



Pemetaan Kemampuan Matematis Peserta *International Kangaroo Mathematics Contest* Menggunakan *K-Means Clustering*

Elke Annisa Octaria^a, Raden Ridwan Hasan Saputra^b

^{a,b}Pendidikan Matematika, Universitas Ibn Khaldun Bogor, Indonesia

Abstrak

Heterogenitas kemampuan matematis siswa Indonesia memerlukan pendekatan berbasis data untuk merancang pembelajaran yang lebih tepat sasaran. Penelitian ini bertujuan memetakan kemampuan matematis peserta Indonesia dalam *International Kangaroo Mathematics Contest* (IKMC) 2024 menggunakan *K-Means Clustering* pada enam tingkatan: *Pre Ecolier* (kelas 1-2), *Ecolier* (kelas 3-4), *Benjamin* (kelas 5-6), *Cadet* (kelas 7-8), *Junior* (kelas 9-10), dan *Student* (kelas 11-12). Data 92.529 peserta dianalisis melalui pembersihan data, standardisasi *z-score*, penentuan *cluster* optimal dengan *Elbow Method* dan *Silhouette Score*, kemudian pemodelan *K-Means* secara global dan per tingkatan. Hasil menunjukkan tiga *cluster* optimal pada setiap tingkatan dengan *Silhouette Score* 0.58-0.64, mengindikasikan kualitas *clustering* yang baik. Secara global, 48.5% peserta berada pada kategori rendah, 41.6% sedang, dan 9.9% tinggi. Analisis per tingkatan mengungkap dominasi *cluster* rendah-sedang (80-90%) dengan gap skor melebar dari 33,5 poin (*Pre Ecolier*) menjadi 44,2 poin (*Cadet*), menunjukkan divergensi kemampuan yang semakin tajam pada jenjang lebih tinggi. *Clustering* berbasis data kompetisi terbukti efektif memetakan heterogenitas kemampuan matematis secara representatif dan memberikan dasar empiris untuk pengembangan pembelajaran terdiferensiasi serta pembinaan matematika berbasis data.

Kata kunci: *K-Means Clustering*; kemampuan matematis; *International Kangaroo Mathematics Contest*; pembelajaran terdiferensiasi

Abstract

The heterogeneity of Indonesian students' mathematical abilities necessitates a data-driven approach to design more targeted learning strategies. This study aims to map the mathematical abilities of Indonesian participants in the 2024 International Kangaroo Mathematics Contest (IKMC) using K-Means Clustering across six levels: Pre Ecolier (grades 1-2), Ecolier (grades 3-4), Benjamin (grades 5-6), Cadet (grades 7-8), Junior (grades 9-10), and Student (grades 11-12). Data from 92,529 participants were analyzed through data cleaning, z-score standardization, optimal cluster determination using Elbow Method and Silhouette Score, followed by global and per-level K-Means modeling. Results revealed three optimal clusters at each level with Silhouette Scores of 0.58-

Submitted: 01-01-2026 Approved: 10-01-2026. Published: 13-01-2026

Corresponding author's e-mail: elke@uika-bogor.ac.id

ISSN: Print 2722-1504 | ONLINE 2721-1002

<https://ejournal.uika-bogor.ac.id/index.php/jpg/index>

0.64, indicating good clustering quality. Globally, 48.5% of participants were in the low category, 41.6% medium, and 9.9% high. Per-level analysis uncovered dominance of low-medium clusters (80-90%) with score gaps widening from 33.5 points (*Pre Ecolier*) to 44.2 points (*Cadet*), showing increasingly sharp ability divergence at higher grades. Competition data-based clustering proved effective in representatively mapping mathematical ability heterogeneity and provides an empirical basis for developing differentiated instruction and data-driven mathematics coaching.

Keywords: K-Means Clustering; mathematical ability; International Kangaroo Mathematics Contest; differentiated instruction

INTRODUCTION

Kompetisi matematika tingkat internasional seperti *International Kangaroo Mathematics Contest* (IKMC) berperan penting dalam menumbuhkan budaya matematika sekaligus memberikan gambaran mengenai pola kemampuan matematis siswa di berbagai negara. Kompetisi ini dimulai di Prancis pada tahun 1991 dengan nama *Kangourou*, yang terinspirasi dari *Australian Mathematics Competition*. Saat ini, IKMC telah berkembang pesat dengan partisipasi lebih dari enam juta siswa dari hampir 90 negara yang bertanding secara serentak setiap Kamis ketiga bulan Maret (Akveld et al., 2020). Berbeda dengan olimpiade matematika yang umumnya berfokus pada pencarian talenta matematis unggul, IKMC mengusung pendekatan yang lebih inklusif dengan tujuan mempopulerkan matematika dan menghadirkan pengalaman belajar yang menyenangkan bagi seluruh siswa. Hal ini diwujudkan melalui soal-soal non-rutin yang menekankan pemikiran logis, kreativitas, dan strategi pemecahan masalah, dengan cakupan materi meliputi geometri, teori bilangan, kombinatorika, dan aljabar (*Association Kangourou Sans Frontières*, 2024). Di sisi lain, hasil PISA 2022 menunjukkan bahwa capaian matematika siswa Indonesia masih berada di bawah rata-rata OECD dan ditandai oleh kesenjangan kemampuan yang cukup besar antar kelompok siswa (OECD, 2023). Berbagai penelitian menegaskan bahwa peningkatan kemampuan matematis tidak ditentukan oleh kompleksitas materi yang diajarkan, tetapi juga oleh pendekatan pembelajaran yang kontekstual dan partisipatif. Tungson dan Saputra (2024) menunjukkan bahwa pendekatan matematika nalaria realistik yang dipadukan dengan strategi pembelajaran adaptif mampu meningkatkan partisipasi dan hasil belajar siswa. Temuan tersebut menegaskan pentingnya pemetaan kemampuan matematis siswa secara objektif sebagai dasar untuk merancang pembelajaran yang lebih tepat dan berbasis data.

Penelitian tentang analisis kemampuan matematis menggunakan pengelompokan data (*clustering*) berkembang pesat dalam beberapa tahun terakhir sebagai bagian dari *Educational Data Mining* (EDM). Romero dan Ventura (2020) menyatakan bahwa teknik ini dapat digunakan untuk mengidentifikasi pola tersembunyi dalam data pendidikan tanpa harus mengikuti kategori yang telah ditentukan sebelumnya. Xu et al. (2019) menggunakan K-Means *Clustering* untuk menganalisis 15.000 siswa di China dan mengidentifikasi empat profil kemampuan berbeda yang tidak terdeteksi melalui analisis

konvensional. Dutt et al. (2017) menemukan bahwa *clustering* dapat mengidentifikasi siswa berisiko dengan akurasi 15-20% lebih tinggi dibanding metode tradisional. Dalam konteks kompetisi, Freiman dan Leikin (2021) menemukan tiga profil berbeda dalam strategi pemecahan masalah: pemikir algoritmik, penalar visual-spasial, dan pemecah masalah kreatif. Han dan Kim (2023) membandingkan beragam algoritma dan menemukan K-Means memberikan hasil paling stabil dan mudah diinterpretasikan untuk dataset besar.

Namun demikian, penelitian-penelitian tersebut memiliki beberapa keterbatasan yang memberikan peluang untuk melakukan riset lebih lanjut. Pertama, mayoritas studi berfokus pada data seperti nilai rapor, ujian nasional, atau tes terstandar. Data kompetisi internasional yang dirancang khusus untuk mengukur keterampilan berpikir tingkat tinggi, kreativitas matematis, dan pemecahan masalah non-rutin masih sangat jarang diteliti (Liao et al., 2021). Sementara itu, soal-soal IKMC memiliki karakteristik unik yang menekankan pemikiran logis, keterampilan dalam komputasi, imajinasi, dan beragam strategi pemecahan masalah yang berbeda (Akveld et al., 2020). Kedua, analisis komparatif antar jenjang pendidikan secara bersamaan jarang dilakukan, kebanyakan penelitian hanya menganalisis satu atau dua tingkatan saja atau tanpa adanya perbedaan jenjang, padahal pemahaman tentang bagaimana struktur kemampuan matematis berkembang dari pendidikan dasar awal hingga menengah atas sangat krusial dalam perkembangan pembelajaran matematika (Schoenfeld, 2016; Geary et al., 2019). Ketiga, penelitian berbasis *clustering* pada domain matematika di Indonesia masih relatif terbatas, khususnya yang memanfaatkan data kompetisi matematika internasional. Setiawan dan Kusumah (2021) menerapkan *K-Means Clustering* pada data Ujian Nasional matematika, namun analisis difokuskan pada tingkat wilayah sehingga belum mampu menggambarkan profil kemampuan individual siswa. Sementara itu, Nuraeni et al. (2023) menggunakan pendekatan pengelompokan untuk menganalisis kemampuan pemecahan masalah matematis, tetapi terbatas pada satu jenjang kelas dengan ukuran sampel yang relatif kecil sehingga generalisasi temuan masih terbatas. Dalam konteks kompetisi internasional, Saputra dan Rahmah (2025) telah mengkaji performa siswa sekolah Islam Indonesia pada IKMC dan menunjukkan adanya variasi capaian *skor* antar peserta. Namun, penelitian tersebut masih bersifat deskriptif dan belum menggali struktur pengelompokan kemampuan matematis secara mendalam. Oleh karena itu, diperlukan kajian lanjutan yang memanfaatkan pendekatan *unsupervised learning* untuk mengidentifikasi pola pengelompokan kemampuan matematis secara lebih objektif dan komprehensif.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk memetakan kemampuan matematis peserta Indonesia IKMC 2024 menggunakan *K-Means Clustering* pada enam tingkatan, mengidentifikasi dan mengkarakterisasi *cluster* optimal yang terbentuk, serta menganalisis pola perkembangan kemampuan antartingkatan sebagai dasar pembelajaran terdiferensiasi dan pembinaan olimpiade.

METHOD

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data peserta *International Kangaroo Mathematics Contest* (IKMC) 2024 untuk membangun model pengelompokan kemampuan matematis berbasis *K-Means Clustering*. Dataset mencakup skor 92.529 peserta yang terdistribusi ke dalam enam tingkatan kompetisi, yaitu *Pre Ecolier* (kelas 1–2), *Ecolier* (kelas 3–4), *Benjamin* (kelas 5–6), *Cadet* (kelas 7–8), *Junior* (kelas 9–10), dan *Student* (kelas 11–12). Data yang digunakan memuat informasi kelas, sekolah, kota asal, identitas peserta, serta skor total kompetisi. Seluruh proses analisis menggunakan *Python* dengan memanfaatkan pustaka *pandas*, *scikit-learn*, dan *matplotlib*.

Tahap awal penelitian adalah pembersihan data (*data cleaning*) untuk memastikan dataset siap dianalisis. Proses ini meliputi pemeriksaan nilai yang hilang (*missing values*), kesesuaian antara kelas peserta dan tingkatan IKMC, serta validasi konsistensi data. Identifikasi *outlier* dilakukan menggunakan metode *Interquartile Range* (IQR) dan *Z-score* dengan ambang batas ± 3 standar deviasi. Namun, *outlier* ekstrem tidak dieliminasi karena dalam konteks kompetisi matematika, nilai ekstrem justru mencerminkan variasi kemampuan peserta dan merupakan bagian dari karakteristik distribusi skor kompetisi.

Penentuan jumlah *cluster* optimal (k) untuk setiap tingkatan dilakukan menggunakan dua metode. Metode pertama adalah *Elbow Method*, yang mengidentifikasi titik “siku” pada kurva *Within-Cluster Sum of Squares* (WCSS) terhadap jumlah *cluster*, yaitu kondisi ketika penambahan *cluster* tidak lagi menghasilkan penurunan WCSS yang signifikan (Aggarwal & Reddy, 2014). Metode kedua adalah *Silhouette Score*, yang mengukur tingkat kesesuaian setiap data point dengan *cluster* tempatnya berada dibandingkan dengan *cluster* terdekat lainnya. Nilai *Silhouette Score* dihitung menggunakan rumus:

$$s(i) = \frac{(b(i) - a(i))}{\max(a(i), b(i))}$$

dengan $a(i)$ menyatakan rata-rata jarak data point ke seluruh anggota *cluster* yang sama (cohesion), dan $b(i)$ menyatakan rata-rata jarak data point ke *cluster* terdekat (separation). Nilai *Silhouette Score* berada pada rentang -1 hingga 1 , di mana nilai yang lebih tinggi menunjukkan kualitas *clustering* yang lebih baik (Rousseeuw, 1987). Jumlah *cluster* optimal dipilih berdasarkan kombinasi nilai *Silhouette Score* tertinggi dan titik *elbow* yang paling jelas pada kurva WCSS. Meskipun grafik *elbow* dapat menunjukkan lebih dari satu kandidat nilai k , interpretasi akhir memprioritaskan solusi dengan tiga *cluster* untuk memudahkan kategorisasi kemampuan matematis ke dalam kelompok rendah, sedang, dan tinggi.

Proses pengelompokan dilakukan menggunakan algoritma *K-Means Clustering*, yaitu metode *unsupervised learning* yang mengelompokkan data ke dalam k cluster dengan meminimalkan nilai WCSS (Aggarwal & Reddy, 2014). Algoritma ini bekerja secara iteratif dengan memperbarui posisi *centroid* berdasarkan rata-rata skor anggota *cluster* hingga mencapai kondisi konvergen. Jarak antar data dihitung menggunakan jarak *Euclidean* yang dirumuskan sebagai berikut:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2}$$

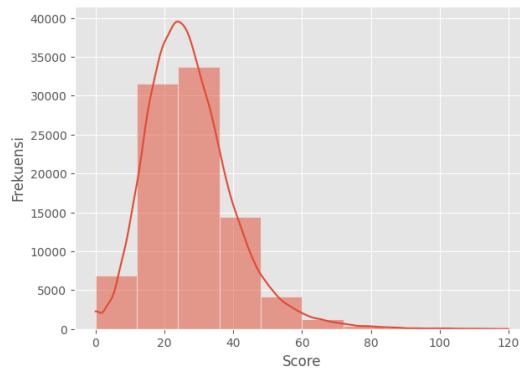
dengan x_i menyatakan data point ke- i , c_j adalah *centroid cluster* ke- j , dan n merupakan jumlah dimensi fitur yang digunakan. Setelah proses *clustering* selesai, setiap peserta memperoleh label *cluster* yang merepresentasikan tingkat kemampuan matematisnya. Analisis deskriptif dilakukan untuk setiap *cluster* dengan menghitung jumlah peserta, nilai rata-rata, nilai minimum dan maksimum, serta sebaran skor. *Cluster* dengan nilai *centroid* terendah diinterpretasikan sebagai kelompok kemampuan rendah, sedangkan *cluster* dengan nilai *centroid* tertinggi diinterpretasikan sebagai kelompok kemampuan tinggi.

Hasil analisis selanjutnya divisualisasikan dalam bentuk grafik *Elbow Method*, diagram sebar *cluster*, serta diagram batang distribusi kategori kemampuan, baik secara global maupun per tingkatan. Visualisasi ini bertujuan untuk memperjelas pemetaan kemampuan matematis antar *cluster* serta menjadi dasar dalam pembahasan hasil penelitian pada bagian selanjutnya.

RESULTS AND DISCUSSION

Temuan penelitian

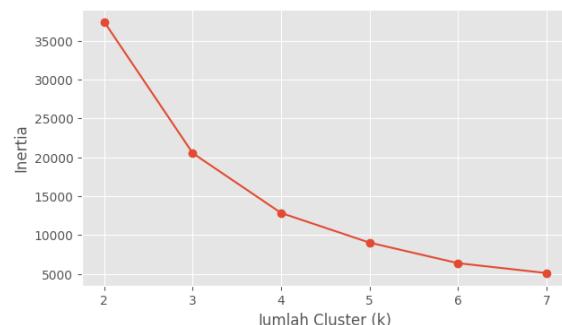
Sebagai tahap awal analisis, distribusi skor peserta IKMC 2024 dianalisis untuk memperoleh gambaran umum mengenai sebaran dan keragaman kemampuan matematis peserta. Visualisasi distribusi skor ini disajikan pada Gambar 1.



Gambar 1. Distribusi Skor Peserta IKMC 2024

Berdasarkan Gambar 1, skor peserta IKMC 2024 tersebar pada rentang nilai 0 hingga 120. Distribusi skor menunjukkan pola yang tidak simetris, dengan konsentrasi skor yang lebih besar pada rentang rendah hingga menengah. Sebagian kecil peserta memperoleh skor tinggi, yang terlihat dari ekor distribusi pada nilai skor tinggi. Data ini menunjukkan adanya variasi skor yang cukup besar antar peserta IKMC 2024. Distribusi skor yang tidak simetris tersebut menunjukkan indikasi bahwa data tidak mengikuti pola distribusi normal. Oleh karena itu, analisis lanjutan dilakukan menggunakan pendekatan *unsupervised learning* untuk mengelompokkan peserta berdasarkan kemiripan skor secara objektif.

Setelah memperoleh gambaran awal distribusi skor, analisis dilanjutkan dengan pengelompokan kemampuan matematis peserta IKMC 2024 secara global menggunakan algoritma *K-Means Clustering*. Penentuan jumlah cluster optimal dilakukan menggunakan kombinasi *Elbow Method* dan *Silhouette Score* untuk setiap tingkatan. Gambar 2 menunjukkan kurva *Elbow Method* yang memvisualisasikan penurunan nilai *Within-Cluster Sum of Squares* (WCSS) terhadap penambahan jumlah *cluster*.



Gambar 2. *Elbow Method* Penentuan Jumlah *Cluster*

Analisis *Elbow Method* menunjukkan titik "siku" yang relatif jelas pada $k=3$ untuk semua tingkatan, mengindikasikan bahwa penambahan *cluster* lebih dari tiga tidak memberikan penurunan WCSS yang signifikan. Tabel 1 menyajikan ringkasan hasil evaluasi untuk penentuan *cluster* optimal pada setiap tingkatan.

Tabel 1. Hasil Evaluasi Penentuan *Cluster* Optimal

Tingkatan	Jumlah Peserta	K Optimal	Silhouette Score	WCSS
Pre Ecolier	26.654	3	0,62	3.847.293
Ecolier	26.530	3	0,64	3.124.567
Benjamin	23.610	3	0,59	3.976.432
Cadet	10.045	3	0,58	2.621.876
Junior	4.240	3	0,61	1.187.654
Student	1.450	3	0,63	434.109

Nilai *Silhouette Score* berkisar antara 0,58-0,64 pada semua tingkatan, mengindikasikan kualitas *clustering* yang baik dengan pemisahan *cluster* yang jelas dan

kohesi tinggi dalam *cluster* (Rousseeuw, 1987). Konsistensi nilai $k=3$ pada semua tingkatan memfasilitasi komparabilitas hasil antar jenjang dan memudahkan interpretasi sebagai kelompok kemampuan rendah, sedang, dan tinggi.

Clustering global terhadap seluruh 92.529 peserta menghasilkan tiga *cluster* dengan karakteristik berbeda. Tabel 2 menyajikan statistik deskriptif clustering global.

Tabel 2. Statistik Deskriptif *Cluster Global Kemampuan Matematis*

Cluster	Jumlah Peserta	Persentase	Rata-rata Skor	Range	Kategori
C1	44.912	48,5%	17,5	0-25	Rendah
C2	38.450	41,6%	32,7	26-43	Sedang
C3	9.167	9,9%	53,8	44-120	Tinggi

Temuan menunjukkan bahwa hampir setengah peserta (48,5%) berada pada *cluster* kemampuan rendah dengan rata-rata skor 17,5, diikuti oleh *cluster* sedang (41,6%) dengan rata-rata 32,7, dan *cluster* tinggi hanya mencakup 9,9% peserta dengan rata-rata skor 53,8. Distribusi ini mengkonfirmasi adanya heterogenitas kemampuan yang signifikan dalam populasi peserta IKMC Indonesia. Rentang skor pada *cluster* tinggi yaitu 44-120 mengindikasikan adanya subkelompok peserta dengan kemampuan matematis sangat menonjol yang mampu menyelesaikan soal-soal tingkat kesulitan maksimal.

Untuk memperoleh gambaran yang lebih rinci, analisis *clustering* kemudian dilakukan secara terpisah pada setiap tingkatan kompetisi. Analisis per tingkatan ini bertujuan melihat variasi kemampuan matematis sesuai jenjang pendidikan dan mengidentifikasi pola perkembangan kemampuan sejak tahap awal pembelajaran. Pembahasan diawali dengan tingkatan *Pre Ecolier* yang merepresentasikan fase awal penguasaan konsep matematika dasar.

Tabel 3. Karakteristik *Cluster Pre Ecolier*

Cluster	Jumlah Peserta	Persentase	Rata-rata Skor	Range	Kategori
C1	9.456	35,5%	13,0	0-19	Rendah
C2	12.396	46,5%	27,0	20-36	Sedang
C3	4.802	18,0%	46,5	37-92	Tinggi

Pada tingkatan *Pre Ecolier*, mayoritas peserta (46,5%) berada pada *cluster* sedang, sementara 35,5% pada cluster rendah dan 18,0% pada cluster tinggi. Meski berada pada jenjang pendidikan paling awal, sudah terbentuk diferensiasi kemampuan yang jelas dengan gap 33,5 poin antara cluster rendah dan tinggi. Temuan ini sejalan dengan penelitian Leikin (2021) yang menunjukkan bahwa talenta matematis dapat teridentifikasi sejak pendidikan usia dini. Skor maksimum 92 pada *cluster* kemampuan tinggi mengindikasikan adanya siswa dengan kemampuan matematis tingkat lanjut pada tahap awal pendidikan.

Tabel 4. Karakteristik *Cluster Ecolier*

Cluster	Jumlah Peserta	Persentase	Rata-rata Skor	Range	Kategori
C1	13.382	50,4%	16,3	0-23	Rendah
C2	10.814	40,8%	29,8	24-39	Sedang
C3	2.334	8,8%	48,5	40-93	Tinggi

Tingkatan Ecolier menunjukkan adanya pergeseran pola kemampuan, ditandai dengan meningkatnya proporsi *cluster* kemampuan rendah menjadi 50,4%, sementara *cluster* kemampuan tinggi menurun cukup tajam menjadi 8,8%. Penurunan proporsi *cluster* tinggi dari 18,0% pada tingkat *Pre Ecolier* menjadi 8,8% mengindikasikan bahwa kompleksitas soal pada jenjang ini mulai menjadi tantangan kognitif yang signifikan bagi sebagian peserta. Selain itu, jarak skor antar *cluster* melebar hingga 32,2 poin, yang menunjukkan semakin jelasnya divergensi kemampuan matematis antar siswa. Pola ini sejalan dengan temuan penelitian mutakhir yang menyatakan bahwa perbedaan kemampuan matematis cenderung semakin melebar seiring peningkatan jenjang pendidikan (Geary et al., 2019).

Tabel 5. Karakteristik *Cluster Benjamin*

Cluster	Jumlah Peserta	Persentase	Rata-rata Skor	Range	Kategori
C1	11.014	46,6%	20,0	0-27	Rendah
C2	10.407	44,1%	34,5	28-44	Sedang
C3	2.189	9,3%	53,7	45-116	Tinggi

Pada tingkatan *Benjamin*, proporsi *cluster* tinggi sedikit meningkat menjadi 9,3% dengan skor maksimum mencapai 116, tertinggi di antara semua tingkatan. Gap skor antara *cluster* rendah dan tinggi melebar menjadi 33,7 poin. Peningkatan skor maksimum ini mengindikasikan bahwa pada akhir jenjang sekolah dasar, kelompok siswa dengan kemampuan matematis tinggi mampu menangani soal-soal yang sangat kompleks yang menuntut penalaran abstrak dan kreativitas pemecahan masalah.

Tabel 6. Karakteristik *Cluster Cadet*

Cluster	Jumlah Peserta	Persentase	Rata-rata Skor	Range	Kategori
C1	4.814	47,9%	21,2	0-29	Rendah
C2	4.214	42,0%	37,8	30-51	Sedang
C3	1.017	10,1%	65,4	52-120	Tinggi

Tingkatan *Cadet* menunjukkan peningkatan signifikan pada rata-rata skor *cluster* tinggi menjadi 65,4, dengan gap antar *cluster* melebar menjadi 44,2 poin terlebar di antara semua tingkatan. Proporsi *cluster* tinggi meningkat menjadi 10,1%, menandai titik kritis dimana kemampuan matematis tinggi mulai terbentuk lebih stabil.

Tabel 7. Karakteristik *Cluster Junior*

Cluster	Jumlah Peserta	Persentase	Rata-rata Skor	Range	Kategori

C1	1.701	40,1%	21,2	0-28	Rendah
C2	2.074	48,9%	36,5	29-48	Sedang
C3	465	11,0%	61,4	49-112	Tinggi

Struktur *cluster* pada tingkatan Junior menunjukkan polarisasi kemampuan yang jelas dengan *cluster* sedang mendominasi (48,9%). Proporsi *cluster* tinggi mencapai puncaknya di 11,0%, namun dengan jumlah absolut hanya 465 peserta. Gap skor antara *cluster* sedang dan tinggi mencapai 24,9 poin, mencerminkan perbedaan kemampuan yang semakin tajam pada jenjang pendidikan menengah atas.

Tabel 8. Karakteristik *Cluster Student*

Cluster	Jumlah Peserta	Persentase	Rata-rata Skor	Range	Kategori
C1	454	31,3%	15,4	0-21	Rendah
C2	698	48,1%	28,4	22-36	Sedang
C3	298	20,6%	45,6	37-120	Tinggi

Tingkatan *Student* menunjukkan pola distribusi yang berbeda dengan proporsi *cluster* tinggi meningkat menjadi 20,6% tertinggi di antara semua tingkatan. Namun, rata-rata skor *cluster* tinggi justru menurun menjadi 45,6, jauh lebih rendah dibanding *Cadet* (65,4) dan *Junior* (61,4).

Jika ditinjau secara keseluruhan, dominasi *cluster* kemampuan rendah hingga sedang pada peserta IKMC 2024 tidak dapat dilepaskan dari karakteristik soal IKMC yang menekankan penalaran logis, kreativitas, dan fleksibilitas strategi pemecahan masalah, bukan sekadar keterampilan prosedural (Akveld et al., 2020). Kondisi ini membuat mayoritas peserta mengalami kesulitan mencapai skor tinggi, terutama jika pengalaman belajar matematika sebelumnya masih berfokus pada latihan rutin dan penerapan algoritmik. Temuan ini konsisten dengan hasil PISA 2022 yang menunjukkan bahwa kemampuan *mathematical reasoning* dan *problem-solving* siswa Indonesia dalam konteks tidak familiar masih relatif rendah (OECD, 2023). Dengan demikian, rendahnya proporsi peserta pada *cluster* kemampuan tinggi mencerminkan tantangan pembelajaran matematika yang bersifat sistemik, bukan semata-mata akibat tingkat kesulitan kompetisi.

Analisis lintas tingkatan memperlihatkan adanya pola perkembangan kemampuan yang tidak linear. Transisi *Pre Ecolier* ke *Ecolier* (kelas 2-3) menunjukkan penurunan proporsi *cluster* tinggi dari 18,0% menjadi 8,8%, mengindikasikan bahwa peningkatan kompleksitas operasi aritmatika menjadi hambatan awal yang signifikan. Pada jenjang *Benjamin* ke *Cadet* (kelas 6-7), gap antar cluster melebar dari 33,7 menjadi 44,2 poin terlebar di antara semua tingkatan. Tingkatan *Junior* ke *Student* (kelas 10-11) memperlihatkan penurunan rata-rata *cluster* tinggi dari 61,4 menjadi 45,6, mencerminkan adanya lonjakan tuntutan kompleksitas dalam penguasaan matematika tingkat lanjut 20-35% peserta pada *cluster* kemampuan rendah di hampir semua

tingkatan menunjukkan bahwa kesulitan belajar matematika cenderung bersifat kumulatif dan semakin menguat apabila tidak disertai dukungan pembelajaran yang tepat sejak jenjang awal (Nelson & Powell, 2018).

Dibandingkan dengan penelitian berbasis ujian reguler, struktur *cluster* pada data IKMC tampak lebih kontras (Xu et al., 2019), mengindikasikan bahwa kompetisi matematika lebih sensitif dalam mengungkap perbedaan kemampuan, khususnya pada aspek pemecahan masalah non-rutin. Perbedaan ini juga menjelaskan mengapa proporsi peserta berkemampuan tinggi dalam penelitian ini (9,9% secara global) lebih kecil dibandingkan kompetisi olimpiade yang bersifat selektif, dimana proporsi mencapai 15-20% (Freiman & Leikin, 2021). Dengan demikian, IKMC berhasil menarik peserta dengan kemampuan lebih beragam, namun pada waktu yang bersamaan juga memperbesar heterogenitas kemampuan matematis.

Secara keseluruhan, pembahasan ini menunjukkan bahwa pemetaan kemampuan matematis berbasis *clustering* pada data kompetisi memberikan gambaran yang lebih tajam dan realistik mengenai heterogenitas kemampuan siswa. Temuan ini membuka ruang bagi pengembangan strategi pembelajaran terdiferensiasi, intervensi dini pada fase transisi kritis, serta pembinaan peserta berkemampuan tinggi yang lebih terarah dan berbasis bukti empiris.

Temuan penelitian ini secara langsung menjawab tiga tujuan yang ditetapkan. Pertama, pemetaan kemampuan matematis berhasil mengidentifikasi struktur tiga *cluster* pada setiap tingkatan dengan kualitas baik (*Silhouette Score* 0.58-0,64). Kedua, karakterisasi *cluster* mengungkap dominasi kemampuan rendah-sedang (80-90%) dengan *gap* skor yang melebar dari 33,5 poin (*Pre Ecolier*) menjadi 44,2 poin (*Cadet*). Ketiga, analisis pola perkembangan mengidentifikasi perubahan signifikan pada tiga fase transisi: penurunan proporsi *cluster* tinggi pada fase awal pendidikan dasar, pelebaran *gap* terbesar pada transisi ke sekolah menengah pertama, dan penurunan performa *cluster* tinggi pada transisi ke sekolah menengah atas.

CONCLUSION

Penelitian ini bertujuan memetakan kemampuan matematis peserta *International Kangaroo Mathematics Contest* (IKMC) 2024 di Indonesia secara objektif menggunakan pendekatan *K-Means Clustering*. Berdasarkan analisis terhadap 92.529 peserta yang mencakup enam tingkatan kompetisi, penelitian ini menunjukkan bahwa kemampuan matematis peserta bersifat heterogen dan membentuk pengelompokan alami yang konsisten. Pengelompokan tersebut menghasilkan tiga *cluster* utama yang merepresentasikan kemampuan matematis rendah, sedang, dan tinggi, baik secara global maupun pada setiap tingkatan. Temuan ini menjawab tujuan penelitian dengan menegaskan bahwa pendekatan *clustering* mampu menggambarkan struktur kemampuan matematis peserta secara lebih representatif dibandingkan kategorisasi

berbasis batas nilai konvensional. Selain itu, pemetaan ini memperlihatkan bahwa perbedaan kemampuan matematis telah muncul sejak jenjang awal dan berkembang seiring peningkatan tingkat pendidikan.

Implikasi dari temuan ini menunjukkan bahwa pemetaan kemampuan matematis berbasis data kompetisi dapat dimanfaatkan sebagai dasar perancangan pembelajaran terdiferensiasi, identifikasi dini peserta yang memerlukan dukungan tambahan, serta pembinaan peserta berkemampuan tinggi secara lebih terarah. Dengan demikian, penelitian ini berkontribusi dalam pengembangan analisis data pendidikan matematika dan menyediakan landasan bagi pengambilan keputusan yang lebih tepat sasaran.

BIBLIOGRAPHY

- Aggarwal, C. C., & Reddy, C. K. (2014). *Data clustering: Algorithms and applications*. Boca Raton, FL: CRC Press.
- Akveld, M., Cáceres, L., Nieto, J. H., & Sánchez, R. (2020). The Math Kangaroo competition. *Espacio Matemático*, 1(2), 74-91.
- Association Kangourou Sans Frontières. (2024). *International Kangaroo Mathematics Contest: Regulations and philosophy*.
- Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61-75). Springer.
- Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, 5, 15991-16005.
- Freiman, V., & Leikin, R. (2021). Problem-solving strategies and mathematical giftedness: Insights from competition data. *ZDM Mathematics Education*, 53(7), 1535-1548.
- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2019). Mathematical cognition deficits in children with learning difficulties. *Journal of Learning Disabilities*, 52(2), 110-123.
- Han, S., & Kim, J. (2023). Comparative analysis of clustering algorithms for student performance data in mathematical competitions. *Education and Information Technologies*, 28(3), 3421-3445.
- Leikin, R. (2021). When practice needs more research: The nature and nurture of mathematical giftedness. *ZDM Mathematics Education*, 53(7), 1579-1589.
- Liao, X., Jiang, S., & Liu, Y. (2021). Educational data mining: A bibliometric analysis of an emerging field. *Interactive Learning Environments*, 31(6), 3520-3536.
- Nelson, G., & Powell, S. R. (2018). A systematic review of instructional interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, 51(2), 131-147.
- Nuraeni, R., Suryadi, D., & Turmudi. (2023). Clustering analysis of students' mathematical problem-solving abilities. *Journal of Mathematics Education*, 14(2), 245-258.
- Organisation for Economic Co-operation and Development. (2023). PISA 2022 results

- (Volume I): The state of learning and equity in education. OECD Publishing.
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.
- Saputra, R. R. H., & Rahmah, H. (2025). Performa siswa sekolah Islam dalam International Kangaroo Mathematics Contest. *JPG: Jurnal Pendidikan Guru*, 6(4), 770–778.
- Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. *Journal of Education*, 196(2), 1-38.
- Setiawan, W., & Kusumah, Y. S. (2021). Pemetaan kemampuan matematika siswa berdasarkan hasil ujian nasional menggunakan K-Means clustering. *Jurnal Pendidikan Matematika*, 15(1), 67–78.
- Tungson, T. T. T., & Saputra, R. R. H. (2024). Efektivitas metode pembayaran “seikhlasnya” dan uji soal matematika nalaria realistik dalam meningkatkan partisipasi dan kemampuan matematika siswa sekolah-sekolah Islam di Surabaya. *JPG: Jurnal Pendidikan Guru*, 5(2), 278–284.
- Xu, D., Wang, H., & Wang, M. (2019). A conceptual model for K-Means clustering of student performance data in mathematics education. *Educational Technology & Society*, 22(4), 122-135.