

SKRIPSI

**ANALISIS KLASSTER NEGARA PESERTA PISA 2022
BERDASARKAN INDIKATOR LITERASI MATEMATIKA
DENGAN *GAUSSIAN MIXTURE MODEL***



**MIRA ARWINI
E0221303**

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SULAWESI BARAT
2025**

SURAT PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Mira Arwini
Tempat/Tanggal Lahir : Rantekatoan, 14 September 2003
NIM : E0221303
Program Studi : Statistika

Menyatakan bahwa tugas akhir dengan judul “Analisis Klaster Negara Peserta PISA 2022 Berdasarkan Indikator Literasi Matematika dengan *Gaussian Mixture Model*” disusun berdasarkan prosedur ilmiah yang telah melalui pembimbingan dan bukan merupakan plagiat dari karya ilmiah/naskah yang lain. Apabila di kemudian hari terbukti bahwa pernyataan ini tidak benar, maka saya bersedia menerima sanksi sesuai yang berlaku.

Majene, 31 Oktober 2025

yang menyatakan,



Mira Arwini

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Mira Arwini
Nim : E0221303
Judul : Analisis Klaster Negara Peserta PISA 2022 Berdasarkan Indikator Literasi Matematika dengan *Gaussian Mixture Model*

Telah berhasil dipertanggungjawabkan dihadapan Tim Penguji (SK Nomor 150/UN55.7/HK.04/2025, tanggal 31 Oktober 2025) dan diterima sebagai bagian persyaratan memperoleh gelar sarjana S1 pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sulawesi Barat.

Disahkan oleh:

Dekan FMIPA

Universitas Sulawesi Barat

Musafira, S.Si., M.Sc.

NIP: 19770911200660422002

Tim Penguji:

Ketua Penguji : Musafira, S.Si., M.Sc.

Sekretaris : Putri Indi Rahayu, S.Si., M.Stat.

Pembimbing 1 : Retno Mayapada, S.Si., M.Si.

Pembimbing 2 : Muh. Hijrah, S.Pd., M.Si.

Penguji 1 : Reski Wahyu Yanti, S.Si., M.Si.

Penguji 2 : Muhammad Hidayatullah, S.Pd., M.Kom

Penguji 3 : Putri Indi Rahayu, S.Si., M.Stat.

ABSTRAK

Literasi matematika merupakan aspek penting dalam penilaian PISA yang mencerminkan kemampuan siswa menerapkan konsep matematika dalam kehidupan nyata. Penelitian ini bertujuan menerapkan metode *Gaussian Mixture Model* (GMM) untuk mengelompokkan negara peserta PISA 2022 berdasarkan indikator literasi matematika serta mengidentifikasi karakteristik tiap klaster. Analisis dilakukan menggunakan data PISA 2022 yang diperoleh dari OECD dengan empat indikator persentase yang merepresentasikan aspek literasi matematika siswa. Hasil analisis menunjukkan bahwa GMM menghasilkan tiga klaster optimal dengan model terbaik adalah EEI, yaitu model dengan bentuk elipsoid yang memiliki volume dan orientasi sama antar klaster yang dipilih berdasarkan nilai *Bayesian Information Criterion* (BIC). Berdasarkan hasil identifikasi klaster pertama mencakup negara-negara dengan nilai tinggi pada seluruh indikator literasi matematika, ditandai dengan kemampuan memahami representasi visual yang baik, kepercayaan diri tinggi, dan pemerataan kemampuan antar siswa yang relatif merata. Klaster kedua terdiri atas negara dengan tingkat pencapaian menengah antar indikator literasi matematika yang cenderung stabil, di mana siswa masih mampu mengaitkan konsep matematika dengan konteks nyata meskipun pemanfaatan alat bantu visual belum optimal. Sementara itu, klaster ketiga menunjukkan negara-negara dengan tingkat capaian yang relatif rendah dibandingkan dua klaster lainnya, di mana meskipun pemahaman siswa yang cukup baik terhadap informasi visual matematika, namun tingkat kepercayaan diri siswa dalam menginterpretasi solusi matematika dan menerapkannya pada konteks kehidupan nyata masih rendah. Temuan ini menunjukkan adanya variasi karakteristik dari masing-masing variabel antarnegara peserta PISA 2022 berdasarkan indikator literasi matematika.

Kata kunci: Literasi Matematika, PISA 2022, *Gaussian Mixture Model*, Klaster

ABSTRACT

Mathematical literacy is an essential aspect of the PISA assessment, reflecting students' ability to apply mathematical concepts in real-life situations. This study aims to apply the Gaussian Mixture Model (GMM) method to cluster PISA 2022 participating countries based on mathematical literacy indicators and to identify the characteristics of each cluster. The analysis was conducted using PISA 2022 data obtained from the OECD, consisting of four percentage-based indicators representing various aspects of students' mathematical literacy. The results indicate that GMM produced three optimal clusters, with the best model being EEI—an ellipsoidal model with equal volume and orientation across clusters—selected based on the Bayesian Information Criterion (BIC) value. The first cluster includes countries with high scores across all mathematical literacy indicators, characterized by strong visual representation skills, high self-confidence, and relatively even performance among students. The second cluster comprises countries with moderate achievement levels across indicators, where students are still able to connect mathematical concepts with real-world contexts, although the use of visual aids remains suboptimal. Meanwhile, the third cluster represents countries with relatively lower performance compared to the other two clusters, where students demonstrate adequate understanding of visual mathematical information but exhibit lower levels of self-confidence in interpreting and applying mathematical solutions in real-life contexts. These findings highlight variations in the characteristics of mathematical literacy indicators among PISA 2022 participating countries.

Keywords: *Mathematical Literacy, PISA 2022, Gaussian Mixture Model, Cluster*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan merupakan pilar utama dalam pembangunan sumber daya manusia yang berkualitas. Dalam menghadapi tantangan abad ke-21 yang ditandai dengan kemajuan teknologi dan persaingan global, diperlukan kompetensi esensial seperti literasi matematika, yang mencerminkan kemampuan individu dalam merumuskan, menerapkan, dan menafsirkan konsep serta prosedur matematika dalam berbagai konteks kehidupan sehari-hari secara efektif (Anzilah & Amir, 2021). Literasi ini menjadi dasar penting bagi individu untuk dapat membuat keputusan yang rasional dan bertanggung jawab.

Salah satu instrumen internasional yang secara berkala mengukur kompetensi literasi matematika adalah *Programme for International Student Assessment* (PISA) yang diselenggarakan oleh *Organization for Economic Co-operation and Development* (OECD). Penilaian ini menargetkan siswa berusia 15 tahun dan dilaksanakan setiap tiga tahun sekali, bertujuan untuk memberikan gambaran menyeluruh mengenai kualitas sistem pendidikan di berbagai negara. PISA 2022 menjadi acuan terbaru dalam memotret kemampuan literasi matematika siswa dari berbagai negara peserta. Berdasarkan laporan PISA 2018, nilai rata-rata literasi matematika siswa Indonesia adalah 379, masih di bawah rata-rata OECD sebesar 489 (Azid et al., 2023). Pada siklus berikutnya, hasil PISA 2022 menunjukkan bahwa skor Indonesia kembali mengalami penurunan menjadi 366, sementara rata-rata global mencapai 472 (Hamdiyanti et al., 2024). Selain itu, performa Indonesia menunjukkan tren penurunan sejak tahun 2000 (Putrawangsa & Hasanah, 2022). Kondisi ini menunjukkan bahwa Indonesia masih berada pada capaian level yang rendah dalam literasi matematika. Negara-negara dengan skor tinggi sering dijadikan rujukan dalam kebijakan pendidikan, sedangkan negara dengan skor rendah, seperti Indonesia, perlu melakukan evaluasi menyeluruh.

Literasi matematika tidak hanya berkaitan dengan kemampuan kognitif, tetapi juga aspek afektif dan sikap terhadap matematika. Oleh karena itu, PISA 2022

menyertakan berbagai indikator yang mewakili dimensi afektif dan kognitif siswa. Beberapa indikator tersebut meliputi: persentase siswa yang mampu memahami informasi matematika dari diagram, grafik, atau simulasi; persentase kepercayaan diri siswa dalam memahami informasi matematika dari diagram, grafik, atau simulasi; kepercayaan diri siswa dalam menginterpretasi solusi matematika dalam konteks kehidupan nyata; serta selisih persentase siswa dengan kepercayaan diri tinggi dan rendah terhadap keterampilan matematika abad ke-21. Indikator ini merepresentasikan kombinasi keterampilan teknis dan keyakinan diri siswa yang berpengaruh terhadap performa matematika.

Berbagai penelitian sebelumnya menunjukkan bahwa literasi matematika berperan penting dalam keberhasilan akademik dan sosial. Hamdiyanti et al. (2024) menyatakan bahwa literasi matematika memungkinkan siswa untuk memahami dan menerapkan konsep-konsep matematika dalam kehidupan sehari-hari, serta berpikir logis dalam memecahkan masalah. Mereka juga menekankan bahwa kombinasi antara literasi matematika dan motivasi belajar sangat penting untuk menunjang keberhasilan pendidikan dan peningkatan performa akademik. Sejalan dengan itu, Azid et al. (2023) menyoroti pentingnya peningkatan literasi matematika di Indonesia berdasarkan temuan PISA terbaru, yang menunjukkan masih rendahnya kemampuan siswa dalam aspek tersebut.

Dalam rangka memahami pola performa negara-negara peserta PISA terhadap indikator tersebut, pendekatan klasterisasi (*clustering*) menjadi metode analitik yang relevan. *Clustering* memungkinkan terbentuknya kelompok negara berdasarkan kemiripan karakteristik literasi matematika, sehingga dapat memberikan gambaran yang lebih mendalam dan tersegmentasi. Sejumlah metode klasterisasi telah digunakan dalam berbagai penelitian, seperti *K-Means*, *Hierarchical Clustering*, DBSCAN, dan *Gaussian Mixture Model* (GMM).

Beberapa studi memberikan perbandingan sistematis atas metode-metode tersebut. Penelitian oleh Greve et al. (2015) membandingkan *K-Means*, *Ward's Hierarchical*, dan GMM melalui studi simulasi serta penerapan pada data pola makan dari 1.791 anak usia 2–9 tahun. Data tersebut dikumpulkan melalui *Food Frequency Questionnaire* (FFQ) dalam studi *Identification and Prevention of*

Dietary- and Lifestyle-Induced Health Effects in Children and Infants (IDEFICS) yang dilaksanakan di Jerman. Hasilnya menunjukkan bahwa GMM memberikan hasil paling konsisten dalam mengenali struktur klaster, khususnya pada data dengan bentuk dan orientasi klaster yang beragam, dengan performa unggul pada lebih dari 90% kasus untuk data kompleks.

Di Indonesia, Andayani et al. (2025) membandingkan DBSCAN dan GMM dalam mengelompokkan provinsi berdasarkan indikator kesejahteraan. GMM menunjukkan nilai evaluasi klaster yang lebih baik berdasarkan *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. Sementara itu, penelitian oleh Airlangga (2025) yang membandingkan empat metode (*K-Means*, *Agglomerative*, DBSCAN, dan GMM) dalam menganalisis data pariwisata, juga menunjukkan performa kompetitif dari GMM, terutama dalam menangani distribusi data yang kompleks dan tumpang tindih. GMM merupakan metode klasterisasi berbasis probabilistik yang tidak menetapkan suatu objek secara mutlak ke dalam satu klaster saja. Sebaliknya, setiap objek diberi nilai probabilitas keanggotaan terhadap masing-masing klaster yang tersedia. Pendekatan ini memberikan fleksibilitas dalam menangkap struktur data yang kompleks atau saling tumpang tindih, karena satu objek dapat secara simultan memiliki kemungkinan untuk berada di lebih dari satu klaster (Bishop, 2006).

Melalui pertimbangan pentingnya literasi matematika dan perlunya pengelompokan negara berdasarkan indikator yang relevan, penelitian ini bertujuan untuk menerapkan GMM dalam mengelompokkan negara peserta PISA 2022 berdasarkan skor literasi matematika, serta mengidentifikasi karakteristik masing-masing klaster yang terbentuk. Dengan pemetaan tersebut, diharapkan dapat diperoleh informasi yang lebih tersegmentasi mengenai posisi Indonesia dan negara lain dalam spektrum literasi matematika global, sehingga dapat menjadi dasar untuk merumuskan strategi peningkatan kualitas pendidikan secara lebih tepat sasaran.

1.2 Rumusan Masalah

Berdasarkan latar belakang maka rumusan masalah dari penelitian ini adalah sebagai berikut.

1. Bagaimana hasil penerapan metode GMM dalam mengelompokkan negara-negara peserta PISA 2022 berdasarkan indikator literasi matematika?
2. Bagaimana karakteristik masing-masing klaster hasil pengelompokan negara peserta PISA 2022 berdasarkan literasi matematika?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah tujuan pada penelitian ini adalah sebagai berikut.

1. Mengetahui hasil penerapan metode GMM dalam mengelompokkan negara-negara peserta PISA 2022 berdasarkan indikator literasi matematika
2. Mengidentifikasi dan mendeskripsikan karakteristik masing-masing klaster hasil pengelompokan negara peserta PISA 2022 berdasarkan literasi matematika

1.4 Manfaat Penelitian

Dengan tercapainya tujuan penelitian, penelitian ini diharapkan dapat memberikan manfaat:

1. Memberikan gambaran mengenai penerapan metode GMM dalam pengelompokan negara-negara peserta PISA 2022 berdasarkan indikator literasi matematika, serta menjadi acuan dalam penggunaan metode analisis klaster berbasis statistik multivariat.
2. Menyediakan informasi mengenai hasil pengelompokan dan karakteristik masing-masing klaster negara, yang dapat digunakan oleh pihak terkait di bidang pendidikan untuk memahami pola performa negara, termasuk Indonesia, dalam capaian literasi matematika.

1.5 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Penelitian ini hanya berfokus pada pengelompokan negara-negara peserta PISA 2022 berdasarkan indikator literasi matematika yang diambil dari data

kuesioner siswa, tanpa menganalisis hubungan kausal antarindikator maupun keterkaitannya dengan domain literasi lainnya seperti membaca dan sains.

2. Metode yang digunakan dalam penelitian ini terbatas pada GMM dengan pendekatan statistik multivariat, di mana jumlah kluster optimal ditentukan menggunakan BIC. Data yang digunakan merupakan data sekunder yang tersedia secara publik dari OECD untuk tahun 2022.

BAB II

KAJIAN PUSTAKA

2.1 *Data mining*

Data mining merupakan suatu teknik yang digunakan untuk menemukan pola-pola tersembunyi dalam data, serta sering diterapkan dalam berbagai metode statistika, *machine learning*, dan kecerdasan buatan (*artificial intelligence*) (Saputri, 2025). Teknik ini berfungsi untuk mengolah informasi dalam jumlah besar dan membentuk pola-pola yang menarik dari data yang telah ada. Hasil dari pengolahan tersebut dapat menghasilkan pengetahuan baru yang berasal dari data historis, yang selanjutnya disimpan dalam media seperti *database* atau data *warehouse*. Pengetahuan ini kemudian dapat dimanfaatkan untuk mendukung proses pengambilan keputusan secara lebih terarah dan berbasis data (Ernawati & Wahyuni, 2024).

Menurut Gupta dan Devanand (2013), *Data mining* memiliki beberapa metode utama yang disesuaikan dengan jenis data dan tujuan analisis. Beberapa metode tersebut meliputi klasifikasi, klasterisasi, asosiasi, regresi, dan peramalan.

1. Klasifikasi (*Classification*) merupakan salah satu metode yang paling umum digunakan dalam analisis data, terutama untuk memprediksi kategori atau kelas tertentu berdasarkan atribut yang diketahui. Metode ini termasuk dalam pendekatan *supervised learning*, karena proses pemodelannya menggunakan data yang telah memiliki label atau target yang jelas..
2. Klasterisasi (*Clustering*) atau segmentasi adalah metode *unsupervised* dalam *data mining* yang digunakan untuk mengidentifikasi kelompok alami dalam data berdasarkan kesamaan karakteristik. Tidak seperti klasifikasi, metode ini tidak menggunakan atribut target, melainkan memperlakukan semua atribut secara setara dalam proses pembelajaran. Klasterisasi sangat berguna dalam mengeksplorasi struktur data yang tidak teramati secara langsung. Dalam penelitian ini, metode klasterisasi digunakan untuk mengelompokkan negara-negara peserta PISA berdasarkan indikator literasi matematika dengan pendekatan GMM.

3. Asosiasi (*Association*) berfokus pada identifikasi hubungan antar item dalam dataset, yang sering digunakan dalam analisis keranjang pasar untuk mendeteksi produk-produk yang sering dibeli secara bersamaan.
4. Regresi (*Regression*) bertujuan untuk memprediksi nilai numerik berdasarkan hubungan antar variabel, bukan memetakan ke dalam kelas seperti pada klasifikasi. Contoh umum dari metode ini adalah *linear regression* dan *logistic regression*.
5. Peramalan (*Forecasting*) digunakan untuk memprediksi nilai di masa depan berdasarkan pola historis dalam data deret waktu. Teknik ini menggabungkan pendekatan statistik dan pembelajaran mesin untuk mengidentifikasi tren, musiman, dan *noise* dalam data.

2.2 Clustering

Clustering adalah suatu teknik untuk mengidentifikasi objek-objek yang serupa dengan mempertimbangkan beberapa kriteria tertentu, kemudian mengelompokkannya ke dalam beberapa klaster (Sari, 2020). Teknik ini termasuk ke dalam pendekatan *unsupervised learning*, karena tidak memerlukan label kelas sebelumnya. Tujuan utama *clustering* adalah untuk menemukan pola atau struktur tersembunyi dalam data melalui pengelompokan berdasarkan kemiripan karakteristik. Berbagai metode *clustering* yang umum digunakan antara lain *k-means*, *hierarchical clustering*, dan DBSCAN dan GMM, yang masing-masing memiliki karakteristik dan algoritma tersendiri. Setiap metode memiliki keunggulan tersendiri tergantung pada jenis data dan tujuan analisis (Gu, 2021).

Meskipun demikian *clustering* juga memiliki tantangan salah satunya adalah penentuan jumlah klaster yang optimal. Evaluasi performa algoritma *clustering* pun penting dilakukan menggunakan metrik seperti *silhouette score*, *Davies–Bouldin Index*, atau *Calinski–Harabasz Index*, agar hasil pengelompokan dapat diukur kualitas dan efektivitasnya (Ozturk & Demirel, 2023). Sementara itu, untuk model-model berbasis probabilistik, evaluasi yang lebih tepat dapat menggunakan BIC, karena metrik ini menyeimbangkan antara tingkat akurasi model dan jumlah parameter yang digunakan. Penerapan metode klasterisasi yang

sesuai dapat membantu mengelompokkan data berdasarkan kemiripan karakteristik, sehingga pola-pola tersembunyi dapat dikenali dan dimanfaatkan untuk mendukung analisis serta pengambilan keputusan yang lebih tepat.

Metode klasterisasi yang digunakan dalam penelitian ini, yakni GMM, dipilih karena kemampuannya dalam memodelkan distribusi data secara probabilistik dan mengakomodasi keberagaman karakteristik antar kelompok.

2.3 *Gaussian Mixture Model*

Gaussian Mixture Model (GMM) merupakan salah satu metode yang termasuk dalam kategori model berbasis distribusi dan digunakan dalam pendekatan *clustering* Deofanny et al. (2022). GMM merupakan metode klasterisasi *non-hierarki* yang memodelkan data berdasarkan distribusi *gaussian*, dengan mempertimbangkan parameter seperti *mean* dan variansi. Klaster dalam GMM dibentuk berdasarkan probabilitas fungsi kepadatan, sehingga memungkinkan pengelompokan data yang lebih fleksibel dan mampu menangkap struktur yang kompleks di dalam data (Saragih et al., 2024).

Dalam *Gaussian Mixture Model* (GMM), terdapat sejumlah komponen *gaussian* sebanyak K , yang merepresentasikan jumlah *cluster* yang ingin diidentifikasi. Setiap komponen *gaussian* ke- k memiliki bobot π_k , yang menunjukkan proporsi dari komponen tersebut dalam model dengan syarat sebagai berikut:

$$\sum_{k=1}^K \pi_k = 1 \quad (2.1)$$

Nilai bobot π_k dapat diperkirakan melalui proporsi jumlah data dalam masing-masing komponen, yaitu:

$$\pi_k = \frac{N_k}{N} \quad (2.2)$$

dimana N_k adalah jumlah data yang termasuk pada komponen ke- k dan N adalah total jumlah data. Setiap komponen *gaussian* dimodelkan menggunakan distribusi *gaussian multivariat* yang dinotasikan sebagai:

$$N(\mathbf{x} | \mu_k, \Sigma_k) \quad (2.3)$$

dengan μ_k sebagai vektor rata-rata dan Σ_k sebagai matriks kovarian dari komponen tersebut.

Fungsi densitas probabilitas total dari data \mathbf{x} yang merupakan campuran dari beberapa distribusi *gaussian*, dinotasikan sebagai $p(\mathbf{x})$. Dengan seluruh parameter tersebut, fungsi distribusi total GMM dapat dituliskan sebagai berikut:

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \cdot N(\mathbf{x} | \mu_k, \Sigma_k) \quad (2.4)$$

Model ini memungkinkan setiap data memiliki probabilitas keanggotaan pada lebih dari satu klaster, sehingga memberikan fleksibilitas dalam pengelompokan data yang tidak terpisah secara tegas. Pendekatan ini banyak digunakan dalam berbagai bidang, termasuk pengenalan pola, klasifikasi suara, dan segmentasi data (Alqahtani & Kalantan, 2020).

GMM memiliki tiga parameter, yaitu *mean*, kovarian dan koefisien campuran dari seluruh komponen GMM. Ketiga parameter ini akan diestimasi secara optimal menggunakan algoritma *Expectation Maximization* (EM). Algoritma ini memanfaatkan data yang tersedia untuk mencari nilai optimal dari variabel, kemudian menentukan parameter model. Menurut Imro'ah dkk (2022), setiap iterasi algoritma EM mencakup dua tahapan, yaitu tahap Ekspektasi (*E-Step*) yang digunakan untuk mencari suatu fungsi yaitu ekspektasi dari fungsi *log-likelihood* yang dinotasikan dengan:

$$E[\log[L(\theta)] | \mathbf{x}_n, \hat{\theta}^{r-1}] \quad (2.5)$$

Dimana $r-1$ menandakan parameter sebelum diestimasi. Persamaan *log-likelihood* pada proses *E-Step* diturunkan terhadap $\hat{\theta}$ sehingga nilai taksiran dari n_k^r, μ_{ik}^r dan σ_{ik}^r diperoleh pada masing-masing parameter. Tahap berikutnya yaitu Maksimisasi (*M-Step*), pada tahap ini nilai parameter pada model *gaussian*, seperti rata-rata, variansi, dan bobot, diperbarui secara iteratif berdasarkan hasil perhitungan yang diperoleh dari langkah *Expectation* (*E-step*).

Pada implementasinya, metode GMM menggunakan paket *mclust* yang menyediakan beberapa variasi model kovarians untuk membedakan klaster

berdasarkan volume, bentuk, dan orientasi distribusi *gaussian* (Scrucca et al., 2016).

Ketiga komponen tersebut memiliki fungsi berbeda namun saling berkaitan. Volume menunjukkan besar kecilnya sebaran data dalam kluster; volume sama berarti ukuran sebaran antar kluster setara, sedangkan volume berbeda menunjukkan variasi ukuran. Bentuk menggambarkan perbandingan penyebaran data antar variabel; bentuk sama menunjukkan pola sebaran seragam, sedangkan bentuk berbeda menandakan penyebaran data yang bervariasi. Orientasi menunjukkan arah atau rotasi kluster terhadap sumbu koordinat; orientasi sama berarti arah penyebaran antar kluster serupa, sedangkan orientasi berbeda menunjukkan arah penyebaran yang khas sesuai struktur data. Tabel berikut merangkum model yang digunakan:

Tabel 2.1 Variasi model dan struktur kovarians pada gmm

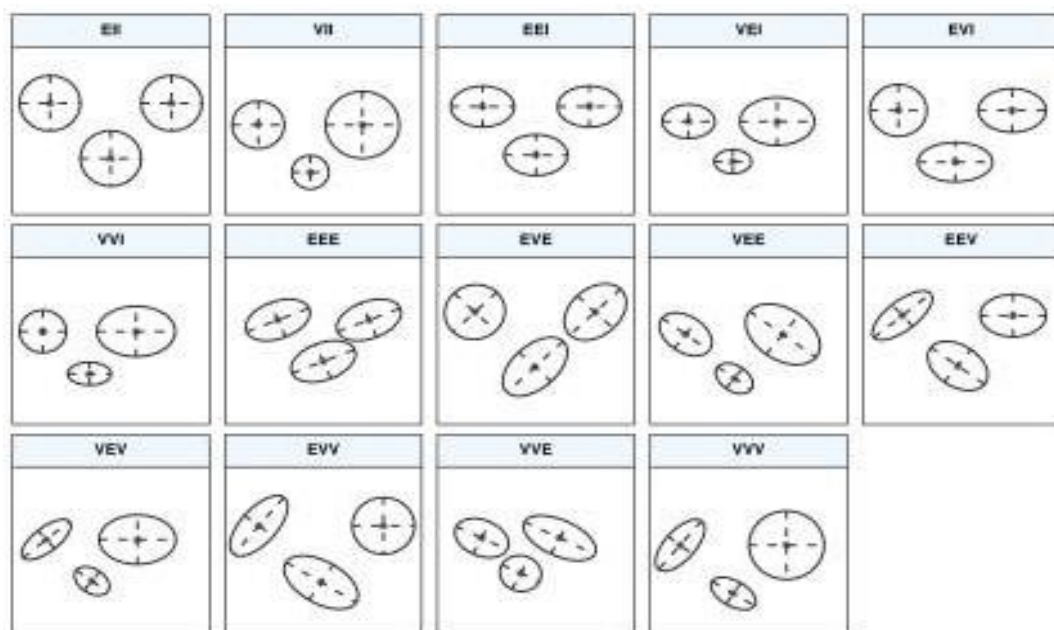
Model	Distribution	Volume	Shape	Orientation
EII	Bola (Spherical)	Sama	Sama	—
VII	Bola (Spherical)	Berbeda	Sama	—
EEI	Diagonal	Sama	Sama	Sumbu koordinat
VEI	Diagonal	Berbeda	Sama	Sumbu koordinat
EVI	Diagonal	Sama	Berbeda	Sumbu koordinat
VVI	Diagonal	Berbeda	Berbeda	Sumbu koordinat
EEE	Elipsoidal	Sama	Sama	Sama
EVE	Elipsoidal	Sama	Berbeda	Sama
VEE	Elipsoidal	Berbeda	Sama	Sama
VVE	Elipsoidal	Berbeda	Berbeda	Sama
EEV	Elipsoidal	Sama	Sama	Berbeda
VEV	Elipsoidal	Berbeda	Sama	Berbeda
EVV	Elipsoidal	Sama	Berbeda	Berbeda
VVV	Elipsoidal	Berbeda	Berbeda	Berbeda

Sumber: Scrucca et al. (2016)

Tabel 2.1 menyajikan variasi model GMM beserta struktur kovariansnya yang digunakan dalam analisis kluster. Struktur kovarians pada tiap model menentukan bentuk, ukuran (volume), dan orientasi kluster dalam ruang

multidimensi. Tujuan penyajian model ini adalah memberikan gambaran umum mengenai seluruh variasi model yang tersedia agar pembaca memahami karakteristik setiap model sebelum dilakukan pemilihan model terbaik sesuai dengan data penelitian.

Perlu diperhatikan bahwa pada model EII dan VII, kolom orientasi keduanya dikosongkan karena kluster pada kedua model tersebut berbentuk bola (spherical). Bentuk bola memiliki simetri sempurna sehingga tidak memiliki arah atau kemiringan tertentu yang dapat dibandingkan antar-kluster



Gambar 2.1 Variasi model *gaussian* pada paket *mclust*

Sumber: Scrucca et al. (2016)

Model EII dan VII tergolong spherical, dengan kluster berbentuk bola tanpa mempertimbangkan arah orientasi. EII mengasumsikan seluruh kluster memiliki volume sama, sedangkan VII memperbolehkan volume berbeda. Pada kedua model ini, komponen bentuk dan orientasi tidak relevan karena sebaran datanya simetris sempurna dan tidak memiliki arah rotasi. Model EEI, VEI, EVI, dan VVI termasuk diagonal dengan orientasi sejajar sumbu koordinat. EEI mengasumsikan volume dan bentuk sama; VEI memperbolehkan variasi volume; EVI memperbolehkan variasi bentuk; sedangkan VVI mengizinkan variasi pada keduanya.

Delapan model berikutnya, yaitu EEE, EVE, VEE, VVE, EEV, VEV, EVV, dan VVV, merupakan model ellipsoidal yang mempertimbangkan orientasi kluster. EEE menggambarkan volume, bentuk, dan orientasi yang sama; sedangkan VVV adalah model paling fleksibel karena ketiganya berbeda antar kluster. Model EVE memiliki volume dan orientasi sama tetapi bentuk berbeda; VEE memiliki volume berbeda dengan bentuk dan orientasi sama; VVE memiliki volume dan bentuk berbeda namun orientasi sama; EEV memiliki volume dan bentuk sama tetapi orientasi berbeda; VEV mengizinkan volume dan orientasi berbeda dengan bentuk sama; dan EVV memiliki volume sama tetapi bentuk serta orientasi berbeda.

Proses klasterisasi menggunakan GMM perlu memperhatikan penentuan jumlah kluster yang optimal. Oleh karena itu diperlukan suatu pendekatan yang dapat menilai keseimbangan antara kecocokan model terhadap data dan kompleksitas model yang digunakan. Salah satu metode yang digunakan untuk menentukan kluster optimal adalah *Bayesian Information Criterion* (BIC). Persamaan umum BIC dinyatakan sebagai berikut:

$$BIC = -2L + k \cdot \ln(N) \quad (2.6)$$

Dengan L merupakan fungsi *log-likelihood*, k adalah jumlah parameter, dan N adalah jumlah data.

Secara umum, literatur sering menyebutkan bahwa model terbaik dipilih berdasarkan nilai BIC terkecil. Namun, dalam implementasi GMM menggunakan paket *mclust* pada perangkat lunak R, pemilihan model terbaik justru berdasarkan nilai BIC tertinggi. Dalam konteks ini, model yang memiliki nilai BIC paling tinggi dibandingkan dengan model lain pada setiap iterasi dipilih sebagai model yang paling optimal (Hasnida & Kusumawati, 2023). Perbedaan arah interpretasi tersebut pada dasarnya tidak mengubah makna utama, karena baik dalam literatur umum maupun pada *mclust*, tujuan akhirnya tetap sama, yaitu memperoleh model yang paling sesuai dengan data dengan mempertimbangkan keseimbangan antara kecocokan model dan kompleksitas parameter keseimbangan antara kecocokan model dan kompleksitas parameter.

2.4 Literasi Matematika dalam PISA 2022

Salah satu aspek penting dalam menilai kemampuan peserta didik dalam konteks internasional adalah literasi matematika. Literasi ini menjadi indikator utama untuk melihat sejauh mana individu mampu menghadapi tantangan kehidupan melalui pemahaman dan penerapan konsep matematika. Instrumen internasional yang digunakan untuk mengukur kemampuan literasi matematika siswa berusia 15 tahun secara berkala setiap tiga tahun adalah *Programme for International Student Assessment* (PISA) yang diselenggarakan oleh OECD.

Kemampuan literasi matematika tidak hanya mencakup keterampilan dalam memahami konsep-konsep matematika, tetapi juga melibatkan proses merumuskan, menerapkan, dan menafsirkan matematika dalam berbagai situasi kehidupan nyata. Selain itu, literasi matematika mencakup penguasaan terhadap konsep, prosedur, fakta, dan alat yang digunakan untuk menjelaskan serta memprediksi fenomena, sekaligus menumbuhkan kesadaran akan peran penting matematika dalam kehidupan sehari-hari dan dalam pengambilan keputusan (Afni & Hartono, 2020).

Menurut Kusumawardani et al. (2018), indikator kemampuan literasi matematika dalam kerangka PISA mencakup berbagai aspek penting, seperti kemampuan berpikir dan bernalar secara matematis, membangun argumentasi, serta berkomunikasi dalam bahasa matematika. Indikator lainnya meliputi kemampuan dalam pemodelan, merumuskan dan menyelesaikan masalah, menggunakan representasi yang sesuai, mengoperasikan simbol serta bahasa matematika secara formal dan teknis, serta memanfaatkan alat-alat matematika secara efektif.

Tingkat kemahiran literasi matematika dalam PISA mengalami pembaruan pada siklus 2022. Jika pada edisi-edisi sebelumnya, khususnya sejak tahun 2012, hanya terdapat enam tingkatan kemahiran, maka pada PISA 2022 jumlah tersebut bertambah menjadi total ada delapan tingkatan.

Setiap tingkat kemahiran merepresentasikan rentang skor tertentu yang menggambarkan kemampuan siswa dalam menyelesaikan soal dengan tingkat kompleksitas yang berbeda. Tingkat kemahiran yang lebih tinggi mencerminkan kemampuan siswa dalam menyelesaikan soal yang lebih kompleks, sedangkan

tingkat yang lebih rendah menunjukkan bahwa siswa masih berada pada tahap penguasaan dasar. Tabel berikut menyajikan rentang skor untuk masing-masing tingkat kemahiran dalam domain matematika berdasarkan hasil PISA 2022.

Tabel 2.2 Tingkat kemahiran literasi matematika pisa 2022

Tingkat Kemahiran	Literasi Matematika
Level 1 (1c)	$\text{skor} < 295,47$
Level 1 (1b)	$295,47 \leq \text{skor} < 357,77$
Level 1 (1a)	$357,77 \leq \text{skor} < 420,07$
Level 2	$420,07 \leq \text{skor} < 482,38$
Level 3	$482,38 \leq \text{skor} < 544,68$
Level 4	$544,68 \leq \text{skor} < 606,99$
Level 5	$606,99 \leq \text{skor} < 669,30$
Level 6	$\text{skor} \geq 669,30$

Sumber: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2022.

Hasil PISA 2022 menunjukkan bahwa skor rata-rata literasi matematika siswa Indonesia adalah 366 poin, yang menempatkan mereka pada level 1a dalam skala kemahiran PISA. Level ini berada di bawah tingkat kompetensi minimum, yaitu level 2, yang digunakan secara internasional untuk menunjukkan kemampuan dasar dalam menerapkan konsep matematika dalam konteks kehidupan nyata. Kondisi ini menunjukkan bahwa peningkatan mutu pembelajaran matematika, khususnya pada aspek kompetensi dasar, harus menjadi prioritas strategis dalam kebijakan pendidikan nasional (OECD, 2023).

2.5 Indikator Penelitian

Indikator penelitian dalam studi ini diperoleh dari data kuesioner PISA 2022 yang mencerminkan dimensi kognitif dan afektif dalam literasi matematika. Pemilihan indikator didasarkan pada relevansinya dengan aktivitas pembelajaran. Keempat indikator yang digunakan dijelaskan lebih lanjut pada bagian berikutnya.

2.5.1 Pemahaman Informasi Matematika dari Visualisasi

Memahami informasi matematika dari grafik, diagram, atau simulasi merupakan salah satu keterampilan penting dalam literasi matematika. PISA 2022 menekankan pentingnya kemampuan ini sebagai bagian dari kompetensi interpretatif siswa dalam menyelesaikan persoalan kontekstual berbasis data visual. Dalam konteks ini, siswa diminta untuk mengamati, menafsirkan, dan menghubungkan informasi yang disajikan secara grafis dengan konsep matematika yang relevan. Berdasarkan hasil penelitian yang dikutip oleh Kurniawati & Juandi (2023), siswa dengan gaya belajar visual cenderung lebih mampu menyampaikan ide melalui bentuk visual seperti histogram, diagram garis, dan tabel. Sebaliknya, siswa dengan gaya belajar auditori atau kinestetik menunjukkan kecenderungan lebih rendah dalam menggunakan representasi visual matematika.

2.5.2 Interpretasi Solusi Matematika dalam Konteks Kehidupan Nyata

Interpretasi solusi matematika dalam konteks soal kontekstual melibatkan proses memahami informasi yang diberikan dalam bentuk cerita nyata, kemudian mengubahnya menjadi model dan strategi penyelesaian matematis. Berdasarkan temuan (Tuhumury et al., 2024), siswa menunjukkan kemampuan tersebut dengan baik. Peneliti menyatakan bahwa subjek dapat menjelaskan permasalahan yang ada pada semua soal dengan kalimat matematika yang baik dan benar, yang mencerminkan proses menginterpretasikan konteks menjadi solusi matematis.

Interpretasi solusi kontekstual membantu siswa menghubungkan konsep abstrak dengan situasi kehidupan sehari-hari, dan menjadikan matematika lebih aplikatif. OECD juga menekankan bahwa paparan terhadap konteks dunia nyata penting dalam mendorong pemahaman mendalam, serta mendukung pemilihan strategi penyelesaian masalah yang tepat.

2.5.3 Kepercayaan Diri dalam Memahami Informasi Visual Matematika

Kepercayaan diri berkaitan erat dengan paparan terhadap praktik pembelajaran yang menekankan aktivasi kognitif dan pemecahan masalah visual. OECD mencatat bahwa kepercayaan diri yang tinggi berhubungan dengan kinerja

matematika yang lebih baik, serta kecenderungan untuk terlibat aktif dalam proses belajar yang menantang.

Faktor-faktor seperti pengalaman belajar sebelumnya dan lingkungan belajar yang mendukung turut membentuk persepsi siswa terhadap kemampuannya dalam memahami informasi visual matematis. Ketika siswa terbiasa menghadapi representasi visual seperti grafik, diagram, atau simbol matematis dalam konteks yang bermakna, mereka akan lebih mudah membangun keyakinan terhadap kemampuannya sendiri. Sebaliknya, kurangnya paparan terhadap tugas visual atau ketidakpahaman terhadap representasi tersebut dapat menurunkan rasa percaya diri siswa.

Yunita et al. (2020) menyatakan bahwa melalui pembelajaran berbasis masalah terhadap kasus visual matematis, terjadi peningkatan *self-efficacy* siswa dalam menggunakan visual thinking matematis. Hal ini menunjukkan bahwa siswa yang memiliki persepsi diri yang kuat cenderung lebih percaya diri dalam mengekstraksi informasi dari representasi visual matematika.

2.5.4 Selisih Kepercayaan Diri Tinggi dan Rendah terhadap Keterampilan Matematika Abad ke-21

Indeks persepsi diri terhadap keterampilan matematika abad ke-21 disusun untuk menangkap kesenjangan antara siswa dengan tingkat kepercayaan diri tinggi dan rendah. PISA 2022 mencatat bahwa lebih dari separuh siswa yang percaya diri (54%) menunjukkan penguasaan tinggi dalam tugas-tugas seperti visualisasi dan pemecahan masalah dunia nyata, dibandingkan hanya 38% dari mereka yang tidak percaya diri.

Selisih yang signifikan ini menunjukkan bahwa keyakinan terhadap kemampuan diri memiliki implikasi penting dalam proses belajar. Perbedaan persepsi ini menjadi indikator penting dalam mengidentifikasi potensi ketimpangan dalam efektivitas pembelajaran antar kelompok siswa, terutama dalam konteks kesiapan menghadapi tantangan abad ke-21 yang menuntut literasi data dan matematika secara luas.

Berdasarkan penelitian yang dilakukan oleh (Semilarski et al., 2021), siswa menunjukkan peningkatan rasa percaya diri terhadap keterampilan abad ke-21, terutama dalam kemampuan berpikir kritis, pemecahan masalah, dan penerapan pengetahuan matematika dalam konteks kehidupan nyata. Peningkatan ini muncul setelah siswa mengikuti pembelajaran yang berbasis pada situasi kehidupan sehari-hari, yang mendorong keterlibatan aktif dalam pemecahan masalah secara bermakna dan aplikatif.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai pengelompokan negara-negara peserta PISA 2022 berdasarkan indikator literasi matematika dengan metode *Gaussian Mixture Model* (GMM), diperoleh kesimpulan sebagai berikut:

- a) Metode *Gaussian Mixture Model* (GMM) berhasil diterapkan untuk mengelompokkan negara-negara peserta PISA 2022 berdasarkan skor literasi matematika. Hasil analisis menunjukkan bahwa model terbaik yang diperoleh adalah model EEI (diagonal, volume dan bentuk sama) dengan tiga kluster sebagai struktur pengelompokan yang paling sesuai dengan data.
- b) Berdasarkan hasil penerapan metode *Gaussian Mixture Model* (GMM), karakteristik masing-masing kluster negara peserta PISA 2022 berdasarkan indikator literasi matematika menunjukkan pola yang berbeda pada tiap variabel penelitian.
 1. Kluster 1 mencerminkan negara-negara dengan rata-rata tertinggi pada seluruh variabel penelitian, ditandai dengan kemampuan memahami representasi visual matematika dan kepercayaan diri siswa yang baik serta pemerataan kemampuan antar siswa yang relatif merata.
 2. Kluster 2 menggambarkan kelompok negara dengan pola rata-rata menengah dan distribusi kemampuan yang seimbang, di mana siswa masih mampu mengaitkan konsep matematika dengan konteks nyata meskipun pemanfaatan alat bantu visual belum optimal.
 3. Kluster 3 menunjukkan kelompok negara dengan pemahaman yang cukup baik terhadap visualisasi matematika, namun tingkat kepercayaan diri siswa dalam menginterpretasi solusi matematika dan menerapkannya pada konteks kehidupan nyata masih rendah, sehingga capaian literasi matematikanya relatif rendah dibandingkan kluster lainnya

5.2 Saran

Berdasarkan hasil penelitian yang telah diperoleh, beberapa saran yang dapat diajukan antara lain:

- a) Bagi penelitian selanjutnya, disarankan untuk memasukkan variabel tambahan dari domain lain seperti membaca dan sains, serta faktor kontekstual seperti latar belakang sosial ekonomi atau strategi pembelajaran, agar hasil pengelompokan lebih komprehensif.
- b) Bagi pihak terkait dalam bidang pendidikan, hasil pengelompokan berdasarkan data kuesioner literasi matematika ini dapat dijadikan bahan pertimbangan awal untuk memahami pola karakteristik siswa dari berbagai negara. Temuan ini dapat membantu dalam merancang pendekatan pembelajaran atau kebijakan pendidikan yang lebih sesuai dengan kondisi dan kebutuhan siswa sebagaimana tergambar dari respon kuesioner, bukan dari hasil tes semata.

DAFTAR PUSTAKA

- Afni, N., & Hartono. (2020). Contextual Teaching and Learning (CTL) as a strategy to improve students mathematical literacy. *Journal of Physics: Conference Series*, 1581(1). <https://doi.org/10.1088/1742-6596/1581/1/012043>
- Airlangga, G. (2025). A Comparative analysis of clustering algorithms for expedia's rtavel dataset. *sinkron*, 9(1), 476–483. <https://doi.org/10.33395/sinkron.v9i1.14343>
- Alqahtani, N. A., & Kalantans, Z. I. (2020). Gaussian mixture models based on principal components and applications. *mathematical problems in engineering*, 2020. <https://doi.org/10.1155/2020/1202307>
- Andayani, S., Retnani, N., Yusri, T. A. S., & Marwoto, B. S. H. (2025). A comparative analysis of dbscan and gaussian mixture model for clustering indonesian provinces based on socioeconomic welfare indicators. *Barekeng*, 19(3), 2039–2056. <https://doi.org/10.30598/barekengvol19iss3pp2039-2056>
- Anzilah, W; Amir, M. F. (2021). Kemampuan literasi matematika siswa sekolah dasar melalui implementasi pembelajaran PAR (Preparation–Assistance–Reflection). *Jurnal Cendekia:S Jurnal Pendidikan Matematika*. <https://doi.org/10.31004/cendekia.v5i1.483>
- Azid, A., Zamnah, L. N., & Solihah, S. (2023). Mengapa literasi matematis penting dan diperhatikan? *Prosiding Galuh Mathematics National Conference*, 3(1), 7–10.
- Bishop, C. M. (2006). Pattern recognition and machine learning. In *Journal of Machine and Computing* (Vol. 4, Issue 1). Springer. <https://doi.org/10.53759/7669/jmc202404020>
- Deofanny, N. F., Rohmawati, A. A., & Indwiarti. (2022). Model gaussian mixture pada distribusi kecepatan angin dengan algoritma Em. *E-Proceedings of Engineering*, 9(3), 1978–1984.
- Ernawati, A., & Wahyuni, S. (2024). Analisis data mining pola penggunaan seluler dan klasifikasi perilaku pengguna di berbagai perangkat menggunakan metode c4 . 5. *Bulletin of Information Technology (BIT)*, 5(4), 162–168.

<https://doi.org/10.47065/bit.v5i2.1689>

- Greve, B., Pigeot, I., Huybrechts, I., Pala, V., & Börnhorst, C. (2016). A comparison of heuristic and model-based clustering methods for dietary pattern analysis. *Public Health Nutrition*, 19(2), 255–264. <https://doi.org/10.1017/S1368980014003243>
- Gu, J. (2021). Comparative analysis based on clustering algorithms. *Journal of Physics: Conference Series*, 1994(1). <https://doi.org/10.1088/1742-6596/1994/1/012024>
- Gupta, V., & Devanand, P. (2013). A survey on data mining : tools, techniques, applications, trends and issues . *Internation Journal of Scientific & Engineering Research*, 4(3), 1–14.
- Saputri, E. (2025). *IT-EXPLORE Teknik dan aplikasi data mining di Indonesia : tinjauan literatur satu dekade (2015-2024)*. 04, 138–149. <https://doi.org/10.24246/itexplore.v4i2.2025.pp138-149>
- Hasnida, I. S. D., & Kusumawati, R. (2023). Penerapan model-based clustering pada pengelompokan saham berdasarkan rasio keuangan. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 15(1), 37–50. <https://doi.org/10.34123/jurnalasks.v15i1.510>
- Kurniawati, R., & Juandi, D. (2023). Systematic literature review : kemampuan representasi visual matematis pada pembelajaran matematika. *ALGORITMA: Journal of Mathematics Education*, 5(1), 26–36. <https://doi.org/10.15408/ajme.v5i1.32603>
- Kusumawardani, D. R., Wardono, & Kartono. (2018). Pentingnya penalaran matematika dalam meningkatkan kemampuan literasi matematika [The importance of mathematical reasoning in improving mathematical literacy skills]. *PRISMA, Prosiding Seminar Nasional Matematika*, 1(1), 588–595.
- Hamdiyanti, M., Rodiana, I., Laelasar, & Subroto, T. (2024). Systematic literature review: mathematical literacy skills in terms of mathematics learning motivation. *IJCER (International Journal of Chemistry Education Research)*, 8, 104–112. <https://doi.org/10.20885/ijcer.vol8.iss2.art3>
- OECD. (2023). PISA 2022 results indonesia. *factsheets*, 1–9.

- <https://www.oecd.org/publication/pisa-2022-results/country-notes/malaysia-1dbe2061/>
- Ozturk, F. E., & Demirel, N. (2023). *Ver i b i l i m i d e r g i s i*. 6(1), 34–45
- PISA. (2023). Hasil PISA 2022 (Volume I). In *Oecd*.
https://www.oecd.org/en/publications/pisa-2022-results-volume-i_53f23881-en.html
- Putrawangsa, S., & Hasanah, U. (2022). Analisis capaian siswa indonesia pada pisa dan urgensi kurikulum berorientasi literasi dan numerasi bagaimana trend capaian tersebut? dan sejauh mana perubahan kurikulum selama ini berdampak pada. *Jurnal Studi Pendidikan Dan Pembelajaran*, 1(1), 1–12.
- Saragih, A. R., Arnita., Pinem, M. Z., Suhendra, D. P. S., & Siregar, N. A. (2024). Pengelompokan data penjualan produk monitor di amazon menggunakan algoritma gaussian mixture model. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKSI)*, 10(11), 10–15.
- Sari, M. R. (2020). Implementation of data mining using clustering methods for analysis of dangerous disease data. *International Journal of Research and Review (Ijrrjournal.Com)*, 7(April), 4
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *R Journal*, 8(1), 289–317. <https://doi.org/10.32614/rj-2016-021>
- Semilarski, H., Soobard, R., & Rannikmäe, M. (2021). Promoting students' perceived self-efficacy towards 21st century skills through everyday life-related scenarios. *Education Sciences*, 11(10).
<https://doi.org/10.3390/educsci11100570>
- Tuhumury, J. F., Moma, L., & Huwaa, N. C. (2024). Analisis representasi matematis peserta didik dalam menyelesaikan soal kontekstual. *Jurnal Pendidikan Matematika Unpatti*, 5(1), 28–34.
<https://doi.org/10.30598/jpmunpatti.v5.i1.p28-34>
- Yessica, F., Kusnandar, D., & Imro'ah, N. (2022). Implementasi metode latent class cluster analysis dalam pengelompokan wilayah berdasarkan indikator indeks pembangunan manusia. *Buletin Ilmiah Math. Stat Dan Terapannya*

(Bimaster), 11(2), 213–220.

Yunita, M. R., Surya, E., & Syahputra, E. (2020). Pengembangan perangkat pembelajaran berbasis masalah untuk kemampuan visual thinking matematis dan self efficacy siswa kelas vii smp swasta raja garuda mas besitang. *Paradikma: Jurnal Pendidikan Matematika*, 13(2), 18–29. <https://doi.org/10.24114/paradikma.v13i3.22913>