebih efektif guna meminimalkan ketidaksesuaian jurusan serta meningkatkan keberhasilan dan kepuasan akademik mahasiswa.

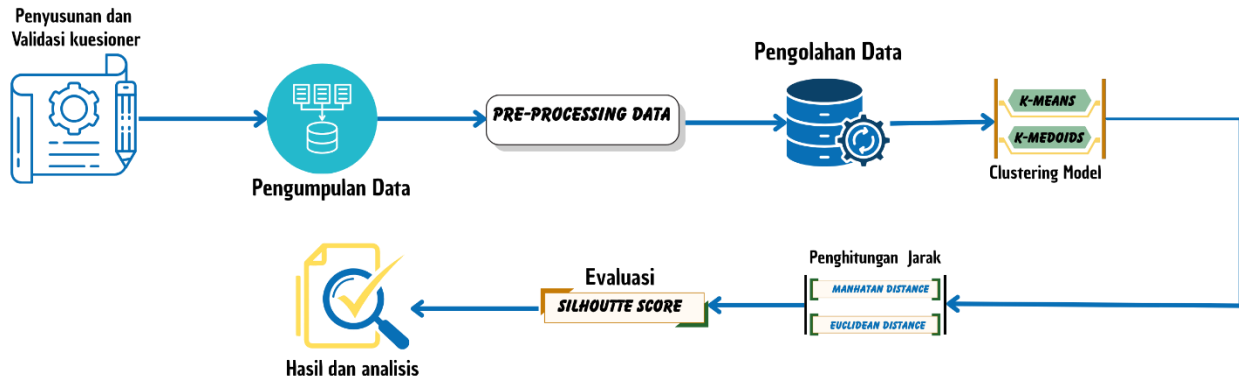
II. METODE PENELITIAN

Penelitian ini diawali dengan penyusunan instrumen kuesioner berbasis dimensi kepribadian introversi dan ekstrovert yang selanjutnya divalidasi melalui uji ahli (*expert judgment*) untuk memastikan kesesuaian setiap butir pernyataan dengan konstruk kepribadian yang diukur. Setelah instrumen dinyatakan layak, data dikumpulkan dari responden calon mahasiswa. Tahap pra-proses (pre-processing) dilakukan dengan mengonversi setiap respons kuesioner berbasis skala Likert ke dalam bentuk numerik agar dapat diolah oleh algoritma klusterisasi. Proses konversi ini bertujuan untuk mentransformasikan data kategorikal menjadi data numerik tanpa mengubah makna jawaban responden, sehingga dataset siap digunakan dalam analisis klusterisasi menggunakan algoritma *K-Means* dan *K-Medoids*.

Tahap analisis dilakukan dengan menerapkan dua algoritma klusterisasi, yaitu *K-Means* dan *K-Medoids*, yang masing-masing dievaluasi menggunakan dua metrik jarak, yaitu *Euclidean Distance* dan *Manhattan Distance*. Pemilihan kedua metrik jarak tersebut didasarkan pada karakteristik data numerik hasil konversi skala Likert, yang memungkinkan penggunaan berbagai ukuran jarak dalam proses klusterisasi. *Euclidean Distance* digunakan untuk mengukur jarak geometris antar objek secara langsung, sedangkan *Manhattan Distance* menghitung jarak berdasarkan jumlah selisih absolut antar atribut, sehingga memberikan perspektif pengukuran jarak yang berbeda (Aggarwal et al., 2001) [16]. Kombinasi ini menghasilkan empat skenario klusterisasi, yaitu *K-Means* dengan *Euclidean Distance*, *K-Means* dengan *Manhattan Distance*, *K-Medoids* dengan *Euclidean Distance*, dan *K-Medoids* dengan *Manhattan Distance*. Pendekatan komparatif ini digunakan untuk mengidentifikasi konfigurasi algoritma dan metrik jarak yang menghasilkan kualitas kluster terbaik berdasarkan nilai *Silhouette Score*.

Keempat skenario klusterisasi dievaluasi menggunakan *Silhouette Score*, yaitu metrik validasi internal yang mengukur tingkat koherensi intra-kluster (cohesion) dan keterpisahan antar-kluster (separation) secara simultan. Nilai *Silhouette* berada pada rentang -1 hingga $+1$, di mana nilai yang mendekati $+1$ menunjukkan bahwa suatu objek berada dalam kluster yang kompak dan terpisah dengan baik dari kluster lain, sedangkan nilai yang mendekati nol atau bernilai negatif mengindikasikan adanya tumpang tindih antar-kluster atau penempatan objek yang kurang tepat (Rousseeuw, 1987) [17]. Nilai rata-rata *Silhouette Score* dari masing-masing kombinasi algoritma dan metrik jarak kemudian dibandingkan untuk menentukan konfigurasi yang menghasilkan struktur klusterisasi paling optimal. Alur

metodologi penelitian secara keseluruhan disajikan pada Gambar 1.



Gambar 1. Langkah Metode Penelitian

A. Pengumpulan dan Sumber Data

Dataset dalam penelitian ini diperoleh melalui survei menggunakan instrumen kuesioner kepribadian yang dikembangkan berdasarkan konsep introversi dan ekstrovertasi menurut teori Eysenck, yang memandang dimensi ekstrovertasi-introversi sebagai spektrum temperamen biologis yang merefleksikan perbedaan individu dalam merespons stimulasi eksternal [18]. Pengembangan instrumen juga mempertimbangkan perspektif psikologis, khususnya terkait orientasi energi psikis dan preferensi interaksi sosial, sebagai landasan pelengkap dalam operasionalisasi konstruk kepribadian [19]. Berdasarkan kedua kerangka teoretis tersebut, lima aspek utama kepribadian, yaitu orientasi energi, preferensi sosial, sensitivitas terhadap rangsangan, pola aktivitas, dan gaya pengambilan keputusan, diadaptasi menjadi butir pernyataan yang relevan dan kontekstual. Instrumen kuesioner yang digunakan terdiri atas 30 item yang diukur menggunakan skala Likert lima tingkat, dengan rentang respons mulai dari “1 Sangat Tidak Setuju” hingga “5 Sangat Setuju”.

Sebelum digunakan dalam pengumpulan data utama, instrumen kuesioner terlebih dahulu divalidasi melalui pendekatan *content validity* oleh ahli di bidang psikologi kepribadian untuk memastikan kesesuaian setiap butir pernyataan dengan konstruk teoretis introversi-ekstrovertasi. Proses validasi dilakukan dengan menilai keterwakilan butir terhadap dimensi karakteristik kepribadian yang relevan, seperti orientasi energi, preferensi sosial, dan respons terhadap stimulasi. Masukan dan saran dari para ahli digunakan sebagai dasar untuk melakukan perbaikan redaksi dan penyempurnaan item instrumen. Setelah instrumen dinyatakan layak, kuesioner disebar kepada mahasiswa yang berasal dari wilayah Bojonegoro dan sekitarnya sebagai responden penelitian, dengan teknik pengambilan sampel purposive sampling berdasarkan kriteria mahasiswa aktif yang bersedia berpartisipasi dalam penelitian.

B. Pre-processing Data

Data yang telah dikumpulkan selanjutnya melalui tahap pre-processing untuk menyiapkan data agar dapat diolah oleh algoritma klusterisasi. Tahap praproses yang dilakukan dalam penelitian ini mencakup proses transformasi data, yaitu mengonversi setiap respons kuesioner berbasis skala Likert ke dalam bentuk numerik dengan rentang nilai 1–5. Proses konversi ini bertujuan untuk mengubah data kualitatif menjadi data numerik tanpa mengubah makna jawaban responden, sehingga dataset dapat diproses secara matematis oleh algoritma K-Means dan K-Medoids yang berbasis perhitungan jarak.

Karena seluruh item berada pada skala Likert 1–5 yang seragam dan memiliki rentang pengukuran yang sama, proses normalisasi tambahan tidak dilakukan. Seluruh variabel telah berada dalam skala yang homogen sehingga tidak menimbulkan bias skala dalam perhitungan jarak Euclidean maupun Manhattan. Dengan kondisi tersebut, data dinilai telah memenuhi prasyarat untuk langsung digunakan dalam proses klusterisasi.

C. Evaluasi Klusterisasi

Data mining merupakan proses penggalian informasi penting dari sejumlah data yang besar dengan menggunakan metode, algoritma, dan teknik statistik tertentu [20]. Proses ini bertujuan untuk menemukan pola, hubungan, atau struktur tersembunyi yang tidak terlihat secara langsung pada data mentah [21], [22]. Dalam konteks penelitian ini, data mining digunakan untuk mengelompokkan kecenderungan kepribadian calon mahasiswa berdasarkan jawaban kuesioner.

1) Perhitungan Algoritma K-Means Clustering:

Algoritma *K-Means* adalah metode klusterisasi yang bertujuan membagi data ke dalam k kelompok berdasarkan tingkat kemiripan antar objek. Setiap kluster direpresentasikan oleh sebuah centroid [23], [24], yaitu titik pusat yang

diperoleh dari nilai rata-rata seluruh data dalam kluster tersebut. Proses klusterisasi dilakukan dengan menghitung jarak antara setiap data dan centroid, kemudian menempatkan data pada kluster yang jaraknya paling dekat.

Keunggulan *K-Means* adalah prosesnya yang sederhana, komputasi cepat, serta efektif diterapkan pada dataset berukuran besar.

Fungsi objektif *K-Means* dirumuskan sebagai berikut:

$$dJ = \sum_{j=1}^k \sum_{x_i \in C_j}^{n_j} \|x_i^{(j)} - c_j\|^2 \quad (1)$$

Keterangan:

J = total jarak dalam seluruh cluster (fungsi objektif yang diminimalkan),

$x_i^{(j)}$ = data ke-i pada cluster ke-j,

c_j = centroid dari cluster ke-j,

n_j = jumlah data dalam cluster ke-j.

Nilai J menunjukkan seberapa baik data dikelompokkan; semakin kecil nilai J , semakin baik hasil pengelompokan. Setelah seluruh data dikelompokkan, centroid baru dihitung menggunakan rata-rata dari semua anggota cluster:

$$c_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i^{(j)} \quad (2)$$

Keterangan:

c_j = centroid baru pada cluster ke-j

n_j = jumlah data dalam cluster j

$x_i^{(j)}$ = data ke-i pada cluster j

K-Means terus mengulang proses penghitungan jarak dan pembaruan centroid hingga tidak terjadi perubahan signifikan (konvergen).

2) Perhitungan Algoritma *K-Medoids Clustering*:

K-Medoids merupakan metode klusterisasi yang prinsip kerjanya mirip dengan *K-Means*, namun pusat kluster bukan nilai rata-rata, melainkan medoid, yaitu salah satu data aktual dalam kluster yang memiliki total jarak minimum terhadap anggota lainnya[10], [25]. Pemilihan medoid membuat metode ini lebih stabil terhadap outlier, karena tidak dipengaruhi nilai ekstrem seperti pada rerata[26].

Rumus *Euclidean Distance*:

$$J = \sum_{j=1}^k \sum_{x_i \in C_j} d = (x_i, m_j) \quad (3)$$

Keterangan:

J = total jarak dalam seluruh cluster,

x_i = data ke-i yang termasuk dalam cluster C_j ,

m_j = medoid dari cluster ke-j,

$d(x_i, m_j)$ = jarak antara data ke-i dan medoid.

Rumus ini menunjukkan bahwa *K-Medoids* berusaha meminimalkan total jarak absolut antara seluruh data dan medoid cluster-nya. Jika pertukaran medoid menyebabkan penurunan nilai J , maka medoid baru akan dipilih.

3) Perhitungan Jarak *Euclidean Distance*

Euclidean Distance adalah ukuran jarak geometris paling umum yang menggambarkan jarak garis lurus antara dua titik dalam ruang berdimensi- n [27]. Metode ini bekerja berdasarkan prinsip teorema Pythagoras, sehingga sangat baik digunakan ketika hubungan antar atribut bersifat linier dan data tidak banyak mengandung outlier.

Rumus *Euclidean Distance*:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (4)$$

Keterangan:

p_i, q_i = nilai atribut ke-i dari objek p dan q,

n = jumlah atribut yang dibandingkan.

Semakin kecil nilai $d(p, q)$, semakin tinggi kemiripan antara kedua data tersebut.

4) Perhitungan Jarak *Manhattan Distance* :

Manhattan Distance (atau *City Block Distance*) menghitung jarak antara dua titik dengan menjumlahkan selisih absolut tiap atributnya[28]. Perhitungannya mirip pola jalan kotak-kotak di kota Manhattan, sehingga jarak yang diukur merupakan total langkah horizontal + vertikal.

$$d(p, q) = \sum_{i=1}^n |p_i - q_i| \quad (5)$$

Keterangan:

p_i, q_i = nilai atribut ke-i dari objek p dan q,

n = jumlah atribut yang digunakan.

Semakin kecil nilai jarak $d(p, q)$, semakin besar kemiripan antar data tersebut

Kelebihan : lebih tahan terhadap outlier dibanding *Euclidean*.

Kekurangan : dapat memberikan nilai jarak lebih besar ketika atribut berskala besar.

D. Data Mining

Evaluasi hasil proses klusterisasi dilakukan dengan menggunakan *Silhouette Score*, yaitu metode yang digunakan untuk menilai sejauh mana suatu data ditempatkan pada kluster yang paling sesuai[29]. Nilai *Silhouette* menggambarkan tingkat kedekatan suatu objek terhadap kluster sendiri dibandingkan dengan kedekatannya terhadap kluster lain. Adapun rumus *Silhouette Score* adalah sebagai berikut:

$$S_i = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Keterangan:

- a(i) = rata-rata jarak data ke-i dengan seluruh anggota dalam klaster yang sama.
 - b(i) = rata-rata jarak data ke-i dengan anggota klaster terdekat lainnya.
 - S(i) = nilai Silhouette untuk data ke-i.
- Interpretasi nilai Silhouette:
- S(i) = +1 → data sangat sesuai berada dalam klasternya (klaster sangat baik).
 - S(i) = 0 → data berada di area perbatasan antara dua klaster.
 - S(i) = -1 → data kemungkinan besar salah ditempatkan.

Nilai rata-rata *Silhouette Score* digunakan untuk menilai kualitas klasterisasi secara keseluruhan. Semakin tinggi nilai rata-rata (mendekati 1), semakin baik struktur klaster yang terbentuk pada algoritma yang digunakan.

III. HASIL DAN PEMBAHASAN

Hasil validasi kuesioner kepribadian yang digunakan dalam penelitian, serta pembahasan mengenai proses expert judgment yang dilakukan untuk memastikan kelayakan instrumen sebelum digunakan pada tahap pengumpulan data.

A. Penyusunan dan Validasi kuesioner

Proses validasi kuesioner kepribadian dilakukan melalui penilaian oleh ahli (expert judgment), yaitu Khoirotus Silfiyah, M.Psi., Psikolog. Validasi ini bertujuan untuk memastikan bahwa butir pernyataan dalam kuesioner telah sesuai dengan indikator kepribadian introvert dan ekstrovert, yang penentuan dimensinya mengacu pada hasil kajian dan temuan dari beberapa penelitian terdahulu, serta mampu mengukur konstruk kepribadian yang dimaksud secara tepat. Sebagai dasar penilaian, struktur dimensi, indikator, serta sebaran item kuesioner disusun dalam bentuk blueprint instrumen yang disajikan pada Tabel di bawah ini.

TABEL I
TABEL DIMENSI INTROVERT

No	Dimensi introvert	Item		jumlah
		Favorabel	Unfavorabel	
1	Preferensi kesendirian	4	19	2
2	Gaya komunikasi	25, 26	23, 24	4
3	Kenyamanan kelompok kecil	5	2, 20, 21	4
4	Kontrol diri	6, 13, 14, 29, 30	11, 12, 27	8
5	Respons sosial	22	1, 3	3
jumlah		10	11	21

Dalam proses validasi, ahli melakukan penelaahan terhadap setiap butir pernyataan berdasarkan beberapa aspek, meliputi kesesuaian item dengan indikator pada masing-masing dimensi introvert dan ekstrovert sebagaimana tercantum dalam Tabel di atas, kejelasan redaksi kalimat, ketepatan penggunaan bahasa, serta potensi ambiguitas makna yang dapat memengaruhi pemahaman responden.

Penilaian dilakukan secara sistematis untuk memastikan bahwa setiap item dapat merepresentasikan karakteristik kepribadian secara akurat.

TABEL II
TABEL DIMENSI EKSTROVERT

No	Dimensi Ekstrovert	Item		jumlah
		Favorabel	Unfavorabel	
1	Inisiatif komunikasi	19	4	2
2	Kenyamanan dalam kelompok besar	2, 20, 21	22	4
3	Ekspresi diri	23, 24	25, 26	4
4	Aktivitas sosial	1, 3	5	3
5	Gaya interaksi	11, 12, 27	6, 13, 14	6
jumlah		11	8	19

Berdasarkan hasil penilaian yang dilakukan oleh Khoirotus Silfiyah, M.Psi., Psikolog, kuesioner dinyatakan valid secara isi (content validity) dengan beberapa catatan perbaikan minor. Perbaikan tersebut terutama berkaitan dengan penyederhanaan redaksi kalimat agar lebih mudah dipahami oleh responden. Seluruh saran yang diberikan telah ditindaklanjuti dengan melakukan revisi pada item terkait tanpa mengubah tujuan pengukuran. Setelah proses revisi dilakukan, kuesioner dinyatakan layak digunakan dalam pengumpulan data penelitian.

Selain uji validitas isi, instrumen juga diuji reliabilitasnya untuk mengukur konsistensi internal antar item menggunakan koefisien Cronbach's Alpha. Hasil perhitungan menunjukkan nilai Cronbach's Alpha sebesar 0,798, yang berada di atas batas minimal 0,70. Nilai ini menunjukkan bahwa instrumen memiliki tingkat reliabilitas yang baik (good reliability) dan konsistensi internal yang memadai dalam mengukur konstruk kepribadian introversi-ekstroversi. Dengan demikian, kuesioner dinyatakan valid dan reliabel untuk digunakan dalam tahap pengumpulan data dan analisis klasterisasi.

B. Hasil Pengumpulan Data

Berikut ini merupakan hasil pengumpulan data yang diperoleh dari mahasiswa perguruan tinggi negeri dan perguruan tinggi swasta, yang selanjutnya digunakan sebagai dasar analisis penelitian.

TABEL III
TABEL JUMLAH RESPONDEN

No	Nama Kampus	Jumlah Responden
1	Uin Jember	1
2	Stkip Moderen Ngawi	1
3	Uns	1
4	Uin Sunan Ampel	1
5	Uin 1 Tulungagung	2
6	Pkn Stan	1
7	Utr Cepu	1
8	Uin Maulana Malik Ibrahim	1
9	Sekolah T Ilmu Keperawatn	1
10	Uin Padjajaran	1
11	Iain Tuban	1
12	Upn Veteran	1

13	Uin Kalijaga Yogja	1
14	Institut Al Mujaddi	1
15	Itats	1
16	Universitas Trunojoyo M	1
17	Ut	2
18	Al-Muhammad Cepu	2
19	Institut Attanwir	3
20	Uin Walisongo	2
21	Ugm	1
22	Unesa	2
23	Ikip	7
24	Unugiri	73
Jumlah		118

C. Hasil Tahap Pre-processing

Tahap pre-processing dilakukan untuk menyiapkan data agar dapat diolah dalam proses klusterisasi. Pada penelitian ini, tahap praproses terbatas pada proses transformasi data, yaitu mengonversi setiap jawaban kuesioner ke dalam bentuk numerik menggunakan skala Likert, dengan ketentuan Sangat Tidak Setuju bernilai 1, Tidak Setuju bernilai 2, Netral bernilai 3, Setuju bernilai 4, dan Sangat Setuju bernilai 5. Proses konversi ini bertujuan untuk mengubah data kualitatif menjadi data numerik tanpa mengubah makna respons responden, sehingga data dapat digunakan sebagai input dalam algoritma klusterisasi *K-Means* dan *K-Medoids*.

TABEL IV
DATA SEBELUM KONVERSI

No	Nama	Prodi	Asal Kampus	V1	V2	...	V30
1	Fauziah Dewi Alata	PIAUD	UNUGIRI	Tidak Setuju	setuju	...	Tidak Setuju
2	Mohamad Yusuf Arifin	PJKR	UNUGIRI	Sangat Tidak Setuju	Tidak Setuju	...	Tidak Setuju
...
118	Muhammad Azizul Hakim	HI	UPN	Sangat Tidak Setuju	Sangat Tidak Setuju	...	Sangat Tidak Setuju

TABEL V
DATA SESUDAH KONVERSI

No	Nama	Prodi	Asal Kampus	V1	V2	...	V30
1	Fauziah Dewi Alata	PIAUD	UNUGIRI	2	4	...	2
2	Mohamad Yusuf Arifin	PJKR	UNUGIRI	1	2	...	2
...
118	Muhammad Azizul Hakim	HI	UPN	1	1	...	1

D. Hasil Proses Klusterisasi

Proses klusterisasi pada penelitian ini dilakukan dengan menggunakan empat kombinasi antara algoritma dan metrik

jarak, yaitu *K-Means* dengan Euclidean Distance, *K-Means* dengan Manhattan Distance, *K-Medoids* dengan Euclidean Distance, serta *K-Medoids* dengan Manhattan Distance. Seluruh pengujian dilakukan dengan menetapkan jumlah kluster $K = 2$, yang merepresentasikan dua kecenderungan kepribadian utama, yaitu introvert dan ekstrovert.

Jumlah kluster ditetapkan sebesar $K = 2$ berdasarkan konstruk teoretis Eysenck yang membagi dimensi kepribadian ke dalam spektrum introversi dan ekstrovert. Oleh karena itu, jumlah kluster tidak ditentukan melalui metode eksploratif seperti Elbow Method atau Silhouette Optimization, melainkan ditetapkan secara konseptual (theory-driven clustering). Pendekatan ini dipilih karena penelitian bertujuan menguji kesesuaian hasil klusterisasi dengan kerangka psikologis yang telah ditetapkan sebelumnya.

Dataset yang digunakan dalam proses klusterisasi merupakan data yang telah melalui tahap pre-processing berupa transformasi respons kuesioner ke dalam bentuk numerik. Seluruh atribut berada pada skala pengukuran yang sama, sehingga data dapat langsung digunakan dalam perhitungan jarak pada setiap kombinasi algoritma dan metrik jarak yang diterapkan. Dengan kondisi tersebut, proses klusterisasi dapat dilakukan secara konsisten untuk mengidentifikasi pola kecenderungan kepribadian responden secara objektif.

1) *K-Means* dengan Euclidean Distance

Hasil klusterisasi menggunakan metode *K-Means* dengan *Euclidean Distance* menghasilkan nilai rata-rata Silhouette Coefficient sebesar 0,44013, yang termasuk dalam kategori *moderate structure*. Nilai ini menunjukkan bahwa kluster yang terbentuk memiliki tingkat pemisahan yang cukup baik serta konsistensi pengelompokan yang relatif stabil. Sebelum proses klusterisasi, data terlebih dahulu melalui satu tahap filtrasi untuk memastikan kesesuaian dan relevansi data.

TABEL VI
DATA HASIL *K-MEANS* DENGAN *EUCLIDEAN DISTANCE*

C1	C2	CLUSTER	a(i)	b(i)	s(i)
6,595453	4,626974	C2	4,626974	6,595453	0,29846
7,516648	4,930055	C2	4,930055	7,516648	0,344115
6,892024	4,055043	C2	4,055043	6,892024	0,411633
6,892024	4,205324	C2	4,205324	6,892024	0,389827
...
8,803408	5,001233	C2	5,001233	8,803408	0,431898
Rata-rata Silhouette					0,44013

Hasil perhitungan menunjukkan jarak masing-masing data terhadap pusat kluster (C1 dan C2), nilai a(i) sebagai jarak rata-rata intra-kluster, serta b(i) sebagai jarak rata-rata terdekat terhadap kluster lain. Sebagian besar nilai s(i) bernilai positif, yang mengindikasikan bahwa mayoritas data lebih dekat dengan kluster asalnya dibandingkan dengan kluster lainnya. Secara umum, metode *K-Means* dengan

Euclidean Distance mampu membentuk dua kluster yang cukup terpisah dan sesuai dengan karakteristik data.

2) *K-Means dengan Manhattan Distance*

Hasil klusterisasi menggunakan metode *K-Means* dengan *Manhattan Distance* dilakukan setelah proses filtrasi data sebanyak 18 kali dan menghasilkan nilai rata-rata *Silhouette Coefficient* sebesar -0,20081, yang termasuk dalam kategori *weak structure*. Nilai ini menunjukkan bahwa pemisahan antar kluster yang terbentuk relatif rendah dan terdapat kedekatan yang cukup tinggi antar kluster.

TABEL VII
DATA HASIL *K-MEANS* DENGAN *MANHATTAN DISTANCE*

C1	C2	CLUSTER	a(i)	b(i)	s(i)
21	20	C2	20	21	-0,04762
16	23	C1	16	23	0,304348
17	18	C1	17	18	0,055556
15	18	C1	15	18	0,166667
....
32	17	C2	17	32	-0,46875
Rata-rata Silhouette					-0,20081

Hasil perhitungan menunjukkan bahwa jarak data terhadap pusat kluster C1 dan C2 tidak memiliki perbedaan yang signifikan. Selisih antara nilai a(i) dan b(i) pada sebagian besar data relatif kecil, sehingga menghasilkan nilai s(i) yang mendekati nol. Kondisi ini mengindikasikan bahwa banyak data berada pada area perbatasan antar kluster, yang mencerminkan adanya tumpang tindih karakteristik. Secara keseluruhan, kluster yang dihasilkan dengan *Manhattan Distance* belum menunjukkan struktur pemisahan yang jelas dan memiliki tingkat keterpisahan yang rendah.

3) *K-Medoids dengan Euclidean Distance*

Hasil klusterisasi menggunakan metode *K-Medoids* dengan *Euclidean Distance*, yang dilakukan setelah satu kali proses filtrasi data, menghasilkan nilai rata-rata *Silhouette Coefficient* sebesar 0. Nilai *Silhouette* yang bernilai nol menunjukkan bahwa sebagian besar data berada pada area perbatasan antar kluster, sehingga struktur kluster yang terbentuk belum menunjukkan pemisahan yang optimal.

TABEL VIII
DATA HASIL *K-MEDOIDS* DENGAN *EUCLIDEAN DISTANCE*

C1	C2	CLUSTER	a(i)	b(i)	s(i)
9,327379	7,416198	C2	7,416198	0	0
10,29563	7,745967	C2	7,745967	0	0
9,219544	7,416198	C2	7,416198	0	0
8,774964	6,403124	C2	6,403124	0	0
....
9,055385	7,745967	C2	7,745967	0	0
Rata-rata Silhouette					0

Hasil perhitungan menunjukkan bahwa nilai a(i) pada sebagian besar data cenderung tidak lebih kecil dibandingkan nilai b(i), sehingga menghasilkan nilai s(i) yang mendekati

nol. Selain itu, jarak data terhadap pusat kluster terdekat tidak menunjukkan perbedaan yang signifikan, yang mengindikasikan adanya tumpang tindih karakteristik antar kluster. Secara keseluruhan, metode *K-Medoids* dengan *Euclidean Distance* belum mampu membentuk struktur kluster yang tegas dan optimal berdasarkan evaluasi *Silhouette*.

4) *K-Medoids dengan Manhattan Distance*

Hasil klusterisasi menggunakan metode *K-Medoids* dengan *Manhattan Distance*, yang dilakukan setelah satu kali proses filtrasi data, menghasilkan nilai rata-rata *Silhouette Coefficient* sebesar 0,3632, yang termasuk dalam kategori *weak to moderate structure*. Nilai *Silhouette* yang bernilai positif menunjukkan bahwa sebagian besar data memiliki kedekatan yang lebih tinggi terhadap kluster asalnya dibandingkan dengan kluster lain, meskipun tingkat pemisahannya belum optimal.

TABEL IX
DATA HASIL *K-MEDOIDS* DENGAN *MANHATTAN DISTANCE*

C1	C2	CLUSTER	a(i)	b(i)	s(i)
41	31	C2	30,7619	51	0,39683
44	36	C2	30,7619	51	0,39683
39	33	C2	30,7619	51	0,39683
37	27	C2	30,7619	51	0,39683
....
38	36	C2	30,7619	51	0,39683
Rata-rata Silhouette					0,3632

Hasil perhitungan menunjukkan bahwa nilai a(i) dan b(i) pada sebagian besar data tidak memiliki perbedaan yang signifikan, sehingga nilai s(i) berada pada kisaran rendah. Kondisi ini mengindikasikan bahwa masih terdapat tumpang tindih karakteristik antar kluster dan batas pemisahan kluster yang kurang tegas. Secara keseluruhan, metode *K-Medoids* dengan *Manhattan Distance* mampu membentuk kluster dengan tingkat pemisahan yang cukup, namun belum menunjukkan struktur kluster yang optimal berdasarkan evaluasi *Silhouette*.

5) *Evaluasi Kualitas Kluster Menggunakan Silhouette Score*

Evaluasi kualitas kluster pada penelitian ini dilakukan menggunakan *Silhouette Score*, yaitu metrik yang mengukur tingkat kesesuaian penempatan setiap objek dalam kluster. Nilai *Silhouette* berada pada rentang -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa data memiliki kesesuaian yang tinggi dengan kluster asalnya dan terpisah dengan baik dari kluster lain, nilai mendekati 0 menunjukkan adanya tumpang tindih antar kluster (*overlapping*), sedangkan nilai negatif mengindikasikan bahwa suatu objek lebih dekat dengan kluster lain dibandingkan kluster tempatnya berada, sehingga termasuk dalam kategori salah kluster (*misclassified*).

TABEL X
PERBANDINGAN NILAI SILHOUETTE SCORE

No	Metode	Silhouette Score	Interpretasi
1	<i>K-Means</i> dengan <i>Euclidean</i>	0,44013	Struktur kluster cukup baik
2	<i>K-Means</i> dengan <i>Manhattan</i>	-0,20081	Struktur kluster sangat lemah dan tidak stabil
3	<i>K-Medoids</i> dengan <i>Euclidean</i>	0	Struktur kluster lemah dan menunjukkan tumpang tindih tinggi antar kluster
4	<i>K-Medoids</i> dengan <i>Manhattan</i>	0,3632	struktur kluster lemah hingga sedang.

Berdasarkan hasil perbandingan yang dilakukan, algoritma *K-Means* dengan metrik jarak *Euclidean* menghasilkan nilai *Silhouette Score* tertinggi sebesar 0,44013, yang menunjukkan bahwa struktur kluster yang terbentuk memiliki tingkat kohesi intra-kluster yang cukup baik serta pemisahan antar-kluster yang relatif lebih jelas dibandingkan metode lainnya. Berdasarkan interpretasi umum *Silhouette Score* menurut Rousseeuw (1987)[30], nilai di atas 0,50 menunjukkan struktur kluster yang kuat, nilai antara 0,26–0,50 menunjukkan struktur sedang (moderate), sedangkan nilai di bawah 0,25 menunjukkan struktur lemah. Dengan demikian, nilai 0,44013 dapat dikategorikan sebagai struktur kluster sedang, yang mengindikasikan masih adanya tumpang tindih karakteristik kepribadian antar responden. Temuan ini sejalan dengan konsep psikologis bahwa dimensi introversi–ekstrovert bersifat kontinu, sehingga batas antar kategori kepribadian tidak sepenuhnya bersifat diskrit atau kategorikal murni.

Hasil penelitian ini juga sejalan dengan temuan [10] dan [13] yang menunjukkan bahwa *K-Means* cenderung memberikan performa lebih baik dibandingkan *K-Medoids* pada dataset dengan *distribusi* relatif homogen dan skala atribut yang seragam. Kondisi ini mendukung bahwa data berbasis skala Likert dengan rentang yang konsisten lebih sesuai diproses menggunakan pendekatan berbasis centroid seperti *K-Means*, yang mampu menangkap kecenderungan pusat data secara lebih stabil dibandingkan metode berbasis medoid.

Sebaliknya, penerapan algoritma *K-Means* dengan *Manhattan Distance* menghasilkan nilai *Silhouette* yang rendah hingga bernilai negatif, yang mengindikasikan struktur kluster yang lemah serta tingginya tingkat tumpang tindih antar-kluster. Sementara itu, kedua skenario klusterisasi menggunakan *algoritma K-Medoids* menghasilkan nilai *Silhouette* yang rendah dan mendekati nol, yang menunjukkan bahwa kluster yang terbentuk belum memiliki pemisahan yang optimal dan masih terdapat banyak data yang berada pada area perbatasan kluster. Secara keseluruhan, hasil evaluasi ini menegaskan bahwa pemilihan algoritma dan metrik jarak berpengaruh terhadap kualitas klusterisasi data, dengan kombinasi *K-Means* dan *Euclidean Distance* menunjukkan performa paling optimal pada penelitian ini.

E. Ground Truth Penelitian

Ground truth dalam penelitian ini ditetapkan berdasarkan penilaian ahli di bidang psikologi kepribadian, yaitu Khoirotus Silfiah, M.Psi., Psikolog, sebagai acuan untuk *menevaluasi* hasil klusterisasi. Penilaian ahli dilakukan berdasarkan hasil pengisian kuesioner kepribadian yang telah melalui proses validasi isi dan revisi instrumen, sehingga layak digunakan untuk mengukur kecenderungan kepribadian introvert dan ekstrovert.

Penentuan ground truth dilakukan dengan mengelompokkan responden ke dalam kategori introvert atau ekstrovert berdasarkan dominasi skor jawaban pada indikator *kepribadian* tertentu. Kategori ini selanjutnya digunakan sebagai label pembanding dalam evaluasi eksternal hasil klusterisasi. Dengan adanya ground truth dari ahli, evaluasi klusterisasi dapat dilakukan secara lebih komprehensif, tidak hanya berdasarkan metrik internal, tetapi juga berdasarkan kesesuaian hasil kluster dengan konsep kepribadian secara psikologis.

1) Hasil Validasi Kluster Introvert dan Ekstrovert

Validasi menggunakan ground truth ahli menunjukkan bahwa kluster introvert yang dihasilkan oleh algoritma *K-Means* dengan metrik jarak *Euclidean* memiliki tingkat *kesesuaian* sebesar 100%, yang mengindikasikan bahwa karakteristik kepribadian introvert relatif homogen dan dapat dikenali secara konsisten. Sebaliknya, kluster ekstrovert menunjukkan tingkat ketidaksesuaian yang lebih tinggi, yang mencerminkan sifat karakteristik ekstrovert yang lebih heterogen dan tumpang tindih, sehingga sulit dipisahkan secara tegas menggunakan pendekatan klusterisasi tanpa supervisi.

TABEL XI
VALIDASI CLUSTER INTROVERT

No	Nama	Prodi	Asal Kampus	Cluster	Validasi
1	Dina Nur Ariyanti	Hukum Ekonomi Syari'ah	Unugiri	Introvert	Introvert
2	Mohammad Ilham Setiawan	TI	Unugiri	Introvert	Introvert

TABEL XII
VALIDASI CLUSTER EKSTROVERT

No	Nama	Prodi	Asal Kampus	Cluster	Validasi
1	Fauziah Dewi Alata	Pendidikan anak usia dini	Unugiri	Ekstrovert	Introvert
2	Mohamad Yusuf Arifin	Pendidikan jasmani kesehatan dan rekreasi	Unugiri Bojonegoro	Ekstrovert	Ekstrovert
3	Adinda mulkhatul fahmy	Pendidikan islam anak usia dini	Unugiri	Ekstrovert	Introvert

4	M.Syahru I Iman	Pendidikan Agama Islam	Unugiri	Ekstrovert	Ekstrovert
...
11 6	Siti Nurfaizah	TI	Unugiri	Ekstrovert	Introvert

Namun, pada kluster ekstrovert, dari total 116 data yang dikelompokkan sebagai ekstrovert, terdapat 30 data yang tidak sesuai dengan label ahli. Ketidaksesuaian yang relatif tinggi pada kluster ekstrovert dapat dijelaskan oleh sifat konstruktif ekstrovert yang lebih heterogen dan situasional. Berbeda dengan introvert yang cenderung stabil dan reflektif, karakteristik ekstrovert dapat muncul dalam berbagai bentuk ekspresi sosial yang tidak selalu konsisten dalam seluruh konteks, sehingga menyebabkan tumpang tindih antar kluster.

Secara keseluruhan, hasil validasi ini menunjukkan bahwa algoritma *K-Means* dengan *Euclidean Distance* memiliki performa yang sangat baik dalam mengidentifikasi kluster introvert, namun masih kurang optimal dalam membedakan kluster ekstrovert, sehingga diperlukan pendekatan lanjutan atau penambahan metode evaluasi untuk meningkatkan akurasi klusterisasi pada kelompok tersebut.

2) Analisis Tingkat Kesesuaian Hasil Klusterisasi

Berdasarkan hasil validasi terhadap ground truth, diperoleh perbedaan tingkat kesesuaian antara kluster introvert dan ekstrovert. Kluster introvert menunjukkan tingkat kesesuaian yang sangat tinggi, yaitu 100%, yang menandakan bahwa algoritma *K-Means* dengan *Euclidean Distance* mampu mengenali pola karakteristik kepribadian introvert secara konsisten. Hal ini mengindikasikan bahwa ciri-ciri introvert cenderung lebih homogen dan mudah dibedakan berdasarkan atribut yang digunakan dalam penelitian.

Sebaliknya, pada kluster ekstrovert ditemukan tingkat ketidaksesuaian yang cukup tinggi, yaitu 30 dari 116 data. Kondisi ini menunjukkan bahwa karakteristik kepribadian ekstrovert memiliki variasi yang lebih luas dan cenderung tumpang tindih dengan karakteristik introvert. Selain itu, perbedaan ini juga mencerminkan keterbatasan metode klusterisasi dalam menangkap kompleksitas kepribadian manusia yang bersifat kontinu dan tidak selalu dapat dipisahkan secara tegas.

Secara keseluruhan, hasil ini menunjukkan bahwa meskipun algoritma *K-Means* dengan *Euclidean Distance* efektif dalam mengidentifikasi kluster introvert, diperlukan pendekatan tambahan atau kombinasi metode lain untuk meningkatkan ketepatan klasifikasi pada kluster ekstrovert.

Secara praktis, model clustering yang dihasilkan dapat diintegrasikan dalam sistem pendukung keputusan berbasis web atau aplikasi sekolah, di mana siswa mengisi kuesioner kepribadian dan sistem secara otomatis mengelompokkan kecenderungan kepribadian sebagai dasar rekomendasi program studi. Integrasi ini memungkinkan konselor pendidikan memperoleh alat bantu objektif dalam proses advis akademik.

IV. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa proses klusterisasi kepribadian calon mahasiswa menggunakan algoritma *K-Means* dan *K-Medoids* dengan berbagai metrik jarak menghasilkan kualitas kluster yang berbeda. Hasil evaluasi menggunakan *Silhouette Score* menunjukkan bahwa algoritma *K-Means* dengan *Euclidean Distance* memberikan nilai tertinggi sebesar 0,44013, sehingga dapat dikatakan sebagai metode terbaik dalam penelitian ini dibandingkan kombinasi metode lainnya.

Evaluasi eksternal menggunakan *ground truth* yang ditetapkan oleh ahli psikologi menunjukkan bahwa kluster introvert yang dihasilkan oleh algoritma *K-Means* dengan *Euclidean Distance* memiliki tingkat kesesuaian 100% dengan label ahli, yang menandakan bahwa algoritma mampu mengidentifikasi karakteristik kepribadian introvert secara konsisten. Namun, pada kluster ekstrovert masih ditemukan tingkat ketidaksesuaian yang cukup tinggi, yaitu 30 dari 116 data, yang mengindikasikan adanya tumpang tindih karakteristik antara kepribadian introvert dan ekstrovert serta keterbatasan algoritma klusterisasi dalam merepresentasikan kompleksitas kepribadian manusia.

Secara keseluruhan, penelitian ini menunjukkan bahwa pemilihan algoritma dan metrik jarak berpengaruh signifikan terhadap kualitas klusterisasi data kepribadian. Algoritma *K-Means* dengan *Euclidean Distance* dinilai paling optimal secara matematis dan cukup relevan secara psikologis untuk mengelompokkan kecenderungan kepribadian calon mahasiswa, meskipun masih terdapat ruang untuk peningkatan pada identifikasi kluster ekstrovert.

Temuan ini menekankan urgensi integrasi kerangka teori psikologi dengan teknik klusterisasi komputasional guna meningkatkan daya interpretasi dan relevansi praktis sistem pendukung keputusan akademik berbasis data.

Penelitian selanjutnya disarankan untuk menggunakan jumlah kluster yang lebih bervariasi atau pendekatan kepribadian yang lebih kompleks, seperti kategori ambivert, guna menangkap karakteristik kepribadian secara lebih komprehensif dan tidak terbatas pada dikotomi introvert-ekstrovert. Disarankan untuk menggabungkan metode klusterisasi dengan metode klasifikasi atau algoritma lain guna meningkatkan akurasi pengelompokan, khususnya pada kluster ekstrovert yang menunjukkan tingkat ketidaksesuaian lebih tinggi.

DAFTAR PUSTAKA

- [1] S. Suharno, F. Lailaturrohman, P. Purwanto, R. Ranto, and M. Akhyar, "Analysis of students with the wrong major based on the metacognitive dimension," *Jurnal Pendidikan Teknologi dan Kejuruan*, vol. 29, no. 1, pp. 71–85, May 2023, doi: 10.21831/jptk.v29i1.53642.
- [2] K. Sylaska and J. D. Mayer, "Major Choices: Students' Personal Intelligence, Considerations When Choosing a Major, and Academic Success," *J. Intell.*, vol. 12, no. 11, Nov. 2024, doi: 10.3390/jintelligence12110115.
- [3] Z. Amin, B. Burhanuddin, T. Fajar Shadiq, and A. Soleh Purba, "How The Choice of Academic Majors and Students' Future

- Achievements According to The Talent Path,” *Nazhruna: Jurnal Pendidikan Islam*, vol. 4, no. 3, pp. 672–684, Nov. 2021, doi: 10.31538/nzh.v4i3.1676.
- [4] H. Jumareng *et al.*, “Introvert and extrovert personality: Is it correlated with academic achievement of Physical Education, Health and Recreation students at university level?,” *Hasanuddin Jumareng*, vol. 6, no. 2, pp. 140–146, 2021, doi: 10.25299/sportarea.2021.vol6(2).6172.
- [5] A. Rahmadani and Y. R. Mukti, “Adaptasi akademik, sosial, personal, dan institutional: studi college adjustment terhadap mahasiswa tingkat pertama,” *Jurnal Konseling dan Pendidikan*, vol. 8, no. 3, p. 159, Oct. 2020, doi: 10.29210/145700.
- [6] A. N. Hadi, Z. Abidin, A. Info, and R. Artikel, “Overview of Anxiety and Coping Methods of High School Students Facing State University Selection Gambaran Kecemasan dan Metode Coping Siswa SMA Menghadapi Seleksi Perguruan Tinggi Negeri,” *Jurnal Imiah Psikologi*, vol. 13, pp. 128–135, 2025, doi: 10.30872/psikoborneo.v13i1.
- [7] M. Ulfah Siregar, “Comparative Study of K-Means Clustering Algorithm and K-Medoids Clustering in Student Data Clustering,” *MEI*, 2022.
- [8] A. R. Pratama, R. Rizky Aryanto, A. Taufiq, M. Pratama, and P. Korespondensi, “Model Klasifikasi Calon Mahasiswa Baru Untuk Sistem Rekomendasi Program Studi Sarjana Berbasis Machine Learning,” vol. 9, no. 4, 2022, doi: 10.25126/jtiik.202294311.
- [9] M. D. Doi, A. Rusgiyono, and T. Wuryandari, “Analisis K-Medoids Dengan Validasi Indeks Pada Ipm Daerah 3t Di Indonesia,” *Jurnal Gaussian*, vol. 12, no. 2, pp. 178–188, Jul. 2023, doi: 10.14710/j.gauss.12.2.178-188.
- [10] D. Dwi Aulia and N. Nurahman, “Comparison Performance of K-Medoids and K-Means Algorithms In Clustering Community Education Levels,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 12, no. 2, pp. 273–282, Jul. 2023, doi: 10.23887/janapati.v12i2.59789.
- [11] H. Lestari Siregar, R. Hidayanthi, A. Langga Dewa Sakti, kan Tapanuli Selatan, J. Sutan Moh Arif, and N. Sumatera, “Implementation of K-Means clustering on student learning achievements based on social economic and social related,” 2024, doi: 10.22219/raden.v4i2.3.
- [12] D. Maryono, C. W. Budiyo, and A. A. Pamungkas, “Implementing of K-Means Clustering for Optimization of Student Grouping Based on Index of Learning Styles in Programming Classes,” *IJIE (Indonesian Journal of Informatics Education)*, vol. 6, no. 2, p. 84, Dec. 2022, doi: 10.20961/ijie.v6i2.68151.
- [13] Yosia and B. Siregar, “Comparative Analysis of K-Means and K-Medoids Algorithms for Product Sales Clustering and Customer,” *Journal of Mathematics, Computations and Statistics*, vol. 7, no. 2, pp. 360–370, Oct. 2024, doi: 10.35580/jmathcos.v7i2.4053.
- [14] N. A. Rizki, K. Kurniawan, I. K. Hasan, and N. Sampe, “Implementasi Algoritma K-Means Untuk Mengelompokkan Mahasiswa Berdasarkan Sumber Belajarnya,” *METIK JURNAL*, vol. 7, no. 2, pp. 62–67, Dec. 2023, doi: 10.47002/metik.v7i2.584.
- [15] H. Mulyani, R. A. Setiawan, and H. Fathi, “Optimization of K Value in Clustering Using Silhouette Score (Case Study: Mall Customers Data),” *Journal of Information Technology and Its Utilization*, vol. 6, no. 2, pp. 45–50, Dec. 2023, doi: 10.56873/jitu.6.2.5243.
- [16] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, “On the Surprising Behavior of Distance Metrics in High Dimensional Space.” [Online]. Available: <http://kops.uni-konstanz.de/volltexte/2009/7007>
- [17] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” 1987.
- [18] N. Izzah and A. Rahman As, “Kreativitas matematis mahasiswa bertipe kepribadian ekstrovert-introvert dalam menyelesaikan masalah geometri,” 2022. [Online]. Available: <http://journal2.um.ac.id/index.php/jkpm>
- [19] M. A. Rohmat and Kusriani, “Penerapan Metode Analytical Hierarchy Process (AHP) Dalam Sistem Pendukung Keputusan Penilaian Kinerja Guru,” *METIK JURNAL*, vol. 5, no. 1, pp. 55–62, Jun. 2021, doi: 10.47002/metik.v5i1.217.
- [20] S. N. Wardah, N. Nurjanah, and D. Suryadi, “Systematic Literature Review: Analisis Tipe Kepribadian Ekstrovert dan Introvert Terhadap Kemampuan Matematis Siswa,” *Indiktika : Jurnal Inovasi Pendidikan Matematika*, vol. 6, no. 2, pp. 294–306, Jun. 2024, doi: 10.31851/indiktika.v6i2.15395.
- [21] A. Dri Hananto, A. Measy Erfiana, B. Lexiani Permata Putri, P. Dwi Putri, and F. Kurniawan, “Algoritma Machine Learning Naive Bayes pada Analisis Sentimen Kesepakatan Polri dan GNPF-MUI pada Aksi Bela Islam III ‘212’ Naive Bayes Machine Learning Algorithm on Sentiment Analysis of Police and GNPF-MUI Agreement on ‘212’ Islamic Defense Action III,” *Technology and Agriculture Journal*, vol. 4, no. 2, pp. 151–160, 2023, doi: 10.37638/sinta.4.2.151-16.
- [22] W. Junthopas and C. Wongoutong, “Pre-Determining the Optimal Number of Clusters for k-Means Clustering Using the Parameters Package in R and Distance Metrics,” *Applied Sciences (Switzerland)*, vol. 15, no. 21, Nov. 2025, doi: 10.3390/app152111372.
- [23] T. Hardiani, “Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 2, pp. 156–165, Aug. 2022, doi: 10.23887/janapati.v11i2.45376.
- [24] B. Laksono, Y. Syahidin, and Y. Yunengsih, “Implementasi Data Mining Klasterisasi Data Pasien Rawat Inap dengan Algoritma K-Means Clustering,” *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 2, pp. 621–627, Apr. 2024, doi: 10.32493/jtsi.v7i2.39354.
- [25] J. Homepage, D. Kurmiati, M. Zakiy Fauzi, and A. Falegas, “MALCOM: Indonesian Journal of Machine Learning and Computer Science Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm Klasterisasi Daerah Rawan Gempa Bumi di Indonesia Menggunakan Algoritma K-Medoids,” vol. 1, pp. 47–57, 2021.
- [26] M. Minarni, E. I. Sari, A. Syahrani, and P. Mandarani, “Klasterisasi Penyakit Menggunakan Algoritma K-Medoids pada Dinas Kesehatan Kabupaten Agam,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 10, no. 3, p. 137, Dec. 2021, doi: 10.23887/janapati.v10i3.34904.
- [27] T. Prompook *et al.*, “Impact of distance measures in adaptive K-means clustering on load profiles and spatial patterns of distributed substations in Thailand,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-07475-8.
- [28] R. Stepanus Ginting, H. Hamdani, A. Septiariani, and F. Alameka, “The Clustering Tindak Kekerasan Dalam Rumah Tangga Di Kota Samarinda Menggunakan Algoritma K-Means,” *METIK JURNAL*, vol. 6, no. 2, pp. 172–177, Dec. 2022, doi: 10.47002/metik.v6i2.378.
- [29] D. Zahro Putri, R. De Pani, S. Ginting, and L. Efrizoni, “Penerapan Algoritma K-Medoids dalam Menganalisis Pola Pelanggan untuk Strategi Pemasaran,” 2025. [Online]. Available: <http://jurnal.mdp.ac.id>
- [30] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” 1987.