

Penerapan Algoritma K-Means Untuk Mengelompokkan Tingkat Stres Akademik Pada Mahasiswa

Lusi Diah Wiranti¹, Elvia Budianita^{1,*}, Alwis Nazir¹, Fitri Insani¹, Reni Susanti²

¹Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

² Fakultas Psikologi, Prodi Psikologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹12150121581@students.uin-suska.ac.id, ^{2,*}elvia.budianita@uin-suska.ac.id, ³alwis.nazir@uin-suska.ac.id,

⁴fitri.insani@uin-suska.ac.id, ⁵reni.susanti@uin-suska.ac.id,

Email Penulis Korespondensi: elvia.budianita@uin-suska.ac.id

Submitted: 20/05/2025; Accepted: 13/06/2025; Published: 13/06/2025

Abstrak—Stres akademik merupakan permasalahan yang sering dialami oleh mahasiswa akibat berbagai tekanan akademik dalam lingkungan pendidikan tinggi. Tekanan-tekanan tersebut mencakup banyaknya tugas yang harus diselesaikan dalam waktu singkat, ekspektasi tinggi dari dosen maupun orang tua, ketidakmampuan dalam mengelola waktu secara efektif, serta persepsi negatif terhadap kemampuan diri sendiri. Apabila stres ini tidak cepat ditangani, dapat berdampak pada penurunan prestasi belajar dan kesehatan mental mahasiswa. Studi ini bertujuan bertujuan untuk mengelompokkan mahasiswa sesuai dengan tingkat stres akademik mereka melalui penerapan algoritma *K-Means*. Data penelitian diperoleh dari 507 responden melalui kuesioner *Perception of Academic Stress Scale* (PASS) yang telah dimodifikasi sesuai dengan konteks penelitian. Data kemudian diproses melalui tahap *preprocessing* dan transformasi ke dalam bentuk numerik agar dapat dianalisis lebih lanjut. Implementasi algoritma *K-Means* dilakukan dengan menggunakan bahasa pemrograman *Python* melalui platform *Google Colab*. Evaluasi hasil *clustering* dilakukan dengan menggunakan *Davies-Bouldin Index* (DBI) dan *Silhouette Coefficient*. Semakin rendah nilai DBI, semakin baik tingkat kekompakan dan keterpisahan antar cluster, sedangkan *Silhouette Coefficient* yang lebih tinggi menunjukkan kualitas pemisahan *cluster* yang lebih optimal. Berdasarkan hasil pengujian, diperoleh jumlah *cluster* terbaik adalah 2, dengan nilai DBI sebesar 1.43 dan *Silhouette Coefficient* sebesar 0.27. Namun, hasil tersebut masih belum optimal dikarenakan karakteristik data responden yang berasal dari lima jurusan berbeda di Fakultas Sains dan Teknologi, serta ketidakseimbangan jumlah pengisi kuesioner dari masing-masing angkatan 2021-2023. *Cluster* 1 sebanyak 229 mahasiswa berada pada kelompok tingkat stres rendah, ditunjukkan oleh skor kuesioner yang cenderung rendah. Sebaliknya, *cluster* 2 sebanyak 278 mahasiswa dengan tingkat stres tinggi, yang ditandai skor tinggi (rentang 3-5) pada pernyataan *favorable*.

Kata Kunci: Stres Akademik; Mahasiswa; Clustering; K-Means; Davies-Bouldin Index; Silhouette Coefficient

Abstract—Academic stress is a prevalent concern among university students, often arising from various challenges within the academic environment. These challenges may include tight assignment deadlines, elevated expectations from both lecturers and parents, ineffective time management, and negative self-assessment. If left unaddressed, such stress can negatively impact students' academic performance and mental well-being. This study focuses on categorizing student academic stress levels using the K-Means clustering algorithm. Data were collected from 507 participants through a customized version of the Perception of Academic Stress Scale (PASS) questionnaire, adapted to suit the study context. Prior to analysis, the data were preprocessed and converted into a numerical format. Clustering was performed using Python on the Google Colab platform. To assess the clustering performance, two evaluation metrics were used: the Davies-Bouldin Index (DBI) and the Silhouette Coefficient. Lower DBI values suggest that the clusters formed are more compact and distinct from each other, while higher Silhouette values indicate better clustering performance. From the evaluation, the best clustering result was found when the number of clusters was 2, with a DBI score of 1.43 and a Silhouette score of 0.27. Nonetheless, these values still fall short of the ideal range, likely due to the heterogeneous nature of the data, as participants came from five different departments within the Faculty of Science and Technology. Moreover, the number of responses varied across academic years (2021–2023). Cluster 1 comprised 229 students identified as having low levels of academic stress, as shown by their lower questionnaire scores. In contrast, Cluster 2 consisted of 278 students with higher levels of stress, as reflected in their higher scores (ranging from 3 to 5) on positively worded items.

Keywords: Academic Stress; Students; Clustering; K-Means; Davies-Bouldin Index; Silhouette Coefficient

1. PENDAHULUAN

Mahasiswa merupakan individu yang berada pada tahap pendidikan tinggi, dengan rentang usia umumnya antara 18 hingga 25 tahun, yang memiliki tanggung jawab akademik maupun non-akademik. [1] Selama masa studi, mahasiswa tidak hanya menjalani aktivitas akademik seperti perkuliahan dan tugas-tugas, tetapi juga kegiatan non-akademik yang mendukung pengembangan diri. Dalam menjalankan peran tersebut, mahasiswa dihadapkan pada berbagai tuntutan, mulai dari tekanan untuk menyelesaikan studi tepat waktu, ekspektasi dari pihak kampus, harapan orang tua, hingga motivasi pribadi untuk meraih gelar sarjana [2]. Tuntutan-tuntutan tersebut, jika tidak dikelola dengan baik, dapat memicu timbulnya stres.

Stres akademik adalah tekanan mental yang sering dirasakan oleh mahasiswa di lingkungan pendidikan, dan berpotensi memberikan pengaruh buruk terhadap perkembangan pribadi [3]. Menurut Bedewy & Gabriel, stres akademik dalam *Perception of Academic Stress Scale* (PASS) mencakup tiga dimensi utama, yaitu ekspektasi akademik, tuntutan akademik, dan persepsi diri akademik. Ketiganya berkaitan dengan harapan tinggi, beban tugas yang berat, serta pandangan negatif terhadap kemampuan diri. [4]. Jika tidak dikelola dengan baik, stres akademik dapat menimbulkan gangguan fisik dan psikologis. Namun dalam batas wajar, stres akademik juga dapat berdampak

positif, seperti meningkatkan motivasi dan kreativitas. Dengan demikian, penting bagi mahasiswa untuk memahami sumber stres akademik dan mengelola stres tersebut agar tidak menimbulkan dampak yang merugikan [5].

Beberapa data nyata menunjukkan bahwa banyak mahasiswa mengalami stres berat. Penelitian dari Mental Health Foundation di Inggris menemukan bahwa 60% orang berusia 18-24 tahun mengalami stres tinggi. Riset Kesehatan Dasar mencatat bahwa hampir 10% populasi Indonesia dengan usia di atas 15 tahun menunjukkan gejala gangguan emosional [6]. Penelitian lain di Universitas Hasanuddin menunjukkan bahwa 90% mahasiswa tahun pertama mengalami stres berat, disusul 89% pada mahasiswa tahun kedua dan ketiga [7]. Data-data ini menunjukkan bahwa stres akademik adalah permasalahan serius yang perlu diperhatikan, terutama di lingkungan perguruan tinggi.

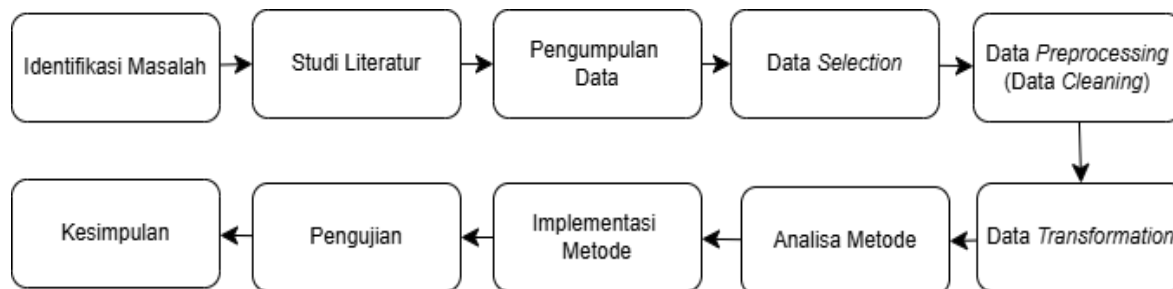
Untuk memahami dan menangani permasalahan stres akademik tersebut secara lebih sistematis, pendekatan yang tepat perlu diterapkan. Salah satunya adalah teknik data *mining*, yaitu proses untuk mengidentifikasi pola, tren, serta informasi penting dari kumpulan data berjumlah besar, sehingga dapat menghasilkan informasi yang relevan guna mendukung pengambilan keputusan secara lebih efektif [8]. Salah satu teknik yang umum diterapkan dalam data *mining* adalah *clustering*, yaitu metode pengelompokan data berdasarkan kemiripan karakteristik tertentu [9]. Metode ini mampu menyederhanakan data kompleks menjadi kelompok-kelompok yang lebih terstruktur, sehingga mendukung pemahaman dan terhadap pengambilan keputusan yang lebih baik terhadap data tersebut. Di antara berbagai metode *clustering*, algoritma *K-Means* banyak dikenal karena kemudahan penerapannya dalam membagi data berdasarkan pusat titik (*centroid*) [10].

Sejumlah penelitian telah berhasil memanfaatkan algoritma *K-Means* untuk mengelompokkan data. Solang dan Nugroho mengelompokkan data stres mahasiswa Universitas Kristen Satya Wacana menjadi tiga kategori: tertekan, sejahtera, harmonis [11]. Penelitian lain oleh Sembiring *et al.* menggunakan *K-Means* untuk mengelompokkan prestasi siswa SD menjadi tiga kelompok: “Sangat Layak”, “Layak Dipertimbangkan”, dan “Tidak Layak” [12]. Penelitian oleh Norshahlan *et al.* juga menggunakan metode serupa untuk membagi data calon siswa baru menjadi dua kelompok, yaitu 47 siswa dalam *cluster* pertama dan 23 siswa dalam *cluster* kedua [13]. Adapun studi dari Wijaya & Januaviani menggunakan gabungan algoritma *K-Means* dan *K-Modes* untuk mengelompokkan faktor stres pada mahasiswa aktif. Hasil analisis metode *K-Means* dengan *cluster* terbesar 41,036% data, yang mayoritas dipicu oleh faktor perkuliahan, individual, dan finansial. Sementara itu, *K-Modes* menunjukkan *cluster* terbesar 36,65% lebih menyoroti faktor perkuliahan dan sosial [14]. Gultom *et al.* juga menggunakan *K-Means* untuk menganalisis tingkat kriminalitas di wilayah Pematang Siantar. Hasil analisis pengelompokan data menunjukkan dua kategori, kriminalitas tinggi dan rendah [15]. Kemudian penelitian oleh Rahmawati & Wijayanto yang menganalisis *cluster* indeks pembangunan manusia dengan menggunakan tiga algoritma yaitu *K-Means*, *Fuzzy C-Means*, dan *Hierarchical Clustering*. Penelitian tersebut menyimpulkan bahwa *K-Means* memberikan hasil pengelompokan yang lebih optimal [16]. Selain itu, penelitian oleh Luchia *et al.* membandingkan *K-Means* dengan *K-Medoids* dalam pembagian data kemiskinan menggunakan *Davies Bouldin Index* (DBI), dimana *K-Means* menunjukkan kinerja lebih baik dengan DBI 0.041 pada $k=8$ dibandingkan *K-Medoids* dengan DBI 0.052 pada $k=2$ [17].

Berdasarkan temuan-temuan tersebut, terbukti bahwa algoritma *K-Means* berhasil dalam pengelompokan data. Namun, penelitian yang memanfaatkan algoritma *K-Means* pada populasi mahasiswa Universitas Islam Negeri Sultan Syarif Kasim Riau menggunakan kuesioner *Perception of Academic Stress Scale* (PASS) versi adaptasi Bahasa Indonesia yang dimodifikasi secara khusus masih sangat terbatas. Kombinasi unik antara populasi lokal dan instrumen yang telah disesuaikan secara kultural ini memberikan kontribusi baru dalam analisis stres akademik yang lebih relevan dan kontekstual. Data penelitian berasal dari mahasiswa Fakultas Sains dan Teknologi menggunakan kuesioner PASS hasil adaptasi Bahasa Indonesia Murdhiono & Vidayanti [18] dan modifikasi oleh psikolog ahli, Reni Susanti, S.Psi, M.Psi. Penggunaan algoritma *K-Means* diharapkan mampu menghasilkan pengelompokan sikap yang lebih tepat, sehingga dapat dijadikan acuan oleh pihak kampus dan lembaga pendidikan lainnya dalam merancang program dukungan untuk membantu mahasiswa mengelola stres akademik secara lebih optimal.

2. METODOLOGI PENELITIAN

Bagian metodologi memaparkan prosedur yang dirancang sebagai solusi terhadap permasalahan penelitian. Rangkaian tersebut disajikan pada Gambar 1 berikut.



Gambar 1. Metodologi Penelitian

2.1 Identifikasi Masalah

Langkah awal dalam proses penelitian adalah mengenali isu utama yang akan menjadi fokus kajian. Fokus utama dalam tahap ini adalah tingkat stres akademik yang dialami mahasiswa. Untuk menganalisis permasalahan tersebut, digunakan algoritma *K-Means* sebagai metode pengelompokan data berdasarkan tingkat stres yang diperoleh dari hasil pengisian kuesioner.

2.2 Studi Literatur

Studi literatur adalah tahapan yang melibatkan analisis mengenai permasalahan yang sedang dihadapi. Literatur pada penelitian ini mencakup beberapa aspek utama. Pertama, teori tentang stres akademik yang mencakup pengertian dan dampaknya terhadap mahasiswa. Kedua, literatur terkait algoritma *K-Means*, yang meliputi prinsip dasar, cara kerja, serta studi kasus pengelompokan data. Selain itu, literatur mengenai metodologi penelitian juga dipelajari, termasuk teknik pengumpulan data melalui survei yang menggunakan instrumen kuesioner untuk mengukur tingkat stres akademik.

2.3 Pengumpulan Data

Pengumpulan data adalah tahapan yang berfungsi untuk mendapatkan informasi terkait permasalahan yang sedang dikaji. Pada penelitian ini, data dikumpulkan melalui kuesioner tentang stres akademik berdasarkan instrumen *Perception of Academic Stress Scale* (PASS) yang dikembangkan oleh Dalia Bedewy dan Adel Gabriel-Adaptasi Bahasa Indonesia oleh Murdhiono dan Vidayanti [18], serta dimodifikasi oleh Reni Susanti, S.Psi, M.Psi, yang merupakan pakar di bidang psikologi, agar sesuai dengan konteks penelitian. Instrumen ini terdiri dari 18 pertanyaan yang disebarluaskan melalui *google form* kepada mahasiswa Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau angkatan 2021-2023. Setiap gejala stres akademik diuraikan dalam panduan PASS Tabel 1 berikut.

Tabel 1. PASS-Manual

No	Aspek	Item	Jumlah
1	Ekspektasi Akademik	6, 8, 12, 13	4
2	Tuntutan Akademik	4, 5, 9, 10, 11, 15, 16, 17	8
3	Persepsi Diri Akademik	1, 2, 3, 7, 14, 18	6
Total			18

2.4 Data Selection

Data *selection* merupakan tahapan memilih data sesuai kebutuhan yang digunakan dalam analisis. Pada studi ini, data yang dianalisis berasal dari responden yang telah mengisi kuesioner. Setiap entri data mencakup atribut awal antara lain nama, NIM, jenis kelamin, angkatan dan program studi, serta jawaban terhadap 18 item kuesioner yang digunakan untuk mengukur tingkat stres akademik. Instrumen ini menggunakan skala likert dengan lima opsi respon, yaitu: Sangat Tidak Setuju (STS), Tidak Setuju (TS), Ragu-Ragu (RR), Setuju (S), dan Sangat Setuju (SS) [18]. Skala likert terdapat dua bentuk pernyataan, yaitu *Favorable* (F) yang selaras dengan variabel penelitian dan *Unfavorable* (UF) yang berlawanan dengan variabel penelitian [19]. Seluruh item dalam kuesioner disusun untuk menggambarkan berbagai aspek stres akademik yang dialami oleh mahasiswa, dan disajikan pada Tabel 2 berikut.

Tabel 2. Kuesioner PASS

No	Pertanyaan	STS	TS	RR	S	SS
1	Saya yakin bahwa saya akan menjadi mahasiswa yang sukses (UF)					
2	Saya memiliki kepercayaan diri yang tinggi bahwa saya akan sukses dalam karir saya di masa depan (UF)					
3	Saya mudah mengambil keputusan terkait kegiatan akademik saya (UF)					
4	Proporsi waktu antara kuliah/bimbingan dengan penugasan yang diberikan menurut saya sudah seimbang (UF)					
5	Saya memiliki waktu yang cukup untuk beristirahat setelah menyelesaikan tugas-tugas saya (UF)					
6	Tuntutan dari dosen terhadap pencapaian akademik saya membuat saya tertekan (F)					
7	Saya takut mengalami kegagalan pada beberapa mata kuliah yang saya jalani pada semester ini (F)					
8	Dosen memiliki harapan yang lebih tinggi terhadap kinerja saya dibandingkan dengan kemampuan saya sendiri (F)					
9	Beban SKS dan praktik yang harus saya tempuh selama berkuliah menurut saya terlalu besar (F)					

No	Pertanyaan	STS	TS	RR	S	SS
10	Menurut saya, jumlah penugasan setiap mata kuliah terlalu banyak (F)					
11	Saya sering tidak dapat menyelesaikan tugas tepat waktu (F)					
12	Harapan orang tua terhadap pencapaian akademik yang lebih baik membuat saya tertekan (F)					
13	Ada persaingan yang ketat dengan teman-teman sekelas saya untuk mendapatkan nilai terbaik (F)					
14	Saya berpikir bahwa kekhawatiran saya akan ujian merupakan salah satu kelemahan saya (F)					
15	Soal-soal ujian yang diujikan biasanya sangat sulit (F)					
16	Menurut saya, alokasi waktu pengerjaan soal-soal ujian terlalu singkat (F)					
17	Saat ujian merupakan salah satu situasi yang menegangkan bagi saya (F)					
18	Bahkan jika saya lulus ujian, saya masih khawatir dengan kepastian mendapatkan pekerjaan (F)					

2.5 Data Preprocessing

Tahap data *preprocessing* berupa pembersihan pembersihan data untuk mengatur data agar siap digunakan dalam analisis lebih lanjut.

2.5.1 Data Cleaning

Pembersihan data merupakan tahap untuk menghilangkan *noise* pada dataset, seperti nilai kosong (*missing value*), data ganda, maupun kesalahan input lainnya. Dalam penelitian ini, pembersihan data dilakukan dengan menghapus data responden yang mengisi kuesioner lebih dari satu kali. Langkah ini diperlukan guna memastikan kualitas data sebelum dianalisis menggunakan algoritma *K-Means*.

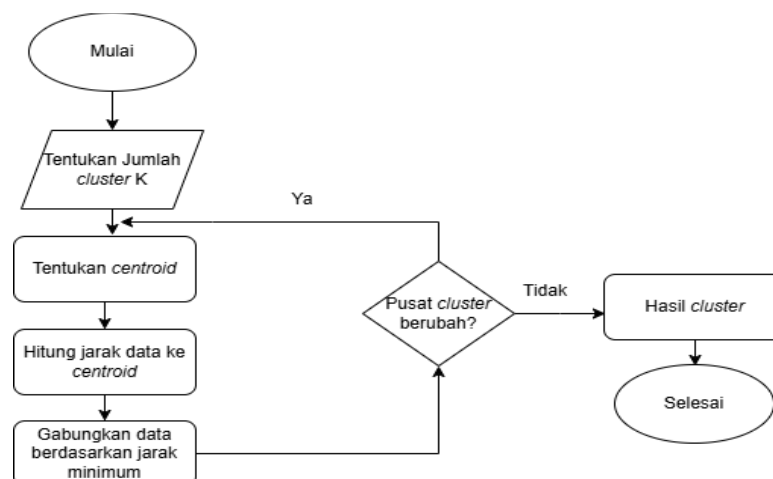
2.6 Data Transformation

Transformasi data merupakan proses mengonversi data mentah menjadi format yang tepat agar dapat diolah lebih lanjut. Proses transformasi data ini dilakukan konversi data yang awalnya berbentuk huruf menjadi angka.

2.7 Analisa Metode

K-Means adalah model *unsupervised learning* untuk mengolah data yang belum pernah diberi label atau klasifikasi sebelumnya. Model ini mengidentifikasi pola yang berulang dalam data dan memberkan respons yang konsisten terhadap pola-pola tersebut pada setiap data *point* [20]. Algoritma *K-Means* bekerja dengan menetapkan jumlah *cluster*, kemudian secara berulang-ulang memperbarui posisi titik pusat (*centroid*) dengan tujuan mengurangi variasi data di dalam setiap *cluster* [21].

Proses penerapan metode *K-Means* disajikan secara terstruktur melalui *flowchart* pada Gambar 2 berikut:



Gambar 2. Flowchart Metode *K-Means*

Berikut penjelasan langkahnya secara sistematis [22]:

- Tentukan jumlah *cluster* (*k*). Nilai *k* ditentukan berdasarkan kebutuhan analisis atau menggunakan metode evaluasi.
- Tentukan *centroid* awal secara acak.

- c. Hitung jarak data ke *centroid* menggunakan persamaan berikut:

$$D(x, y) = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2} \tag{1}$$

Dalam rumus tersebut, $D(x, y)$ menyatakan jarak antara dua titik, yaitu titik data dan *centroid*. Simbol x merepresentasikan data, sedangkan y menunjukkan *centroid*. X_1 dan X_2 adalah komponen dari titik data x , sedangkan Y_1 dan Y_2 adalah komponen dari titik *centroid* y .

- d. Mengelompokkan data ke *cluster* terdekat. Data akan digabungkan ke dalam *cluster* yang memiliki jarak paling dekat dengan setiap *centroid* yang bersangkutan.
- e. Setelah pengelompokan dilakukan, sistem akan menghitung ulang posisi *centroid* baru untuk setiap *cluster* menggunakan persamaan berikut:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \tag{2}$$

Dimana μ_k merupakan *centroid cluster* ke- k , N_k merepresentasikan total data dalam *cluster* tersebut, dan x_q merupakan entri data ke- q yang termasuk dalam *cluster* k .

- f. Jika pusat *cluster* mengalami perubahan, yaitu setelah dilakukan pembaruan *centroid* dan posisi *centroid* baru berbeda dengan sebelumnya, maka algoritma akan terus mengulang proses perhitungan jarak antara seluruh data dengan *centroid* yang baru. Selanjutnya, data akan dikelompokkan ulang ke dalam *cluster* yang paling dekat dengan *centroid* tersebut. Proses ini akan terus diulang hingga posisi *centroid* tidak lagi berubah.
- g. Setelah proses iterasi berhenti, hasil akhir berupa pembagian data ke dalam k *cluster* yang terbentuk.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan pemaparan mengenai hasil dari penerapan metode yang telah dipaparkan pada bagian sebelumnya serta pembahasan yang berkaitan dengan proses dan evaluasi dalam penelitian. Fokus utama pada bab ini adalah pemanfaatan algoritma *K-Means* untuk melakukan pengelompokan tingkat stres akademik mahasiswa berdasarkan data yang dikumpulkan melalui kuesioner. Setiap tahapan yang dilakukan dalam penelitian dijabarkan secara sistematis, mulai dari analisis kebutuhan data, pemilihan data, preprocessing, transformasi data, hingga implementasi dan pengujian algoritma.

3.1 Analisis Kebutuhan Data

Analisis kebutuhan data pada penelitian ini bertujuan untuk menentukan jenis dan sumber data yang dibutuhkan guna mendukung proses pengelompokan tingkat stres akademik pada mahasiswa. Data yang digunakan diperoleh dengan menyebarkan kuesioner kepada para mahasiswa sebagai responden. Kuesioner tersebut dirancang untuk mengukur berbagai indikator yang berkaitan dengan stres akademik, seperti beban tugas, tekanan waktu, kecemasan terhadap nilai, serta kemampuan manajemen waktu. Dari hasil pengisian kuesioner tersebut, terkumpul sebanyak 510 data responden.

3.2 Data Selection

Tahapan data *selection* merupakan bagian penting dalam penelitian. Data diperoleh dari jawaban pertanyaan pada kuesioner PASS yang telah diisi oleh responden. Kuesioner tersebut terdiri dari 18 pernyataan, dimana setiap pertanyaan mencakup aspek stres akademik yang dialami oleh responden. Tabel 3 memperlihatkan data yang menjadi dasar dalam pelaksanaan penelitian.

Tabel 3. Data Selection

No	Q1 (UF)	Q2 (UF)	Q3 (UF)	Q4 (UF)	Q5 (UF)	Q6 (F)	Q7 (F)	Q8 (F)	Q18 (F)
1	Sangat Setuju	Sangat Setuju	Setuju	Setuju	Sangat Setuju	Tidak Setuju	Setuju	Ragu-Ragu	Sangat Setuju
2	Sangat Setuju	Sangat Setuju	Setuju	Setuju	Setuju	Ragu-Ragu	Setuju	Ragu-Ragu	Ragu-Ragu
3	Sangat Setuju	Sangat Setuju	Setuju	Setuju	Sangat Setuju	Setuju	Sangat Setuju	Setuju	Sangat Setuju
4	Sangat Setuju	Sangat Setuju	Sangat Setuju	Setuju	Setuju	Setuju	Ragu-Ragu	Tidak Setuju	Ragu-Ragu
5	Setuju	Setuju	Tidak Setuju	Setuju	Ragu-Ragu	Ragu-Ragu	Sangat Setuju	Ragu-Ragu	Ragu-Ragu
...
506	Sangat Setuju	Setuju	Setuju	Setuju	Ragu-Ragu	Tidak Setuju	Ragu-Ragu	Tidak Setuju	Tidak Setuju



No	Q1 (UF)	Q2 (UF)	Q3 (UF)	Q4 (UF)	Q5 (UF)	Q6 (F)	Q7 (F)	Q8 (F)	Q18 (F)
507	Setuju	Ragu- Ragu	Ragu- Ragu	Tidak Setuju	Tidak Setuju	Setuju	Ragu- Ragu	Setuju	Ragu- Ragu
508	Ragu- Ragu	Ragu- Ragu	Ragu- Ragu	Tidak Setuju	Tidak Setuju	Ragu- Ragu	Ragu- Ragu	Setuju	Ragu- Ragu
509	Setuju	Setuju	Ragu- Ragu	Ragu- Ragu	Tidak Setuju	Ragu- Ragu	Tidak Setuju	Tidak Setuju	Ragu- Ragu
510	Setuju	Setuju	Setuju	Setuju	Ragu- Ragu	Tidak Setuju	Tidak Setuju	Tidak Setuju	Tidak Setuju

3.3 Data Preprocessing

Pada penelitian ini proses data *preprocessing* meliputi data *cleaning*.

3.3.1 Data Cleaning

Dalam konteks penelitian ini, data *cleaning* dilakukan untuk menghilangkan duplikasi dan kesalahan pengisian pada data yang diperoleh dari kuesioner. Data dikumpulkan dari pengisian kuesioner oleh mahasiswa. Setelah data terkumpul, dilakukan pemeriksaan terhadap kemungkinan adanya duplikasi. Hasil pemeriksaan menunjukkan bahwa terdapat tiga data duplikat, sehingga jumlah data awal sebanyak 510 dikurangi menjadi 507 data valid, yang terdiri dari 157 mahasiswa angkatan 2021, 229 angkatan 2022, dan 121 angkatan 2023. Dengan melakukan proses data *cleaning* ini, dataset menjadi lebih bersih dan siap digunakan untuk proses analisis *clustering*.

3.4 Data Transformation

Proses ini melibatkan perubahan data menjadi bentuk numerik. Transformasi ini memastikan bahwa seluruh data bersifat kuantitatif dan seragam, sehingga memudahkan algoritma dalam menghitung jarak antar data serta dalam proses pembentukan *cluster*. Setiap item dalam kuesioner dikategorikan sebagai *Favorable* (F) atau *Unfavorable* (UF). Untuk pernyataan *Favorable*, skor diberikan dari 1 (Sangat Tidak Setuju) hingga 5 (Sangat Setuju). Sebaliknya, untuk *Unfavorable*, skala dibalik, yakni dari 5 (Sangat Tidak Setuju) hingga 1 (Sangat Setuju). Konversi ini ditampilkan pada Tabel 4.

Tabel 4. Bobot Nilai

Kuesioner	Keterangan	Skor F	Skor UF
STS	Sangat Tidak Setuju	1	5
TS	Tidak Setuju	2	4
RR	Ragu-Ragu	3	3
S	Setuju	4	2
SS	Sangat Setuju	5	1

Setelah proses transformasi data dilakukan berdasarkan skema penskoran pada Tabel 4, setiap respons kualitatif dari kuesioner PASS diubah menjadi nilai numerik sesuai dengan kategori *Favorable* (F) dan *Unfavorable* (UF). Tabel 5 berikut menampilkan sebagian hasil data yang telah dikonversi ke dalam bentuk numerik dan siap digunakan sebagai input dalam algoritma *K-Means*.

Tabel 5. Data Setelah di Transformasi

No	Q1 (UF)	Q2 (UF)	Q3 (UF)	Q4 (UF)	Q5 (UF)	Q6 (F)	Q7 (F)	Q8 (F)	Q18 (F)
1	1	1	2	2	1	2	4	3	5
2	1	1	2	2	2	3	4	3	3
3	1	1	2	2	1	4	5	4	5
4	1	1	1	2	2	4	3	2	3
5	2	2	4	2	3	3	5	3	3
...
503	1	2	2	2	3	2	2	2	2
504	2	3	3	4	4	4	3	4	3
505	3	3	3	4	4	3	3	4	3
506	2	2	3	3	4	3	2	2	3
507	2	2	2	2	3	2	2	2	2

3.5 Implementasi Metode K-Means

Algoritma *K-Means* diimplementasikan dengan memanfaatkan *Python* sebagai *platform* pemrograman, karena menyediakan berbagai *library* yang mendukung pengolahan data dan *machine learning* secara efisien. Implementasi

dilakukan dengan memanfaatkan sejumlah *library* antara lain *numpy*, *pandas*, *matplotlib*, *sklearn.cluster*, *sklearn.decomposition* dan *sklearn.metrics*. Dataset yang digunakan diunggah dalam format excel dan dibaca menggunakan *pandas* untuk diproses lebih lanjut.

Centroid awal ditetapkan pada data ke-4 dan ke-6 sebagai dasar awal proses pengelompokan. Algoritma *K-Means* dijalankan dan dilakukan iterasi hingga tercapai konvergensi, yaitu ketika perubahan posisi *centroid* tidak lagi signifikan. Hasil pengelompokan ditambahkan ke dalam data dalam bentuk label *cluster*. Tabel 6 menunjukkan sebagian data yang telah dilengkapi dengan label hasil pengelompokan menggunakan algoritma *K-Means*.

Tabel 6. Hasil Pengelompokan *Cluster*

No	Q1 (UF)	Q2 (UF)	Q3 (UF)	Q4 (UF)	Q5 (UF)	Q6 (F)	Q7 (F)	Q8 (F)	Q18 (F)	Cluster_ KMEANS
1	1	1	2	2	1	2	4	3	5	2
2	1	1	2	2	2	3	4	3	3	2
3	1	1	2	2	1	4	5	4	5	2
4	1	1	1	2	2	4	3	2	3	1
5	2	2	4	2	3	3	5	3	3	2
...
503	1	2	2	2	3	2	2	2	2	1
504	2	3	3	4	4	4	3	4	3	2
505	3	3	3	4	4	3	3	4	3	2
506	2	2	3	3	4	3	2	2	3	1
507	2	2	2	2	3	2	2	2	2	1

Tabel 7. Jumlah Data Setiap *Cluster*

Cluster	Jumlah Data
1	229
2	278

Hasil clustering pada Tabel 7 menghasilkan dua kelompok utama. Cluster 1 terdiri dari 229 mahasiswa yang tergolong dalam kategori stres akademik rendah. Hal ini didasarkan pada rata-rata skor total dari 18 item pernyataan hasil pengisian kuesioner setelah proses transformasi yang cenderung rendah, mencerminkan bahwa mahasiswa dalam kelompok ini tidak terlalu merasakan tekanan dalam kegiatan akademiknya. Sementara itu, cluster 2 mencakup 278 mahasiswa yang termasuk dalam kategori stres akademik tinggi. Kelompok ini menunjukkan rata-rata skor kuesioner yang lebih tinggi, mengindikasikan adanya beban psikologis yang lebih signifikan dalam menjalani aktivitas perkuliahan. Skor tinggi tersebut terutama terlihat pada pernyataan nomor 12 dan 13 kuesioner yang terdapat pada Tabel 2, dengan rentang skor berkisar antara 3 hingga 5. Tingginya skor pada dua pernyataan ini menunjukkan bahwa tekanan dari harapan eksternal (orang tua) dan kompetisi akademik internal antar rekan merupakan faktor dominan yang berkontribusi terhadap peningkatan stres. Hal ini mengindikasikan pentingnya perhatian terhadap faktor sosial dan lingkungan akademik dalam upaya mengurangi stres mahasiswa.

3.6 Pengujian

Setelah implementasi algoritma *K-Means*, langkah selanjutnya adalah melakukan pengujian untuk mengevaluasi seberapa baik hasil *clustering* yang diperoleh. Pengujian ini dilakukan dengan menggunakan beberapa metrik evaluasi yang umum digunakan dalam analisis *clustering*, seperti *Davies-Bouldin Index (DBI)* dan *Silhouette Coefficient*.

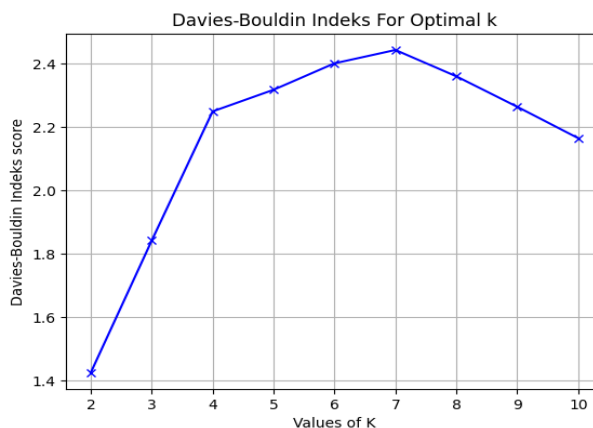
- a. *Davies-Bouldin Index (DBI)* merupakan indikator evaluasi dalam analisis *cluster* yang bertujuan untuk menilai sejauh mana masing-masing *cluster* memiliki batas pemisah yang jelas [23]. DBI mengevaluasi hasil pengelompokan dengan memperhatikan tingkat kekompakan dan jarak setiap *cluster* dengan *cluster* terdekatnya [24]. DBI memiliki hubungan positif terhadap kasus “*within class*” dan hubungan negatif terhadap kasus “*between class*”. DBI digunakan untuk menilai hasil *clustering* karena bersifat umum dan cocok digunakan dalam proses evaluasi. Proses evaluasi ini terbagi menjadi dua jenis utama, yaitu validasi internal dan validasi eksternal yang digunakan untuk melihat seberapa baik hasil pengelompokan data [25]. Berikut hasil pengujian menggunakan DBI:

Tabel 8. Hasil Pengujian Menggunakan DBI

Cluster	Nilai DBI
2	1.43
3	1.84
4	2.25
5	2.32
6	2.40
7	2.44
8	2.36

Cluster	Nilai DBI
9	2.26
10	2.17

Pengujian dilakukan dengan variasi jumlah *cluster* dari 2 hingga 10. Berdasarkan hasil evaluasi, nilai DBI terendah diperoleh saat $k = 2$, yaitu sebesar 1.43. Hal ini menunjukkan bahwa pengelompokan data ke dalam dua *cluster* menghasilkan struktur *cluster* yang paling optimal dibandingkan jumlah *cluster* lainnya. Karena semakin rendah nilai DBI, semakin baik hasil pengelompokan. Berikut hasil visualisasi pengujian menggunakan DBI:



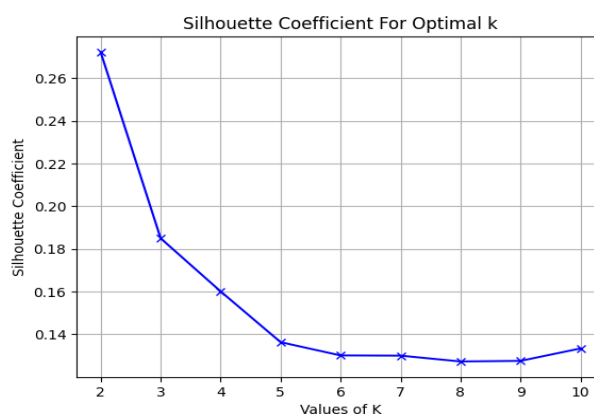
Gambar 3. Visualisasi Hasil Pengujian Menggunakan DBI

- b. *Silhouette Coefficient* merupakan metrik evaluasi yang digunakan untuk menilai tingkat kesamaan antar *cluster*. Semakin mendekati angka 1 pada *Silhouette Coefficient*, maka semakin tinggi kualitas pengelompokan yang telah dilakukan [26]. Metrik ini merupakan kombinasi dari pendekatan *cohesion* dan *separation* [27].

Tabel 9. Hasil Pengujian *Silhouette Coefficient*

Cluster	Nilai Silhouette
2	0.27
3	0.19
4	0.16
5	0.14
6	0.13
7	0.13
8	0.13
9	0.13
10	0.13

Hasil pengujian menggunakan metode *Silhouette Coefficient* untuk jumlah *cluster* antara 2 hingga 10 menunjukkan bahwa nilai tertinggi diperoleh pada jumlah *cluster* 2, yaitu sebesar 0.27. Nilai tersebut mengindikasikan bahwa pengelompokan data paling optimal terjadi saat data dibagi menjadi dua *cluster*. Sementara itu, nilai *Silhouette* cenderung menurun seiring bertambahnya jumlah *cluster*, yang menunjukkan bahwa kualitas pengelompokan menjadi kurang optimal. Visualisasi hasil pengujian menggunakan *Silhouette Coefficient* disajikan sebagai berikut:



Gambar 4. Visualisasi Hasil Pengujian Menggunakan *Silhouette Coefficient*

Berdasarkan hasil pengujian menggunakan *Davies-Bouldin Index* (DBI) dan *Silhouette Coefficient*, pengelompokan paling optimal terjadi saat data dibagi menjadi dua *cluster*. Nilai DBI terendah sebesar 1.43 dan nilai *Silhouette* tertinggi sebesar 0.27 sama-sama diperoleh pada jumlah *cluster* 2, yang menunjukkan bahwa konfigurasi ini memberikan hasil terbaik dibandingkan jumlah *cluster* lainnya. Namun demikian, nilai-nilai tersebut belum cukup optimal dikarenakan karakteristik data responden yang berasal dari lima jurusan berbeda di Fakultas Sains dan Teknologi, serta ketidakseimbangan jumlah pengisi kuesioner dari masing-masing angkatan 2021-2023.

4. KESIMPULAN

Penelitian ini memanfaatkan algoritma *K-Means*, untuk mengelompokkan tingkat stres akademik mahasiswa Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Data diperoleh melalui kuesioner *Perception of Academic Stress Scale* (PASS) yang telah dimodifikasi, kemudian diproses hingga menghasilkan 507 data valid dari total 510 data responden. Tahapan penelitian yang dilakukan meliputi data *selection*, data *preprocessing*, data *transformation*, implementasi metode *K-Means*, dan evaluasi *clustering* dengan menggunakan metrik validasi. Berdasarkan hasil pengujian menggunakan *Davies-Bouldin Index* (DBI) dan *Silhouette Coefficient*, jumlah *cluster* divariasikan secara bertahap mulai dari 2 hingga 10. Hasil evaluasi menunjukkan bahwa jumlah *cluster* terbaik adalah 2. Namun demikian, hasil pengujian yang diperoleh belum menunjukkan pemisahan *cluster* yang kuat. Kondisi ini disebabkan oleh karakteristik data responden yang berasal dari lima jurusan berbeda di Fakultas Sains dan Teknologi, serta ketidaseimbangan jumlah pengisi kuesioner dari masing-masing angkatan 2021 hingga 2023. *Cluster* 1 terdiri dari 229 mahasiswa yang tergolong dalam kategori stres akademik rendah. Kategori ini ditentukan berdasarkan rata-rata skor total 18 item pernyataan hasil pengisian kuesioner yang cenderung rendah, mencerminkan bahwa para responden dalam *cluster* ini tidak terlalu terbebani oleh tekanan akademik. Sementara itu, *cluster* 2 terdiri dari 278 mahasiswa dalam kategori stres akademik tinggi, dengan skor rata-rata kuesioner yang lebih tinggi, yang mencerminkan persepsi tekanan akademik yang lebih besar pada mahasiswa dalam kelompok ini. Dengan demikian, temuan ini dapat dijadikan dasar oleh pihak kampus untuk merancang program dukungan yang lebih tepat sasaran, terutama bagi mahasiswa yang mengalami stres akademik tinggi. Penelitian lanjutan disarankan untuk mengelompokkan tingkat stres akademik berdasarkan program studi.

REFERENCES

- [1] F. Andiarna and E. Kusumawati, "Pengaruh Pembelajaran Daring terhadap Stres Akademik Mahasiswa Selama Pandemi Covid-19 Funsu Andiarna, Estri Kusumawati," *Jurnal Psikologi*, vol. 16, no. 2, pp. 139–150, Dec. 2020, doi: 10.24014/jp.v14i2.9221.
- [2] M. D. Nurmala, T. U. S. H. Wibowo, and A. Rachmayani, "Tingkat Stres Mahasiswa Dalam Pembelajaran Online Pada Masa Pandemi Covid-19," *Jurnal Penelitian Bimbingan dan Konseling*, vol. 5, no. 2, pp. 13–23, Dec. 2020, doi: 10.30870/jpbk.v5i2.10108
- [3] A. Fiqih and V. Ratnawati, "Mengurai Stres Akademik Mahasiswa Tingkat Akhir: Faktor Pemicu, Dampak Dan Strategi Pengelolaan Di Universitas Nusantara PGRI Kediri," *Semdikjar*, vol. 6, pp. 755–765, Aug. 2023.
- [4] D. K. Dewi, S. I. Savira, Y. W. Satwik, and R. N. Khoirunnisa, "Profil Perceived Academic Stress pada Mahasiswa Profile of Perceived Academic Stress in Students," *Jurnal Psikologi Teori dan Terapan*, vol. 13, no. 3, pp. 395–402, 2022, doi: 10.26740/jptt.v13n3.p395-403
- [5] N. M. Yusuf and Yusuf Jannatul Ma'wa, "Faktor-Faktor Yang Mempengaruhi Stres Akademik," *Psyche 165 Journal*, vol. 13, pp. 235–239, Jun. 2020, <https://doi.org/10.35134/jpsy165.v13i2.84>
- [6] D. Damayanti, E. A. Trisus, E. Yunanti, B. L. Ingrid, and T. Panjaitan, "Hubungan Tingkat Stres dengan Siklus Menstruasi Mahasiswi," *Jurnal Kedokteran dan Kesehatan*, vol. 18, pp. 212–219, Jul. 2022, doi: 10.32536/jrki.v7i2.260
- [7] I. Rosyidah, A. R. Efendi, Muh. Am. Arfah, P. A. Jasman, and N. Pratami, "Gambaran Tingkat Stres Akademik Mahasiswa Program Studi Ilmu Keperawatan Fakultas Keperawatan Unhas," *Jurnal Abdi*, vol. 2, no. 1, pp. 33–39, Jan. 2020.
- [8] S. Syam, Y. Tokoro, L. Judijanto, M. Garonga, M. F. Sinaga, and N. Umar, "Data Mining : Teori dan Penerapannya dalam Berbagai Bidang," PT. Sonpedia Publishing Indonesia, Jambi, 2024.
- [9] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [10] D. Haryadi, "Penerapan Algoritma K-Means Clustering Pada Produksi Perkebunan Kelapa Sawit Menurut Provinsi," *Journal of Informatics and Communication Technology (JICT)*, vol. 3, no. 1, pp. 50–64, Jul. 2021, doi: 10.52661/j_ict.v3i1.71.
- [11] T. Solang and A. Nugroho, "Analisis Kesehatan Mental Mahasiswa Universitas Kristen Satya Wacana Menggunakan Metode Clustering Algoritma K-Means," *Jurnal TEKINKOM*, vol. 6, no. 1, 2023, doi: 10.37600/tekinkom.v6i1.641.
- [12] S. Natalia Br Sembiring, H. Winata, S. Kusnasari, S. Informasi, and S. Triguna Dharma, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *Jurnal Sistem Informasi Tgd*, vol. 1, pp. 31–40, Jan. 2022, <https://doi.org/10.53513/jursi.v1i1.4784>
- [13] M. Norshahlan, H. Jaya, and R. Kustini, "Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Calon Siswa Baru," *Jurnal Sistem Informasi TGD*, vol. 2, no. 6, 2023, <https://doi.org/10.30743/INFOTEKJAR.V1I2.70>
- [14] J. Wijaya, T. Magdalena, A. Januavian, and K. Kunci, "Clustering Faktor Stres Pada Mahasiswa Aktif Menggunakan Algoritma K-Means Dan K-Modes," *Multidisciplinary Scientific Journal*, vol. 2, pp. 907–917, Feb. 2024, <https://doi.org/10.57185/mutiara.v2i2.137>



- [15] D. Gultom *et al.*, “Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar,” *Jurnal Teknologi Informasi*, vol. 4, no. 1, 2020, doi: 10.36294/jurti.v4i1.1263
- [16] R. Rahmawati and W. A. Wijayanto, “Analisis Cluster Dengan Algoritma K-Means, Fuzzy C-Means Dan Hierarchical Clustering (Studi Kasus: Indeks Pembangunan Manusia tahun 2019),” *Jurnal Informatika dan Komputer*, vol. 5, no. 2, p. 73, Feb. 2021, doi: 10.26798/jiko.v5i2.422
- [17] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. F. Octavia, “Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, pp. 35–41, Sep. 2022, doi: 10.57152/malcom.v2i2.422.
- [18] W. R. Murdhiono and V. Vidayanti, “Examining Academic Stress and Its Source Among Nursing Professional Students (Ners) Using the Modified Perception of Academic Stress Scale (PAS),” *Indonesian Nursing Journal Of Education And Clinic (Injec)*, vol. 7, no. 1, p. 2, Jun. 2022, doi: 10.24990/injec.v7i1.441.
- [19] M. A. Hudin and S. M. Budiani, “Hubungan antara Workplace Well-Being dengan Kinerja Karyawan pada PT. X di Sidoarjo Hudin,” *Jurnal Penelitian Psikologi*, 2021, <https://doi.org/10.26740/cjpp.v8i4.41192>
- [20] M. Cui, “Introduction to the K-Means Clustering Algorithm Based on the Elbow Method,” *Accounting, Auditing and Finance*, vol. 1, pp. 5–8, Oct. 2020, doi: 10.23977/accaf.2020.010102.
- [21] P. W. Rahayu, I. G. I. Sudipa, Suryani, and A. Ridwan, *BUKU AJAR DATA MINING*. 2024. [Online]. Available: <https://www.researchgate.net/publication/377415198>
- [22] F. Amin, D. S. Anggraeni, and Q. Aini, “Penerapan Metode K-Means dalam Penjualan Produk Souq.Com,” *Applied Information System and Management (AISM)*, vol. 5, no. 1, pp. 7–14, Apr. 2022, doi: 10.15408/aism.v5i1.22534.
- [23] R. Ishak, “Clustering Prestasi Akademik Lulusan Menggunakan Metode K-Means,” *Jambura Journal of Electrical and Electronics Engineering*, vol. 6, pp. 76–81, Jan. 2024, <https://doi.org/10.37905/jjee.v6i1.23967>
- [24] R. K. Dinata, Bustami, and S. Retno, “Optimizing the Evaluation of K-means Clustering Using the Weight Product,” *Revue d’Intelligence Artificielle*, vol. 38, no. 4, pp. 1223–1233, Aug. 2024, doi: 10.18280/ria.380416.
- [25] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, “Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities,” *TEM Journal*, vol. 10, no. 3, pp. 1099–1103, Aug. 2021, doi: 10.18421/TEM103-13.
- [26] T. Rahmawati, Y. Wilandari, and P. Kartikasari, “Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids,” *Jurnal Gaussian*, vol. 13, no. 1, pp. 13–24, Aug. 2024, doi: 10.14710/j.gauss.13.1.13-24.
- [27] S. Paembonan, H. Abduh, and K. Kunci, “Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coefficient,” *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, vol. 6, no. 2, Sep. 2021, https://doi.org/10.51557/pt_jiit.v6i2.659