

# Penggunaan Model Klaster K-Means dan Klasifikasi KNN untuk Identifikasi Pengetahuan Matematika Mahasiswa

Hery Sunandar<sup>1</sup>

<sup>1</sup> Universitas Budi Darma, Jl. Sisingamangaraja No.338, Kota Medan, 20219, Sumatera Utara

## ARTICLE INFORMATION

Received: Februari, 2024

Revised: Maret, 2024

Available online: April, 2024

## KEYWORDS

Klasifikasi, KNN, Klaster, K-MEANS, Pengetahuan\_matematika.

## CORRESPONDENCE

Phone: 081397863774

E-mail: [herysun1975@gmail.com](mailto:herysun1975@gmail.com)

## A B S T R A C T

Students' mathematical knowledge is an important factor that needs to be considered in the learning process majoring in Informatics Engineering. Students who have low mathematical knowledge will definitely have difficulty calculating and applying the algorithms used to solve programming and computing problems. The weaknesses that these students have can become obstacles to the future. By knowing students' mathematical knowledge, the learning process does not have to be adjusted to the student's level of knowledge, but further action is needed to increase students' mathematical knowledge. One way to find out students' mathematical knowledge is to classify students' mathematical knowledge. Classification of students' mathematical knowledge can be done using a very popular method, namely the K-Nearest Neighbor (KNN) method. However, in order for classification to be carried out, grouping must be carried out first to form classes from the results of grouping mathematical knowledge data. The grouping of students' mathematical knowledge data was carried out using the K-MEANS method. Data on students' mathematical knowledge was taken based on course grades for calculus 1, applied physics, calculus 2, statistics and probability, and discrete mathematics. The grades from several of these courses will be processed using K-MEANS and form three groups of mathematical knowledge, namely slow, sufficient, and fast. Thus, it was concluded that to determine the classification of slow, sufficient, and fast mathematical knowledge of students majoring in informatics engineering, it was carried out using the KNN method, but first the data was grouped using the KMEAN method.

## PENDAHULUAN

Pengetahuan matematika merupakan aspek yang krusial dalam pembelajaran mahasiswa jurusan Teknik Informatika. Tingkat pemahaman matematika yang rendah dapat menjadi hambatan serius bagi mahasiswa dalam memahami konsep-konsep pemrograman dan algoritma yang menjadi dasar dalam bidang teknologi informasi [1], [2]. Mahasiswa yang memiliki pemahaman matematika yang kurang memadai sering mengalami kesulitan dalam menyelesaikan masalah pemrograman dan komputasi yang kompleks [3]. Ketidamampuan ini tidak hanya memengaruhi kinerja akademik mahasiswa saat ini, tetapi juga dapat menjadi penghambat bagi perkembangan dan kesuksesan di masa depan [4]. Mahasiswa yang menghadapi kesulitan dengan pengetahuan matematika yang rendah tidak hanya mengalami hambatan dalam proses pembelajaran, tetapi juga dalam menyerap materi kuliah, menyelesaikan tugas-tugas, dan mengikuti ujian selanjutnya [5]. Kesulitan ini sering kali mengarah pada penurunan kualitas kerja, penurunan nilai akademik dan bahkan kegagalan dalam mencapai standar yang diharapkan dalam mata kuliah terkait [6]. Hal ini dapat mempengaruhi motivasi dan percaya diri mahasiswa dalam menghadapi tantangan akademik, serta mengurangi minat dalam bidang studi tersebut [7].

Upaya untuk meningkatkan pengetahuan matematika mahasiswa menjadi tantangan untuk diatasi. Salah satu pendekatan yang dapat dilakukan adalah dengan melakukan klasifikasi terhadap pengetahuan matematika mahasiswa. Klasifikasi ini memungkinkan untuk mengelompokkan mahasiswa ke dalam kategori-kategori berdasarkan tingkat pemahaman. Salah satu metode yang umum digunakan untuk melakukan klasifikasi ini adalah metode K-Nearest Neighbor (KNN) [8], [9], yang memungkinkan untuk mengelompokkan mahasiswa berdasarkan kemiripan pengetahuan matematika dengan mahasiswa lain dalam dataset. Dengan demikian [10], [11], [12], metode KNN memberikan pendekatan yang efektif untuk mengidentifikasi dan memahami pola-pola dalam tingkat pengetahuan matematika mahasiswa, sehingga memungkinkan untuk merancang strategi pembelajaran yang lebih tepat dan efektif sesuai dengan kebutuhan individu masing-masing mahasiswa. Namun, sebelum melakukan klasifikasi dengan metode KNN, diperlukan langkah pengelompokkan (*Clustering*) terlebih dahulu untuk membentuk kelompok-kelompok pengetahuan matematika mahasiswa. Metode *Clustering* yang dapat digunakan adalah metode K-Means, yang dapat membantu dalam mengelompokkan data pengetahuan matematika mahasiswa menjadi kelompok-kelompok berdasarkan karakteristik yang dimiliki [13], [14], [15]. Selain itu teknik indentifikasi yang dianggap mampu menjadi pendukung mesin pembelajaran (*Machine Learning*). *Machine Learning* berupa kecerdasan buatan yang diterapkan ke komputer untuk belajar dari data yang diberikan [16], [17].

Penelitian Suraya dkk yang berkaitan dengan penggunaan *Clustering* K-Means melakukan pengelompokan data untuk menemukan pola atau hubungan antar data dengan konteks evaluasi data yang dilakukan dan dipilih dengan 3 kelompok yang dapat diberi label menjadi Varietas Unggul, Varietas Menengah, dan Varietas Baku [15]. Sedangkan penelitian Andi Bode dkk yang telah mengeksplorasi klasifikasi KNN dengan menerapkan algoritma KNN untuk melakukan prediksi tingkat kepuasan mahasiswa dalam perkuliahan online di Universitas Ichsan Gorontalo [10]. Studi yang terkait dengan Klasifikasi KNN telah menunjukkan keberhasilannya dalam pengujian yang dapat disimpulkan bahwa nilai ketetapan KNN pada klasifikasi tingkat kematangan buah tomat berpengaruh terhadap hasil akurasi, tetapi tidak terlalu signifikan [18].

Dalam mengelompokkan data ke dalam kelompok-kelompok yang berbeda berdasarkan pola-pola yang ada dalam dataset. Algoritma *Clustering* K-Means populer karena algoritmanya yang sederhana dengan konvergensi yang cepat. Nilai K dari *Clustering* perlu diberikan terlebih dahulu dan pemilihan nilai K secara langsung mempengaruhi hasil konvergensi menurut Yuan C [19]. Di bidang kedokteran, Wargijono Utomo menerapkan K-Means untuk mengelompokkan 4 atribut, antara lain kasus konfirmasi, pengobatan, kebutuhan, dan kasus kematian dalam penanganan COVID-19 [20]. Secara keseluruhan, penelitian terdahulu telah menunjukkan bahwa *Clustering* K-Means dan klasifikasi KNN merupakan dua metode yang kuat dan efektif untuk mengelompokkan

dan mengklasifikasikan data dalam berbagai konteks. Integrasi kedua metode ini dalam suatu kerangka kerja yang terstruktur dapat meningkatkan pemahaman dan penggunaan mereka dalam memecahkan masalah-masalah dunia nyata yang kompleks. Penerapan kedua model K-Means dan KNN oleh Mayangsari H, yaitu menggunakan algoritma K-Means untuk menghasilkan klasifikasi karakteristik berdasarkan kemampuan dari mahasiswa dan pendistribusian mahasiswa ke instansi KKL dihasilkan dari kluster K-Means dan di uji kembali menggunakan metode KNN [21].

Data pengetahuan matematika mahasiswa yang digunakan dalam penelitian ini berasal dari nilai-nilai yang diperoleh dari beberapa matakuliah matematika yang relevan. Matakuliah-matakuliah yang menjadi fokus inklusi dalam analisis ini adalah kalkulus 1, fisika terapan, kalkulus 2, statistik dan probabilitas, serta matematika diskrit. Data nilai dari setiap matakuliah ini akan dijadikan sebagai atribut atau fitur dalam analisis klasifikasi menggunakan metode KNN. Dengan memanfaatkan data nilai dari matakuliah-matakuliah tersebut, penelitian ini bertujuan untuk membentuk tiga kelompok pengetahuan matematika yang berbeda, yaitu kelompok lambat, cukup, dan cepat. Proses pembentukan kelompok-kelompok ini akan dilakukan dengan menggunakan metode K-MEANS, yang akan mengelompokkan mahasiswa berdasarkan kesamaan nilai-nilai dalam matakuliah-matakuliah tersebut. Dengan demikian, pembentukan tiga kelompok pengetahuan matematika ini akan memberikan landasan yang kuat bagi analisis lebih lanjut terkait dengan karakteristik dan performa akademik mahasiswa di jurusan Teknik Informatika.

Urgensi dari penelitian ini sangatlah penting mengingat beberapa faktor yang menjadi latar belakang dan masalah yang dihadapi. (1). Pengetahuan matematika merupakan landasan yang vital dalam pembelajaran jurusan Teknik Informatika. Mahasiswa yang memiliki pemahaman matematika yang baik cenderung lebih sukses dalam memahami konsep-konsep pemrograman dan algoritma yang kompleks. Sebaliknya, mahasiswa yang mengalami kesulitan dalam pemahaman matematika dapat menghambat kemajuan mereka dalam studi dan karier di bidang teknologi informasi. (2). Kesulitan dalam pengetahuan matematika mahasiswa dapat berdampak langsung pada proses pembelajaran. Mahasiswa yang mengalami kesulitan dalam memahami materi kuliah, menyelesaikan tugas, dan mengikuti ujian dapat mengalami penurunan nilai dan bahkan kegagalan dalam mata kuliah terkait. Hal ini tidak hanya mempengaruhi kinerja akademik mereka saat ini, tetapi juga masa depan akademik dan profesional mereka.

Dalam konteks ini, penelitian ini mendesak untuk memberikan pemahaman yang lebih baik tentang tingkat pengetahuan matematika mahasiswa jurusan Teknik Informatika. Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang lebih baik tentang distribusi pengetahuan matematika di antara mahasiswa, serta memberikan landasan bagi pengembangan strategi pembelajaran yang lebih adaptif dan efektif untuk mendukung keberhasilan akademik mereka. Dengan demikian, urgensi dari penelitian ini sangatlah penting dalam meningkatkan kualitas pendidikan dan kesuksesan mahasiswa jurusan Teknik Informatika.

## METODE PENELITIAN

Metodologi penelitian penggunaan model kluster K-Means dan klasifikasi KNN untuk identifikasi pengetahuan matematika mahasiswa:

1. Pengumpulan Data  
Data nilai dari matakuliah-matakuliah yang relevan, seperti kalkulus 1, fisika terapan, kalkulus 2, statistik dan probabilitas, serta matematika diskrit, akan dikumpulkan dari rekaman akademik mahasiswa jurusan Teknik Informatika.
2. Preprocessing Data  
Data akan diproses untuk menghilangkan data yang tidak lengkap atau tidak relevan dan normalisasi data akan dilakukan untuk memastikan bahwa setiap variabel memiliki rentang nilai yang seragam.
3. Pengelompokkan Data (*Clustering*) dengan Metode K-MEANS  
Metode K-MEANS akan diterapkan untuk mengelompokkan data pengetahuan matematika mahasiswa menjadi tiga kelompok: lambat, cukup, dan cepat.
4. Klasifikasi dengan Metode KNN  
Metode KNN akan digunakan untuk mengklasifikasikan tingkat pengetahuan matematika mahasiswa berdasarkan kelompok-kelompok yang telah terbentuk melalui K-MEANS.
5. Evaluasi Hasil  
Hasil klasifikasi akan dievaluasi untuk menilai keefektifan metode KNN dalam mengidentifikasi tingkat pengetahuan matematika mahasiswa.

## HASIL DAN PEMBAHASAN

### 3. Penyajian Data

Data-data mahasiswa yang digunakan dalam pengujian diambil dari data nilai dari matakuliah-matakuliah yang sebelumnya yang pernah diikuti oleh mahasiswa jurusan Teknik Informatika. Data tersebut juga harus relevan untuk tujuan identifikasi pengetahuan matematika mahasiswa. Adapun matakuliah-matakuliah yang relevan, yaitu kalkulus 1, fisika terapan, kalkulus 2, statistik dan probabilitas, serta matematika diskrit.

Tabel 1. Nilai mahasiswa yang uji

No.	Nama	Kalkulus 1 (Kal1)	Fisika Terapan (FisT)	Kalkulus 2 (Kal2)	Statistik dan Probalitas (StaP)	Matematika Diskrit (MatD)
1.	AAG	4	3	3	3	3
2.	ARS	3	4	3	4	3
3.	AS	4	3	2	3	3
4.	AP	2	2	3	3	2
5.	AZ	4	4	4	4	3
6.	CAS	4	3	3	3	3
7.	COP	3	3	3	3	3
8.	EDN	3	3	3	3	3

9.	ERS	3	4	3	4	4
10.	ET	2	3	2	2	4
11.	FRI	3	4	2	3	2
12.	FRE	2	2	2	3	3
13.	GAL	3	3	3	3	2
14.	GAN	3	3	3	3	3
15.	GIN	4	2	3	3	3
16.	HAR	4	4	3	3	3
17.	INA	3	4	3	4	4
18.	IND	2	3	3	3	2
19.	IST	3	3	3	2	3
20.	JEL	3	4	4	3	3
21.	JUD	4	3	3	4	4
22.	JUN	4	4	4	4	4
23.	LAN	3	3	4	3	2
24.	LIN	3	3	3	2	2
25.	LIS	4	4	3	4	4
26.	MAS	2	4	3	3	4
27.	MRS	3	3	3	2	3
28.	MRE	4	3	3	2	4
29.	MTT	4	3	3	3	3
30.	NAO	3	3	2	2	2
31.	OLI	2	2	3	3	2
32.	PTA	3	4	2	2	3
33.	PTI	4	3	4	3	2
34.	RAS	2	2	2	2	3
35.	RAN	3	4	3	4	2
36.	RES	2	3	4	4	4
37.	ROS	3	3	3	3	3
38.	TAN	3	4	3	3	3
39.	TUK	4	4	4	4	4
40.	WIN	4	3	3	2	4

Data nilai tabel di atas dijadikan atribut dari setiap matakuliah yaitu : kalkulus 1 (Kal1), fisika terapan (FisT), kalkulus 2 (Kal2), statistik dan probabilitas (StaP), dan matematika diskrit (MatD). Data tabel 1 di atas diambil dari data-data mahasiswa jurusan Teknik Informatika semester 6 sebanyak 40 orang dan nilai yang digunakan dalam tabel berbentuk nilai pembobotan yang sering digunakan dalam penilaian akhir nilai mahasiswa.

## 2. Pembentukan *cluster* (K-Means)

Penghitungan K-means *Clustering* pada dataset tabel 1 di atas, terdiri dari beberapa langkah utama:

### 1. Menentukan jumlah *cluster* (k),

Jumlah Kluster yang dibentuk terdiri dari tiga yaitu TINGGI, SEDANG, dan CUKUP, maka jumlah kluster (k) = 3

- Kluster TINGGI dengan pembentuk nilai kluster 0 dan deskripsi kluster kelompok mahasiswa yang memiliki pengetahuan matematika tinggi, cepat tanggap, dan baik dalam menyelesaikan soal yang berhubungan dengan matematika.
- Kluster SEDANG dengan pembentuk nilai kluster 1 dan deskripsi kluster kelompok mahasiswa yang memiliki pengetahuan matematika sedang, tidak cepat tanggap, namun baik dalam menyelesaikan soal yang berhubungan dengan matematika, dan lebih suka menggunakan kalkulator.
- Kluster CUKUP dengan pembentuk nilai kluster 2 dan deskripsi kluster kelompok mahasiswa yang memiliki pengetahuan matematika cukup untuk tingkat dasar dan namun terlalu lama atau lambat menyelesaikan soal yang berhubungan dengan matematika.

### 2. Inisialisasi *centroid* awal

Memilih sejumlah titik pusat (*centroid*) awal dari data yang telah digunakan sebagai referensi. Pemilihan ini untuk membentuk *cluster* pada iterasi dan jumlah *centroid* yang dipilih sesuai dengan jumlah *cluster* yang diinginkan (k) dengan metode Inisialisasi Acak (*Random Initialization*), maka dipilih:

*Centroid* 1: (4, 3, 3, 3, 3) → diambil dari data mahasiswa No.1 AAG

*Centroid* 2: (2, 2, 3, 3, 2) → diambil dari data mahasiswa No.4 AP

*Centroid* 3: (3, 4, 3, 4, 4) → diambil dari data mahasiswa No.9 ERS

### 3. Hitungan jarak Euclidean dari setiap titik ke *centroid*

Hitungan jarak Euclidean dari setiap titik ke *centroid*, dengan jarak Euclidean antara dua titik  $p = (p_1, p_2, p_3, p_4, p_5)$  dan  $q = (q_1, q_2, q_3, q_4, q_5)$

$$Distance = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2 + (p_4 - q_4)^2 + (p_5 - q_5)^2}$$

Menghitung jarak setiap titik data ke *centroid* 1

$$Distance = \sqrt{(Kal1 - CKal1)^2 + (FisT - CFisT)^2 + (Kal2 - CKal2)^2 + (SatP - CSatP)^2 + (MatD - CMatD)^2}$$

Jarak AAG ke C1

$$= \sqrt{(4-4)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2} = 0$$

Jarak ARS ke C1

$$= \sqrt{(3-4)^2 + (4-3)^2 + (3-3)^2 + (4-3)^2 + (3-3)^2} = 1.732050808$$

dan selanjutnya untuk semua titik data dihitung jaraknya ke *centroid* 1. Begitu juga untuk setiap data masing-masing dihitung jaraknya ke *centroid* 2 dan *centroid* 3. Sehingga, berikut hasil keseluruhan semua titik data ke setiap *centroid*.

Tabel 2. Hasil perhitungan jarak setiap titik data ke setiap *centroid*

No.	Nama	Kal1	FisT	Kal2	StaP	MatD	Jarak ke C1	Jarak ke C2	Jarak ke C3
1	AAG	4	3	3	3	3	0.00	2.45	2.00
2	ARS	3	4	3	4	3	1.73	2.65	1.00
3	AS	4	3	2	3	3	1.00	2.65	2.24
4	AP	2	2	3	3	2	2.45	0.00	3.16
5	AZ	4	4	4	4	3	1.73	3.32	1.73
6	CAS	4	3	3	3	3	0.00	2.45	2.00
7	COP	3	3	3	3	3	1.00	1.73	1.73
8	EDN	3	3	3	3	3	1.00	1.73	1.73
9	ERS	3	4	3	4	4	2.00	3.16	0.00
10	ET	2	3	2	2	4	2.65	2.65	2.65
11	FRI	3	4	2	3	2	2.00	2.45	2.45
12	FRE	2	2	2	3	3	2.45	1.41	2.83
13	GAL	3	3	3	3	2	1.41	1.41	2.45
14	GAN	3	3	3	3	3	1.00	1.73	1.73
15	GIN	4	2	3	3	3	1.00	2.24	2.65
16	HAR	4	4	3	3	3	1.00	3.00	1.73
17	INA	3	4	3	4	4	2.00	3.16	0.00
18	IND	2	3	3	3	2	2.24	1.00	2.65
19	IST	3	3	3	2	3	1.41	2.00	2.45
20	JEL	3	4	4	3	3	1.73	2.65	1.73
21	JU	4	3	3	4	4	1.41	3.16	1.41
22	JUN	4	4	4	4	4	2.00	3.74	1.41
23	LAN	3	3	4	3	2	1.73	1.73	2.65
24	LIN	3	3	3	2	2	1.73	1.73	3.00
25	LIS	4	4	3	4	4	1.73	3.61	1.00
26	MAS	2	4	3	3	4	2.45	2.83	1.41
27	MRS	3	3	3	2	3	1.41	2.00	2.45
28	MRE	4	3	3	2	4	1.41	3.16	2.45
29	MT	4	3	3	3	3	0.00	2.45	2.00
30	NAO	3	3	2	2	2	2.00	2.00	3.16
31	OLI	2	2	3	3	2	2.45	0.00	3.16
32	PTA	3	4	2	2	3	2.00	2.83	2.45
33	PTI	4	3	4	3	2	1.41	2.45	2.83
34	RAS	2	2	2	2	3	2.65	1.73	3.32
35	RAN	3	4	3	4	2	2.00	2.45	2.00
36	RE	2	3	4	4	4	2.65	2.65	1.73
37	ROS	3	3	3	3	3	1.00	1.73	1.73
38	TAN	3	4	3	3	3	1.41	2.45	1.41
39	TUK	4	4	4	4	4	2.00	3.74	1.41
40	WIN	4	3	3	2	4	1.41	3.16	2.45

#### 4. Kelompokkan data berdasarkan *centroid* terdekat

Setelah jarak terdekat data dari ke tiga *centroid* awal tabel 2 didapatkan, selanjutnya menentukan klaster terdekatnya dengan cara membandingkan nilai terkecil dari antara jarak ke C1, jarak ke C2, hingga jarak ke C3.

Pada titik data no. 1 AAG: jarak ke C1= 0, jarak ke C2 = 2.45, jarak ke C3 = 2, maka klaster terdekat adalah C1.

Lakukan cara yang sama untuk setiap data hingga semua titik data dapat diketahui posisi klaster baru. Klaster baru untuk semua titik data mahasiswa disajikan pada tabel 3 berikut.

Tabel 3. Penentuan kelompok data klaster

No.	Nama	Kal1	FisT	Kal2	StaP	MatD	Jarak ke C1	Jarak ke C2	Jarak ke C3	Klaster
1	AAG	4	3	3	3	3	0.00	2.45	2.00	C1
2	ARS	3	4	3	4	3	1.73	2.65	1.00	C3
3	AS	4	3	2	3	3	1.00	2.65	2.24	C1

No.	Nama	Kal1	FisT	Kal2	StaP	MatD	Jarak ke C1	Jarak ke C2	Jarak ke C3	Klaster
4	AP	2	2	3	3	2	2.45	0.00	3.16	C2
5	AZ	4	4	4	4	3	1.73	3.32	1.73	C1
6	CAS	4	3	3	3	3	0.00	2.45	2.00	C1
7	COP	3	3	3	3	3	1.00	1.73	1.73	C1
8	EDN	3	3	3	3	3	1.00	1.73	1.73	C1
9	ERS	3	4	3	4	4	2.00	3.16	0.00	C3
10	ET	2	3	2	2	4	2.65	2.65	2.65	C1
11	FRI	3	4	2	3	2	2.00	2.45	2.45	C1
12	FRE	2	2	2	3	3	2.45	1.41	2.83	C2
13	GAL	3	3	3	3	2	1.41	1.41	2.45	C1
14	GAN	3	3	3	3	3	1.00	1.73	1.73	C1
15	GIN	4	2	3	3	3	1.00	2.24	2.65	C1
16	HAR	4	4	3	3	3	1.00	3.00	1.73	C1
17	INA	3	4	3	4	4	2.00	3.16	0.00	C3
18	IND	2	3	3	3	2	2.24	1.00	2.65	C2
19	IST	3	3	3	2	3	1.41	2.00	2.45	C1
20	JEL	3	4	4	3	3	1.73	2.65	1.73	C1
21	JU	4	3	3	4	4	1.41	3.16	1.41	C1
22	JUN	4	4	4	4	4	2.00	3.74	1.41	C3
23	LAN	3	3	4	3	2	1.73	1.73	2.65	C1
24	LIN	3	3	3	2	2	1.73	1.73	3.00	C1
25	LIS	4	4	3	4	4	1.73	3.61	1.00	C3
26	MAS	2	4	3	3	4	2.45	2.83	1.41	C3
27	MRS	3	3	3	2	3	1.41	2.00	2.45	C1
28	MRE	4	3	3	2	4	1.41	3.16	2.45	C1
29	MT	4	3	3	3	3	0.00	2.45	2.00	C1
30	NAO	3	3	2	2	2	2.00	2.00	3.16	C1
31	OLI	2	2	3	3	2	2.45	0.00	3.16	C2
32	PTA	3	4	2	2	3	2.00	2.83	2.45	C1
33	PTI	4	3	4	3	2	1.41	2.45	2.83	C1
34	RAS	2	2	2	2	3	2.65	1.73	3.32	C2
35	RAN	3	4	3	4	2	2.00	2.45	2.00	C1
36	RE	2	3	4	4	4	2.65	2.65	1.73	C3
37	ROS	3	3	3	3	3	1.00	1.73	1.73	C1
38	TAN	3	4	3	3	3	1.41	2.45	1.41	C1
39	TUK	4	4	4	4	4	2.00	3.74	1.41	C3
40	WIN	4	3	3	2	4	1.41	3.16	2.45	C1

##### 5. Memperbarui *centroid*

Sebelum memperbarui *centroid*, terlebih dahulu mendapatkan nilai rata-rata dari setiap klaster baru.

*Cluster C1* memiliki 27 anggota sesuai nomor urut mahasiswa yaitu:

1, 3, 5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 19, 20, 21, 23, 24, 27, 28, 29, 30, 32, 33, 35, 37, 38, 40

Maka selanjutnya hitung rata-rata untuk setiap kolom nomor tersebut:

$$\text{Kal1} = (4+4+4+4+3+3+2+3+3+3+4+4+3+3+4+3+3+3+4+4+3+3+4+3+3+3+4) / 27 = 3.37$$

$$\text{FisT} = (3+3+4+3+3+3+3+4+3+3+2+4+3+4+3+3+3+3+3+3+4+3+4+3+4+3) / 27 = 3.2$$

$$\text{Kal2} = (3+2+4+3+3+3+2+2+3+3+3+3+4+3+4+3+3+3+2+2+4+3+3+3) / 27 = 2.96$$

$$\text{StaP} = (3+3+4+3+3+3+2+3+3+3+3+2+3+4+3+2+2+2+3+2+2+3+4+3+3+2) / 27 = 2.81$$

$$\text{MatD} = (3+3+3+3+3+3+4+2+2+3+3+3+3+4+2+2+3+4+3+2+3+2+2+3+3+4) / 27 = 2.28$$

*Cluster C2* memiliki anggota sesuai nomor urut mahasiswa yaitu: 4, 12, 18, 31, 34

Maka selanjutnya hitung rata-rata untuk setiap kolom nomor tersebut:

$$\text{Kal1} = (2+2+2+2+2) / 5 = 2.00, \text{FisT} = (2+2+3+2+2) / 5 = 2.2, \text{Kal2} = (3+2+3+3+2) / 5 = 2.6, \text{StaP} = (3+3+3+3+2) / 5 = 2.8, \text{MatD} = (2+3+2+2+3) / 5 = 2.4$$

*Cluster C3* memiliki anggota sesuai nomor urut mahasiswa yaitu: 2, 9, 17, 22, 25, 26, 36, 39

Maka selanjutnya hitung rata-rata untuk setiap kolom nomor tersebut:  $\text{Kal1} = (3+3+3+4+4+2+2+4) / 8 = 3.12$

$$\text{FisT} = (4+4+4+4+4+4+3+4) / 8 = 3.87, \text{Kal2} = (3+3+3+4+3+3+4+4) / 8 = 3.37, \text{StaP} = (4+4+4+4+4+3+4+4) / 8 = 3.87, \text{MatD} = (3+4+4+4+4+4+4+4) / 8 = 3.87$$

Dari semua perhitungan tersebut didapatkan *Centroid* baru C1: (3.37, 3.2, 2.9, 2.81, 2.8), *Centroid* baru C2: (2, 2.2, 2.6, 2.8, 2.4), dan *Centroid* baru C3: (3.12, 3.87, 3.37, 3.87, 3.87)

##### 6. Mengulangi proses hingga konvergensi

Melakukan ulang proses hitung jarak terdekat setiap titik data dengan *centroid* baru C1, C2, dan C3, hingga pembentukan kelompok untuk kluster baru menggunakan cara dan rumus yang sama dari poin 1 hingga 5. Tampilan berikut pada tabel 4 adalah hasil dari proses mendapatkan kluster baru dari iterasi ke 1.

Tabel 4. Hasil pengelompokkan kluster iterasi ke 1

No.	Nama	Kal1	FisT	Kal2	StaP	MatD	Jarak ke C1	Jarak ke C2	Jarak ke C3	Kluster
1	AAG	4	3	3	3	3	0.70	2.28	1.79	C1
2	ARS	3	4	3	4	3	1.47	2.49	0.98	C3
3	AS	4	3	2	3	3	1.19	2.32	2.23	C1
4	AP	2	2	3	3	2	2.05	0.63	3.03	C2
5	AZ	4	4	4	4	3	1.87	3.32	1.40	C3
6	CAS	4	3	3	3	3	0.70	2.28	1.79	C1
7	COP	3	3	3	3	3	0.48	1.48	1.57	C1
8	EDN	3	3	3	3	3	0.48	1.48	1.57	C1
9	ERS	3	4	3	4	4	1.84	2.90	0.45	C3
10	ET	2	3	2	2	4	2.18	2.05	2.73	C2
11	FRI	3	4	2	3	2	1.58	2.19	2.49	C1
12	FRE	2	2	2	3	3	2.08	0.89	2.86	C2
13	GAL	3	3	3	3	2	1.01	1.41	2.28	C1
14	GAN	3	3	3	3	3	0.48	1.48	1.57	C1
15	GIN	4	2	3	3	3	1.39	2.14	2.44	C1
16	HAR	4	4	3	3	3	1.02	2.79	1.57	C1
17	INA	3	4	3	4	4	1.84	2.90	0.45	C3
18	IND	2	3	3	3	2	1.66	1.00	2.54	C2
19	IST	3	3	3	2	3	0.93	1.67	2.28	C1
20	JEL	3	4	4	3	3	1.37	2.57	1.40	C1
21	JU	4	3	3	4	4	1.76	2.97	1.31	C3
22	JUN	4	4	4	4	4	2.17	3.63	1.10	C3
23	LAN	3	3	4	3	2	1.44	1.95	2.34	C1
24	LIN	3	3	3	2	2	1.28	1.61	2.82	C1
25	LIS	4	4	3	4	4	1.91	3.38	0.98	C3
26	MAS	2	4	3	3	4	1.94	2.45	1.48	C3
27	MRS	3	3	3	2	3	0.93	1.67	2.28	C1
28	MRE	4	3	3	2	4	1.53	2.83	2.28	C1
29	MT	4	3	3	3	3	0.70	2.28	1.79	C1
30	NAO	3	3	2	2	2	1.60	1.67	3.11	C1
31	OLI	2	2	3	3	2	2.05	0.63	3.03	C2
32	PTA	3	4	2	2	3	1.53	2.37	2.49	C1
33	PTI	4	3	4	3	2	1.53	2.61	2.49	C1
34	RAS	2	2	2	2	3	2.23	1.18	3.31	C2
35	RAN	3	4	3	4	2	1.71	2.45	1.92	C1
36	RE	2	3	4	4	4	2.38	2.57	1.57	C3
37	ROS	3	3	3	3	3	0.48	1.48	1.57	C1
38	TAN	3	4	3	3	3	0.89	2.19	1.31	C1
39	TUK	4	4	4	4	4	2.17	3.63	1.10	C3
40	WIN	4	3	3	2	4	1.53	2.83	2.28	C1

Hasil tabel kluster awal dan hasil kluster iterasi ke 1 masih mengalami perbedaan, maka langkah yang sama diulang untuk menentukan kluster berikutnya pada iterasi ke 2 untuk mendapatkan konvergensi. Jika pada iterasi ke 2 dan seterusnya hingga batas maksimum iterasi yaitu 100, data posisi kluster awal dianggap konvergensi. Setelah dilakukan iterasi hingga iterasi ke 4 didapatkan hasil yang konvergen yaitu kluster 4 iterasi ke 4 tidak mengalami perubahan dengan kluster ke 3 dari iterasi ke 3. Proses iterasi ke 3 dan ke 4 mendapatkan posisi kluster yang sama. Tampilan semua iterasi pada tabel 5.

Tabel 5. Iterasi untuk mendapatkan kluster yang konvergen

Kluster Awal	Kluster ke 1	Kluster ke 2	Kluster ke 3	Kluster ke 4
C1	C1	C1	C1	C1
C3	C3	C3	C3	C3
C1	C1	C1	C1	C1
C2	C2	C2	C2	C2
C1	C3	C3	C3	C3
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C3	C3	C3	C3	C3

C1	C2	C2	C2	C2
C1	C1	C1	C1	C1
C2	C2	C2	C2	C2
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C3	C3	C3	C3	C3
C2	C2	C2	C2	C2
C1	C1	C1	C1	C1
C1	C1	C3	C3	C3
C1	C3	C3	C3	C3
C3	C3	C3	C3	C3
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C3	C3	C3	C3	C3
C3	C3	C3	C3	C3
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C2	C2	C2	C2	C2
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C2	C2	C2	C2	C2
C1	C1	C1	C1	C1
C3	C3	C3	C3	C3
C1	C1	C1	C1	C1
C1	C1	C1	C1	C1
C3	C3	C3	C3	C3
C1	C1	C1	C1	C1

## 7. Hasil akhir kluster

Hasil akhir kluster terdapat pada kluster ke 4 dari iterasi ke 4. Hasil tersebut digabungkan ditampilkan pada tabel 6.

Tabel 6. Hasil akhir kluster

No.	Nama	Kal1	FisT	Kal2	StaP	MatD	Kluster	Keterangan
1	AAG	4	3	3	3	3	C1	Tinggi
2	ARS	3	4	3	4	3	C1	Tinggi
3	AS	4	3	2	3	3	C1	Tinggi
4	AP	2	2	3	3	2	C3	Cukup
5	AZ	4	4	4	4	3	C3	Cukup
6	CAS	4	3	3	3	3	C1	Tinggi
7	COP	3	3	3	3	3	C1	Tinggi
8	EDN	3	3	3	3	3	C1	Tinggi
9	ERS	3	4	3	4	4	C1	Tinggi
10	ET	2	3	2	2	4	C1	Tinggi
11	FRI	3	4	2	3	2	C1	Tinggi
12	FRE	2	2	2	3	3	C1	Tinggi
13	GAL	3	3	3	3	2	C3	Cukup
14	GAN	3	3	3	3	3	C1	Tinggi
15	GIN	4	2	3	3	3	C1	Tinggi
16	HAR	4	4	3	3	3	C1	Tinggi
17	INA	3	4	3	4	4	C1	Tinggi
18	IND	2	3	3	3	2	C3	Cukup
19	IST	3	3	3	2	3	C2	Sedang
20	JEL	3	4	4	3	3	C2	Sedang
21	JU	4	3	3	4	4	C1	Tinggi
22	JUN	4	4	4	4	4	C1	Tinggi
23	LAN	3	3	4	3	2	C3	Cukup
24	LIN	3	3	3	2	2	C2	Sedang
25	LIS	4	4	3	4	4	C1	Tinggi
26	MAS	2	4	3	3	4	C1	Tinggi
27	MRS	3	3	3	2	3	C2	Sedang
28	MRE	4	3	3	2	4	C2	Sedang



No.	Nama	Kal1	FisT	Kal2	StaP	MatD	Klaster	Keterangan
29	MT	4	3	3	3	3	C1	Tinggi
30	NAO	3	3	2	2	2	C1	Tinggi
31	OLI	2	2	3	3	2	C3	Cukup
32	PTA	3	4	2	2	3	C1	Tinggi
33	PTI	4	3	4	3	2	C3	Cukup
34	RAS	2	2	2	2	3	C1	Tinggi
35	RAN	3	4	3	4	2	C3	Cukup
36	RE	2	3	4	4	4	C1	Tinggi
37	ROS	3	3	3	3	3	C1	Tinggi
38	TAN	3	4	3	3	3	C1	Tinggi
39	TUK	4	4	4	4	4	C1	Tinggi
40	WIN	4	3	3	2	4	C2	Sedang

### 3. Uji Klasifikasi KNN

Hasil klaster dari metode K-Means akan digunakan sebagai alat pemetaan dalam uji klasifikasi KNN. Menggunakan K-Means sebelum KNN dapat meningkatkan performa dan efisiensi dalam melakukan klasifikasi, hal ini terutama pada dataset yang besar. Setiap data poin sekarang telah memiliki label klaster yang sesuai. Sehingga dengan KNN selanjutnya hanya mencari tetangga terdekat dalam klaster yang sama atau klaster yang relevan saja, bukan seluruh dataset.

Sebelum menguji klaster dalam melakukan indentifikasi data, terlebih dahulu tentukan jumlah tetangga terdekat ( $k$ ) = 3 dan data baru yang akan diinput NamaUji: RET, Kal1Uji: 2, FisTUji: 3, Kal2Uji: 4, StaPUji: 4, dan MatDUji: 3

1. Hitung jarak Euclidean antara data baru dan setiap data dalam dataset.

jarak Euclidean antara dua titik  $p = (p_1, p_2, p_3, p_4, p_5)$  dan  $q = (q_1, q_2, q_3, q_4, q_5)$

$$\text{Euclidean Distance} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2 + (p_4 - q_4)^2 + (p_5 - q_5)^2}$$

*Euclidean Distance*

$$= \sqrt{(Kal1 - Kal1Uji)^2 + (FisT - FisTUji)^2 + (Kal2 - Kal2Uji)^2 + (SatP - SatPUji)^2 + (MatD - MatDUji)^2}$$

$$= \sqrt{(4 - 2)^2 + (3 - 3)^2 + (3 - 4)^2 + (3 - 2)^2 + (3 - 3)^2}$$

$$= 2.45$$

Lakukan hitung jarak terdekat data baru dengan setiap titik data dengan cara yang sama, demikian tampilan hasil semua jarak terdekat pada tabel berikut:

Tabel 7. Hasil jarak Euclidean terdekat

No.	Nama	Kal1	FisT	Kal2	StaP	MatD	Kelas	Jarak
1	AAG	4	3	3	3	3	Tinggi	2.45
2	ARS	3	4	3	4	3	Tinggi	2.65
3	AS	4	3	2	3	3	Tinggi	3.00
4	AP	2	2	3	3	2	Cukup	2.00
5	AZ	4	4	4	4	3	Cukup	3.00
6	CAS	4	3	3	3	3	Tinggi	2.45
7	COP	3	3	3	3	3	Tinggi	1.73
8	EDN	3	3	3	3	3	Tinggi	1.73
9	ERS	3	4	3	4	4	Tinggi	2.83
10	ET	2	3	2	2	4	Tinggi	2.24
11	FRI	3	4	2	3	2	Tinggi	2.83
12	FRE	2	2	2	3	3	Tinggi	2.45
13	GAL	3	3	3	3	2	Cukup	2.00
14	GAN	3	3	3	3	3	Tinggi	1.73
15	GIN	4	2	3	3	3	Tinggi	2.65
16	HAR	4	4	3	3	3	Tinggi	2.65
17	INA	3	4	3	4	4	Tinggi	2.83
18	IND	2	3	3	3	2	Cukup	1.73
19	IST	3	3	3	2	3	Sedang	1.41
20	JEL	3	4	4	3	3	Sedang	1.73
21	JU	4	3	3	4	4	Tinggi	3.16
22	JUN	4	4	4	4	4	Tinggi	3.16
23	LAN	3	3	4	3	2	Cukup	1.73
24	LIN	3	3	3	2	2	Sedang	1.73
25	LIS	4	4	3	4	4	Tinggi	3.32
26	MAS	2	4	3	3	4	Tinggi	2.00
27	MRS	3	3	3	2	3	Sedang	1.41
28	MRE	4	3	3	2	4	Sedang	2.45
29	MT	4	3	3	3	3	Tinggi	2.45
30	NAO	3	3	2	2	2	Tinggi	2.45



No.	Nama	Kal1	FisT	Kal2	StaP	MatD	Kelas	Jarak
31	OLI	2	2	3	3	2	Cukup	2.00
32	PTA	3	4	2	2	3	Tinggi	2.45
33	PTI	4	3	4	3	2	Cukup	2.45
34	RAS	2	2	2	2	3	Tinggi	2.24
35	RAN	3	4	3	4	2	Cukup	2.83
36	RE	2	3	4	4	4	Tinggi	2.24
37	ROS	3	3	3	3	3	Tinggi	1.73
38	TAN	3	4	3	3	3	Tinggi	2.00
39	TUK	4	4	4	4	4	Tinggi	3.16
40	WIN	4	3	3	2	4	Sedang	2.45

2. Temukan 3 tetangga terdekat berdasarkan jarak yang sudah dihitung. Setelah mendapatkan hasil keseluruhan perhitungan jarak terdekat, maka didapatkan 3 nilai jarak tetangga yang terdekat, yaitu terdapat pada:

Tabel 8. Nilai jarak terkecil

No.	Nama	Kelas	Jarak
19	IST	Sedang	1.41
27	MRS	Sedang	1.41
7	COP	Tinggi	1.73

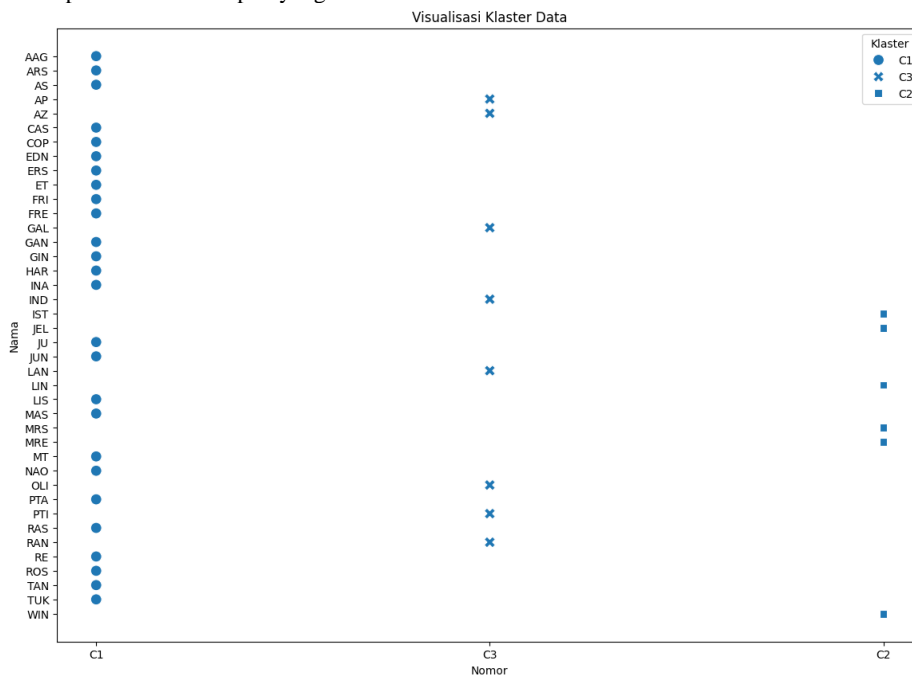
3. Hasil klasifikasi data baru  
Klasifikasikan data baru berdasarkan mayoritas label dari 3 tetangga terdekat tabel 8. Diketahui terdapat 2 Kelas Sedang dan 1 Kelas Tinggi, dengan demikian hasil voting adalah data baru dengan NamaUji: RET, Kal1Uji: 2, FisTUji: 3, Kal2Uji: 4, StaPUji: 2, dan MatDUji: 3 termasuk kelas Sedang.

#### 4. Pembahasan Klaster dan Klasifikasi

K-Means dan KNN dua algoritma yang digunakan dalam analisis data pengetahuan matematika mahasiswa.

1. Penggunaan K-Means untuk Klasterisasi

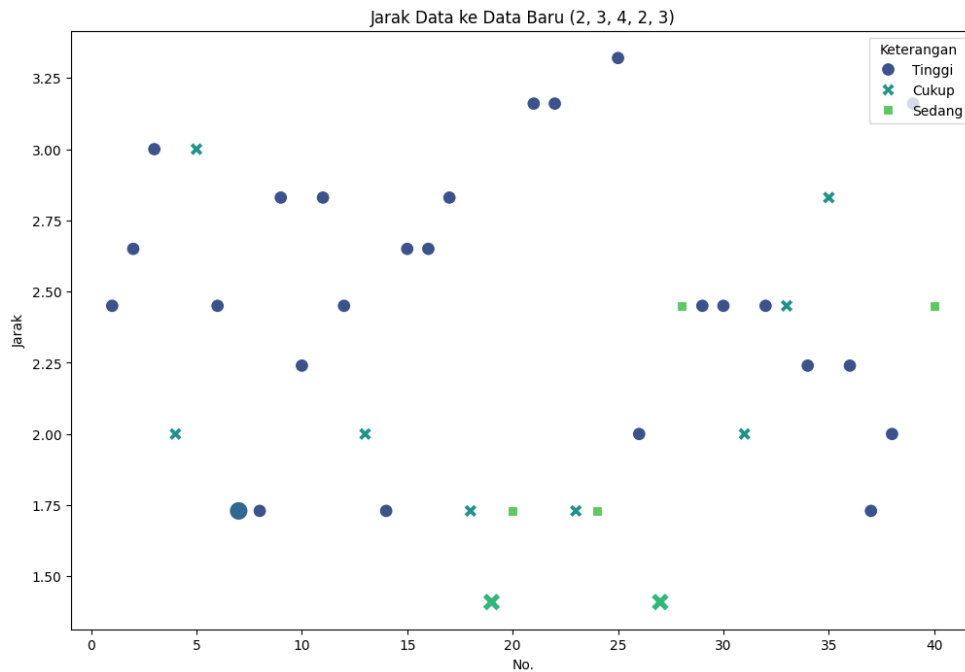
Dalam contoh di atas, K-Means digunakan untuk mengelompokkan data mahasiswa berdasarkan nilai-nilai mereka dalam lima mata pelajaran: Kalkulus 1, Fisika Terapan, Kalkulus 2, Statistik dan Probalitas, dan Matematika Diskrit. Hasilnya, data dikelompokkan ke dalam tiga klaster: C1 (Tinggi), C2 (Sedang), dan C3 (Cukup). Klaster ini membantu dalam memahami distribusi data dan pola dalam kelompok yang lebih besar.



Gambar 1. Plot Klaster

2. Penggunaan KNN untuk Klasifikasi

Setelah data dikelompokkan menggunakan K-Means, KNN digunakan untuk mengklasifikasikan data baru berdasarkan kemiripan dengan data yang sudah dihasil sebelumnya. Dalam hal ini, data baru (2, 3, 4, 2, 3) diklasifikasikan dengan mencari tiga data terdekat ( $k=3$ ) menggunakan jarak Euclidean.



Gambar 1. Jarak setiap data ke data baru

### KESIMPULAN

Kolaborasi antara K-Means dan KNN memberikan pendekatan yang efektif dalam mengelompokkan dan mengklasifikasikan data. Adapun kesimpulan penelitian ini sebagai berikut:

1. K-Means mengelompokkan data mahasiswa ke dalam tiga kluster berdasarkan nilai mahasiswa, yang memberikan wawasan mengenai distribusi dan pola dalam dataset.
2. KNN digunakan untuk mengklasifikasikan data baru dengan mempertimbangkan kedekatannya dengan data yang telah dikelompokkan, memberikan hasil klasifikasi yang tepat dan dapat diandalkan.
3. Penggunaan K-Means untuk klusterisasi memudahkan dalam mengidentifikasi kelompok yang homogen dalam data, sementara KNN memungkinkan untuk prediksi dan klasifikasi berdasarkan data yang ada..

### REFERENCES

- [1] M. Z. bin Mohamed, R. Hidayat, N. N. binti Suhaizi, N. binti M. Sabri, M. K. H. bin Mahmud, and S. N. binti Baharuddin, "Artificial intelligence in mathematics education: A systematic literature review," *International Electronic Journal of Mathematics Education*, vol. 17, no. 3, p. em0694, Jun. 2022, doi: 10.29333/iejme/12132.
- [2] D. Wolfram, "The symbiotic relationship between information retrieval and informetrics," *Scientometrics*, vol. 102, no. 3, pp. 2201–2214, Mar. 2015, doi: 10.1007/s11192-014-1479-0.
- [3] K. R. Devi, "THE ROLE OF MATHEMATICS IN COMPUTER SCIENCE COLLEGE EDUCATION," 2019. [Online]. Available: [www.ijcscst.com](http://www.ijcscst.com)
- [4] P. M. Farib, M. Ikhsan, and M. Subianto, "Proses berpikir kritis matematis siswa sekolah menengah pertama melalui discovery learning," *Jurnal Riset Pendidikan Matematika*, vol. 6, no. 1, pp. 99–117, May 2019, doi: 10.21831/jrpm.v6i1.21396.
- [5] M. Zulyanty, I. Yuwono, D. M. Muksar, and U. N. Malang, "METAKOGNISI SISWA DENGAN GAYA BELAJAR INTROVERT DALAM MEMECAHKAN MASALAH MATEMATIKA," 2017. [Online]. Available: <http://journal2.um.ac.id/index.php/jkpm>
- [6] D. Herawaty, "Model pembelajaran matematika realistik yang efektif untuk meningkatkan kemampuan matematika siswa SMP," *Jurnal Pendidikan Matematika Raflesia*, vol. 3, no. 2, 2018, [Online]. Available: <https://ejournal.unib.ac.id/index.php/jpmr>
- [7] Wahyuddin, "Meningkatkan Hasil Belajar Matematika Melalui Pembelajaran Dengan Pemberian Tugas Terstruktur Disertai Umpan Balik," *Media Pendidikan Matematika*, vol. 8, no. 2, p. 14, Dec. 2020, [Online]. Available: <http://ojs.ikipmataram.ac.id/index.php/jmpm>
- [8] I. Firman Ashari, E. Dwi Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," 2023. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [9] J. B. Chandra and D. Nasien, "Application Of Machine Learning K-Nearest Neighbour Algorithm To Predict Diabetes," Pekanbaru, Indonesia, Jun. 2022. [Online]. Available: <http://www.ijeepse.ejournal.unri.ac.id>
- [10] A. Bode, Z. Y. Lamasigi, and I. C. R. Drajana, "The K-Nearest Neighbor Algorithm using Forward Selection and Backward Elimination in Predicting the Student's Satisfaction Level of University Ichsan Gorontalo toward Online Lectures during the COVID-19 Pandemic," *ILKOM Jurnal Ilmiah*, vol. 15, no. 1, pp. 118–123, Apr. 2023, doi: 10.33096/ilkom.v15i1.1381.118-123.

- [11] A. Naas, S. Na'ima, H. Mulyo, and A. Widiastuti, "Klasifikasi penerima bantuan program rehabilitasi rumah tidak layak huni menggunakan algoritme K-Nearest Neighbor," *Jurnal Teknologi dan Sistem Komputer*, vol. 10, no. 1, pp. 32–37, 2022, doi: 10.14710/jtsiskom.2022.14110.
- [12] F. Ramadhani and O. Krianto Sulaiman, "Implementation of the K-Nearest Neighbor (KNN) Algorithm in Making a Web-Based Article Topic System," *International of Computer Science and Information Technology (AloCSIT) Journal*, vol. 3, no. 1, pp. 1–8, 2021.
- [13] J. Pecuchova and M. Drlik, "Identification of Students with Similar Behavioural Patterns Using *Clustering* Techniques," in *E-learning in the Transformation of Education in Digital Society*, 2022, pp. 257–267. doi: 10.34916/el.2022.14.19.
- [14] F. E. Ozturk, N. Demirel, and M. Bilgisi, "Comparison of the Methods to Determine Optimal Number of *Cluster*," 2023. [Online]. Available: [www.dergipark.gov.tr/veri](http://www.dergipark.gov.tr/veri)
- [15] S. Suraya, M. Sholeh, and U. Lestari, "Evaluation of Data *Clustering* Accuracy using K-Means Algorithm," *International Journal of Multidisciplinary Approach Research and Science*, vol. 2, no. 01, pp. 385–396, Dec. 2023, doi: 10.59653/ijmars.v2i01.504.
- [16] Y. Hasan, "Machine Learning Pengenalan Herpetofauna Dilindungi Di Indonesia," 2023. doi: 10.54367.
- [17] H. Sunandar, "Machine Learning Pengenalan Anura Berdasarkan Corak dan Warna," 2023. doi: 10.54367.
- [18] S. Sanjaya, M. L. Pura, S. K. Gusti, F. Yanto, and F. Syafria, "K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue, Saturation, and Value Colors," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 2, no. 2, p. 101, Nov. 2019, doi: 10.24014/ijaidm.v2i2.7975.
- [19] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means *Clustering* Algorithm," *J (Basel)*, vol. 2, no. 2, pp. 226–235, Jun. 2019, doi: 10.3390/j2020016.
- [20] W. Utomo, "The comparison of k-means and k-medoids algorithms for *Clustering* the spread of the covid-19 outbreak in Indonesia," *ILKOM Jurnal Ilmiah*, vol. 13, no. 1, pp. 31–35, Apr. 2021, doi: 10.33096/ilkom.v13i1.763.31-35.
- [21] H. Mayangsari, D. Kartika, D. Musu, and J. K. Sistem Informasi Universitas Dipa Makassar Jln Perintis Kemerdekaan, "Analisis Sistem Informasi Pendistribusian Tempat KKL Menggunakan Metode K-Nearst Neighbor," 2023.