

# Optimalisasi Analisis Data Peserta Olimpiade Sains Nasional Indonesia Menggunakan Algoritma K-Means Clustering

Agustina Purwatiningsih<sup>1</sup>, Muhammad Habibi<sup>2</sup>

<sup>1</sup>Department of Software Engineering, Universitas Duta Bangsa Surakarta, Surakarta, Indonesia

<sup>2</sup>Department of Informatics, Universitas Jendral Achmad Yani Yogyakarta, Yogyakarta, Indonesia

<sup>1</sup>agustinapurwa@udb.ac.id\*, <sup>2</sup>muhammadhabibi17@gmail.com

## Abstract

*Mastery of science and technology with high integrity values is one of the main requirements for the progress of a nation. One of the main programs for developing students' talents and interests organized by the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) through the National Achievement Center (Puspresnas) is the National Science Olympiad (OSN). The purpose of OSN is to get prospective participants to represent Indonesia in international science competitions and build a national database of talented students in the field of science. The principles of OSN are inclusive, growth, participative and sustainable, namely equal opportunities for all Indonesian students without distinguishing between ethnicity, religion, appearance, and race. As well as intensification of coaching in the regions in order to strive for equal achievement through talent scouting activities. The K-Means clustering algorithm is suitable for finding patterns in educational data such as student performance, coaching effectiveness, or regions with certain levels of achievement. The purpose of this study is to find patterns in OSN results based on the medal acquisition of each region in Indonesia, thus producing strategic information for equalizing coaching. The results of this study, there are 3 clusters for provinces for OSN medal acquisition and 6 clusters of cities and districts according to medal results. Cluster 1 provincial data and cluster 1 district or city data are priority areas for OSN coaching from the National Achievement Center (Puspresnas). There are 33 provinces and 168 districts or cities included in cluster 1, which require special attention. The results of this study can be a reference for Puspresnas in enforcing the OSN principle, namely equal opportunities for all Indonesian students without distinguishing between ethnicity, religion, appearance, and race.*

*Keywords: K-means, National Science Olympiad, data mining, clustering*

## Abstrak

Penguasaan ilmu pengetahuan dan teknologi dengan nilai integritas tinggi merupakan salah satu syarat utama kemajuan sebuah bangsa. Salah satu program utama untuk pengembangan bakat dan minat peserta didik yang diselenggarakan oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek) melalui Pusat Prestasi Nasional (Puspresnas) adalah Olimpiade Sains Nasional (OSN). Tujuan dari OSN mendapatkan calon peserta untuk mewakili Indonesia pada kompetisi sains tingkat internasional dan membangun basis data nasional peserta didik yang bertalenta dalam bidang sains. Prinsip OSN adalah *inclusive, growth, participative* dan *sustain*, yaitu pemerataan kesempatan bagi seluruh peserta didik Indonesia tanpa membedakan suku, agama, rupa, dan ras. Serta intensifikasi pembinaan di daerah dalam rangka mengupayakan pemerataan prestasi melalui kegiatan pencarian dan pemanduan bakat (*talent scouting*). Algoritma K-Means clustering cocok untuk menemukan pola dalam data pendidikan seperti performa siswa, efektivitas pembinaan, atau wilayah dengan tingkat pencapaian tertentu. Tujuan dari penelitian ini adalah untuk menemukan pola hasil OSN berdasarkan perolehan medali tiap wilayah di Indonesia, sehingga menghasilkan informasi strategis untuk pemerataan pembinaan. Hasil dari penelitian ini, terdapat 3 kluster untuk provinsi perolehan medali OSN dan 6 kluster kota dan kabupaten menurut perolehan hasil medali. Kluster 1 data provinsi dan kluster 1 data kabupaten atau kota merupakan daerah prioritas untuk pembinaan OSN dari Pusat Prestasi Nasional (Puspresnas). Terdapat 33 provinsi dan 167 kabupaten atau kota yang termasuk dalam kluster 1, yang memerlukan perhatian khusus. Hasil dari penelitian ini dapat menjadi acuan untuk Puspresnas dalam menegakkan prinsip OSN yaitu pemerataan kesempatan bagi seluruh peserta didik Indonesia tanpa membedakan suku, agama, rupa, dan ras.

Kata kunci: K-Means, Olimpiade Sains Nasional, data mining, kluster

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

## 1. Pendahuluan

Penguasaan ilmu pengetahuan dan teknologi dengan nilai integritas tinggi merupakan salah satu syarat utama kemajuan sebuah bangsa. Salah satu program utama untuk pengembangan bakat dan minat peserta didik yang diselenggarakan oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek) melalui Pusat Prestasi Nasional

(Puspresnas) adalah Olimpiade Sains Nasional (OSN). Diharapkan OSN dapat mengantarkan peserta didik mendapatkan ilmu pengetahuan dan teknologi serta pemerataan prestasi dan memaksimalkan potensi dari seluruh pelosok NKRI. OSN diselenggarakan secara bertingkat, mulai dari tingkat kabupaten/kota hingga nasional, dengan tujuan meningkatkan kualitas pendidikan di Indonesia dengan mendorong siswa untuk menguasai ilmu pengetahuan dan teknologi,

membangun budaya kompetisi yang sehat, dan mewujudkan talenta emas bangsa [1]. Skill ini sekarang sangat penting untuk mengidentifikasi potensi siswa dan mendukung pengembangan SDM sains yang unggul [2]. Bidang pendidikan sangat dipengaruhi oleh penyelenggaraan OSN. Siswa tidak hanya mendapatkan peningkatan dalam kemampuan akademik mereka melalui OSN, tetapi mereka juga memperoleh pemahaman tentang soft skills seperti kreatifitas, olahraga, dan kemampuan berpikir kritis.

Mekanisme bertingkat tersebut merupakan salah satu cara untuk memberikan kesempatan yang sama dan adil bagi peserta didik di seluruh Indonesia untuk berprestasi dan menjadi bibit-bibit talenta potensial [3]. Tujuan dari OSN mendapatkan calon peserta untuk mewakili Indonesia pada kompetisi sains tingkat internasional dan membangun basis data nasional peserta didik yang bertalenta dalam bidang sains. Penyelenggaraan ajang OSN mengikuti *prinsip inclusive, growth, participative* dan *sustain*. Pemerataan kesempatan bagi seluruh peserta didik Indonesia tanpa membedakan suku, agama, rupa, dan ras. Serta intensifikasi pembinaan di daerah dalam rangka mengupayakan pemerataan prestasi melalui kegiatan pencarian dan pemanduan bakat (talent scouting) yang melibatkan seluruh pemangku kepentingan merupakan upaya untuk mencapai prinsip OSN [4].

Namun, pembinaan dan seleksi peserta OSN seringkali menghadapi masalah seperti data yang tidak terstruktur dan kurangnya pendekatan pengambilan keputusan strategis berbasis data. Ketika teknologi big data berkembang, ada peluang untuk mengatasi masalah ini. Dengan keunggulan dalam mengelompokkan data dengan cepat dan efisien, algoritma K-Means *clustering* sangat cocok untuk menemukan pola dalam data pendidikan seperti performa siswa, efektivitas pembinaan, atau wilayah dengan tingkat pencapaian tertentu [5-7].

Beberapa penelitian telah menunjukkan bahwa algoritma ini efektif dalam konteks pendidikan. Misalnya, [5] menunjukkan bagaimana penggunaan K-Means dalam pembuatan model analisis data pendidikan untuk menilai kualitas siswa. Sehingga dapat menemukan pola pembelajaran yang dapat membantu strategi pembinaan yang lebih baik. Penelitian lain oleh [6] juga penggunaan K-Means untuk mengelompokkan peserta OSN tingkat SMA berdasarkan kemampuan mereka. Penelitian ini memberikan hasil yang signifikan dalam menentukan

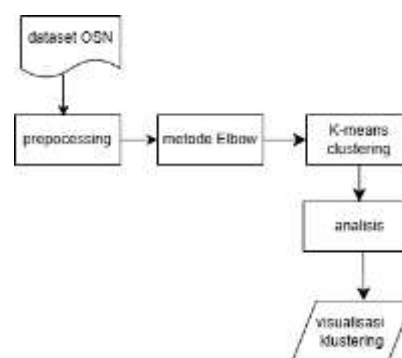
strategi pembinaan peserta. Namun [7], menggunakan multi-clustering untuk melihat kabupaten dan kota di Provinsi Jawa Barat berdasarkan Indeks Desa Membangun, yang memberikan rekomendasi strategis untuk meningkatkan pendidikan di daerah tersebut.

Kemampuan untuk menangani data berukuran besar dengan cepat adalah keunggulan utama K-Means. Namun, kelemahan algoritma ini termasuk sensitivitas terhadap inisialisasi centroid, yang dapat berdampak pada hasil akhir klasterisasi. Metode yang berbeda untuk meningkatkan kekuatan algoritma K-Means dalam analisis data pendidikan dibahas dalam penelitian [8] dan menunjukkan relevansinya dalam berbagai konteks analisis klaster. Banyak penelitian telah menggunakan algoritma K-Means untuk menganalisis data pendidikan,

Namun, masih ada kekurangan dalam bagaimana metode ini dapat diterapkan untuk berbagai elemen dalam konteks OSN, seperti mengelola data siswa, sekolah, dan wilayah secara menyeluruh. Tujuan dari penelitian ini adalah untuk mengisi celah tersebut dan menghasilkan informasi strategis untuk pembinaan dan pengambilan keputusan yang lebih baik dengan menggunakan K-Means dalam analisis data OSN.

## 2. Metode Penelitian

Pada penelitian ini metode klasterisasi data Olimpiade Sains Nasional, dibagi menjadi beberapa tahapan yaitu pengumpulan data, preprocessing, penentuan jumlah klaster, *clustering*, analisis, serta visualisasi klaster. Langkah penelitian dapat terlihat pada Gambar 1 berikut.



Gambar 1. Langkah Penelitian

### 2.1. Dataset

Data yang digunakan pada penelitian ini berupa dataset yang bersifat public dari [9], yaitu data OSN dari tahun 2009 sampai 2024 yang diambil dari platform kaggle. Terdapat 38 provinsi di Indonesia yang mengikuti OSN dan ditambah 1 daerah yaitu luar negeri, dimana

setiap provinsi juga diwakili dari siswa beberapa kota dan kabupaten. Terdapat sembilan mata pelajaran yang diikuti dalam OSN yaitu : matematika, IPA, IPS, informatika, geografi, kebumian, fisika, kimia, dan biologi. Terdapat beberapa penghargaan pemenang Olimpiade Sains Nasional tiap mata pelajaran, diantaranya[4]:

- a. Peringkat 1 sampai 5 mendapat medali emas, sertifikat, dan uang pembinaan 25 juta.
- b. Peringkat 6 sampai 15 mendapat medali perak, sertifikat, dan uang pembinaan 20 juta.
- c. Peringkat 16 sampai 30 mendapat medali perunggu, sertifikat, dan uang pembinaan 15 juta.
- d. Peringkat diatas 30 mendapatkan medali partisipan.

Terdapat 8 atribut data OSN SMP dan SMA dari tahun 2009 sampai 2024, dapat dilihat pada Tabel 1.

Tabel 1. Tabel Atribut OSN

NO	Atribut	Type	Keterangan
1	Nama Peserta	object	Nama peserta OSN
2	Sekolah	object	Sekolah asal peserta
3	Provinsi	object	Provinsi sekolah peserta
4	Kab/Kota	object	Kota asal sekolah peserta
5	Bidang	object	Bidang lomba yang diikuti
6	Jenjang Lomba	object	Jenjang sekolah
7	Medali	object	Jenis medali yang diperoleh
8	Tahun	int64	Tahun lomba OSN

## 2.2. Preprocessing

Preprocessing adalah tahapan awal pada analisis data. Adapun fungsi preprocessing diantaranya, untuk menyederhanakan data yang kompleks, mengidentifikasi dan mengatasi outlier, serta menormalisasi data. Langkah ini merupakan hal penting yang harus dilakukan pada analisis data dengan K-Means. Data perolehan medali OSN tiap tahun diberi data yang unik yang dapat diproses oleh algoritma.

Selain itu, adanya kemungkinan nilai yang hilang dalam kumpulan data dapat mendistorsi atau mengurangi hasil analisis. Oleh karena itu, diperlukan preprocessing untuk nilai-nilai yang hilang sebelum melanjutkan analisis. Salah satu strategi yang umum adalah menghilangkan catatan dengan nilai yang hilang, terutama jika data yang hilang hanya mencakup sebagian kecil dari kumpulan data dan penghapusannya

tidak memiliki pengaruh besar terhadap integritas kumpulan data atau temuan analisis [10].

## 2.3. Metode Elbow

Pada analisis data menggunakan metode K-Means, akan membagi data berkelompok berdasarkan kedekatan titik pusat. *Clustering* merupakan proses pengelompokan sekumpulan objek sehingga objek-objek yang berada dalam kelompok atau cluster yang sama lebih mirip satu sama lain dibandingkan dengan objek-objek dalam kelompok lain. Salah satu tugas terpenting dalam analisis clustering adalah menemukan jumlah cluster terbaik untuk menangkap struktur intrinsik data[11].

Analisis kluster sering kali menghadapi tantangan dalam menentukan jumlah kluster yang optimal. Metode Elbow, yang memanfaatkan *Within Cluster Sum of Squares* (WCSS), merupakan teknik yang banyak digunakan untuk menentukan jumlah cluster yang optimal. Metode Elbow sangat efektif untuk menangani pengelompokan K-Means. Metode ini membuat grafik varians yang dijelaskan terhadap jumlah cluster dan memilih siku kurva sebagai jumlah cluster yang akan digunakan[12]. Persamaan 1 menunjukkan rumus untuk WCSS, di mana  $Y_i$  mewakili pusat massa yang terkait dengan pengamatan  $X_i$ , dan  $n$  menunjukkan jumlah fitur yang digunakan.

$$WCSS = \sum_{i=1}^n (X_i - Y_i)^2 \tag{1}$$

## 2.4. K-Means Clustering

Setelah jumlah kluster telah didapat dengan metode Elbow, maka tahapan selanjutnya adalah melakukan proses *clustering* dengan menggunakan metode K-Means. Metode pengelompokan K-Means clustering merupakan pendekatan non-hierarkis, dimana pengelompokan data ke dalam satu atau lebih kluster atau kelompok [13]. K-means *clustering* adalah strategi partisi yang membagi data menjadi  $k$  subset (atau cluster) yang tidak tumpang tindih yang tidak memiliki struktur internal [14]. Pengelompokan K-means melibatkan langkah-langkah sebagai berikut [15]:

- Secara acak menghasilkan  $k$  centroid dari titik data
- Bentuk  $k$  kluster dengan menugaskan setiap titik data ke centroid terdekat menggunakan jarak euclidean
- Menghitung ulang titik pusat setiap kluster.
- Ulangi langkah 2 dan 3 hingga pusat massa stabil (tidak ada pergerakan besar di lokasi atau jumlah putaran yang telah ditentukan).

- Rumus utama K-means adalah perhitungan jarak euclidean, yang digunakan untuk menetapkan titik data ke kluster.

Rumus utama K-means adalah penghitungan jarak Euclidean, yang digunakan untuk menetapkan titik data ke cluster.

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j + c_j)^2} \tag{2}$$

d = jarak (*distance*)

n = jumlah benda

j = (mulai dari 1 sampai n)

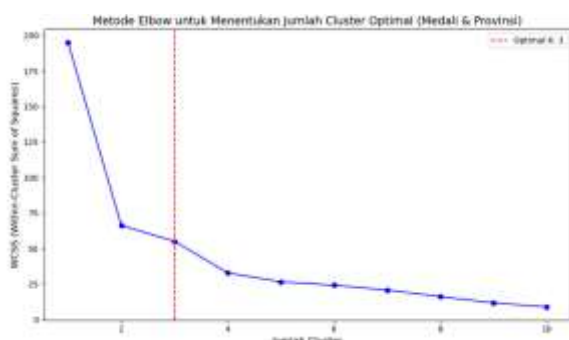
xj = fitur objek j terhadap x

cj = pusat massa fitur j

### 3. Hasil dan Pembahasan

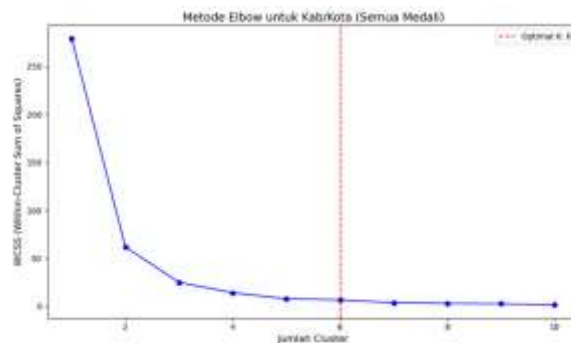
#### 3.1. Validasi Kluster

Hasil percobaan dari metode Elbow menunjukkan bahwa tiga cluster merupakan jumlah yang paling optimal untuk dataset OSN. Saat menggunakan Metode Elbow, eksperimennya melibatkan pemeriksaan berbagai kemungkinan cluster mulai dari 1 hingga 10. Metode ini menghitung jumlah kuadrat dalam cluster (WCSS) untuk setiap jumlah cluster yang mungkin dan mencari titik di mana pengurangan WCSS mulai melambat, menunjukkan penurunan keuntungan dari kekuatan penjelas dari lebih banyak cluster. Dalam situasi ini, plot WCSS menunjukkan 'siku' di k=3. Dimana data yang dipakai adalah perolehan medali tiap Provinsi. Gambar 2 menunjukkan hasil metode Elbow untuk menentukan jumlah kluster provinsi.



Gambar 2. Jumlah Kluster Provinsi

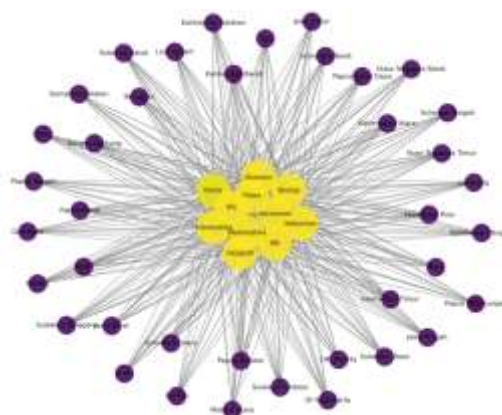
Sedangkan untuk menentukan jumlah kluster kabupaten atau kota didapatkan hasil elbow k=6, dimana terdapat enam kluster kabupaten atau kota dalam perolehan jumlah medali OSN. Gambar 3 menunjukkan hasil metode elbow untuk kluster kabupaten atau kota.



Gambar 3. Jumlah Kluster Kota atau Kabupaten

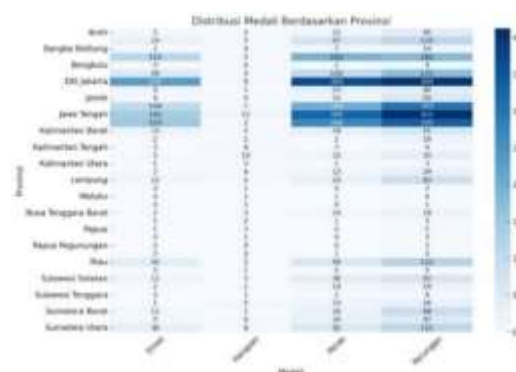
#### 3.2. Analisis Clustering

Terdapat keterkaitan antara provinsi peserta OSN dengan bidang lomba yang diikuti. Dimana peserta OSN adalah siswa SMP, SMA/MA/ sederajat dari seluruh Indonesia beserta satu Sekolah Indonesia Luar Negeri. Terdapat sembilan bidang lomba diantaranya : matematika, IPA, IPS, informatika, geografi, kebumihan, fisika, kimia, dan biologi. Keterkaitan antara provinsi dan bidang lomba yang diikuti terdapat pada Gambar 4.



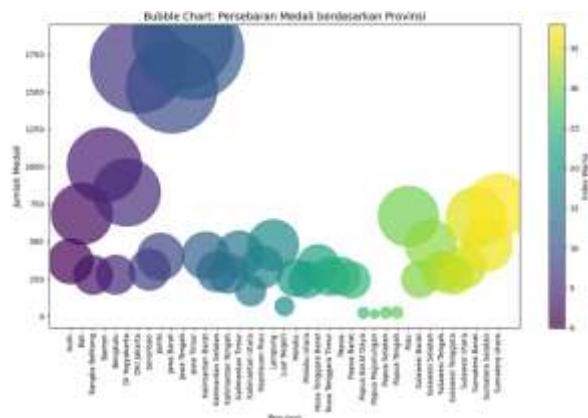
Gambar 4. Grafik Keterkaitan Provinsi dengan Bidang Lomba OSN

Dari data perolehan medali tiap tahun juga dapat dikelompokkan per provinsi. Ada beberapa jenis medali lomba untuk perolehan medali emas, perak, perunggu, dan harapan. Distribusi perolehan medali pada tahun 2024 dapat dilihat pada Gambar 5.



Gambar 5. Distribusi Medali OSN tahun 2024 berdasarkan Provinsi

Sedangkan pada Gambar 6 menunjukkan *buble chart* persebaran medali berdasarkan provinsi.



Gambar 6. *Buble Chart* Persebaran Medali berdasar Provinsi

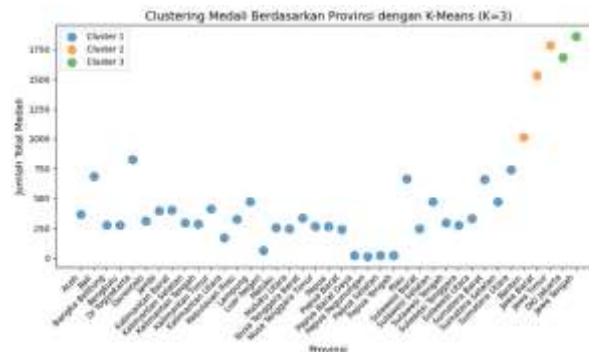
Dari data OSN ini didapatkan 10 top provinsi dengan perolehan medali terbanyak. Top 10 Provinsi peraih medali terbanyak dapat dilihat pada Tabel 2.

Tabel 2. Daftar 10 Provinsi dengan Perolehan Medali Terbanyak

No	Provinsi	Jumlah Medali
1	Jawa Tengah	1859
2	Jawa Timur	1781
3	DKI Jakarta	1680
4	Jawa Barat	1531
5	Banten	1012
6	DI Yogyakarta	828
7	Sumatera Utara	741
8	Bali	685
9	Riau	668
10	Sumatera Barat	663

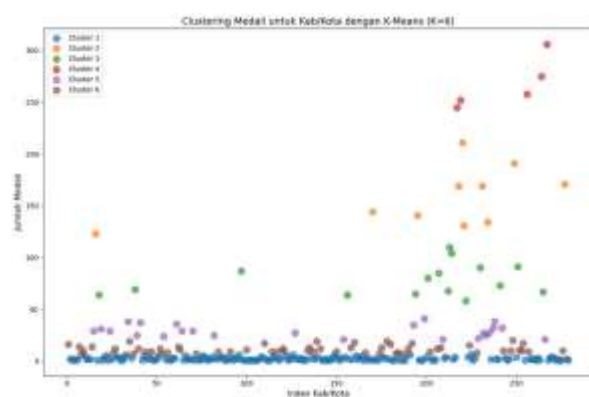
Hasil K-Means dengan  $k = 3$  berupa scatter plot antara perolehan medali dan provinsi asal peserta OSN dapat dilihat pada Gambar 7. Dari Gambar 7 dapat dijelaskan bahwa hasil perolehan medali OSN dapat dibagi menjadi 3 kluster yaitu kluster 1, 2, dan 3. Dimana kluster 3 merupakan provinsi dengan perolehan medali OSN terbanyak yaitu provinsi DKI Jakarta dan Jawa Tengah. Sedangkan kluster 2 merupakan provinsi dengan mendapatkan medali menengah yaitu Banten, Jawa Barat, dan Jawa Timur. Sisanya, 33 provinsi lainnya masuk dalam kluster 1, dimana kluster 1 merupakan segmen yang terdiri dari provinsi dengan perolehan medali paling sedikit. Dengan demikian, apabila Pusat Prestasi Nasional (Puspresnas) berminat untuk menambah pembinaan OSN untuk daerah tertinggal maka dapat memprioritaskan kluster 1 sebagai sasaran utama.

Oeh karena itu, apabila strategi pemerataan pendidikan dan pembinaan OSN lebih fokus pada kuantitas, maka calon peserta OSN diprospek dari segmen kluster 1 dan 2 akan lebih efektif. Hasil *clustering* berdasarkan titik pusat yang berdekatan atau *centroid* dapat dilihat pada Gambar 7.



Gambar 7. Hasil clustering medali OSN provinsi di Indonesia

Segmentasi hasil OSN lainnya, berdasarkan perolehan medali per kota atau kabupaten digambarkan pada Gambar 8. Hasil K-means dengan  $k=6$ , yaitu diperoleh enam *clustering* dari kluster 1 sampai 6. Titik centroid yang berdekatan dikempokkan menjadi satu berdasarkan kedekatan perolehan medali OSN. Jumlah kabupaten dan kota di setiap kluster adalah 168, 10, 15, 5, 24, dan 57, masing-masing untuk kluster 1, kluster 2, kluster 3, kluster 4, kluster 5, dan dan kluster 6. Gambar 8 menunjukkan segmentasi perolehan medali per kabupaten atau kota.



Gambar 8. Clustering Medali per Kota atau Kabupaten

Dari Gambar 8 dapat dijelaskan, bahwa kluster 4 merupakan kabupaten dan kota yang mendapatkan perolehan medali terbanyak, yaitu Kota Jakarta Barat, Kota Jakarta Selatan, Kota Semarang, Kota Surabaya, dan Kota Tangerang Selatan. Kluster 4 merupakan kota atau kabupaten unggulan untuk perolehan medali OSN yang bisa dijadikan sebagai wilayah percontohan pembinaan. Untuk urutan hasil medali selanjutnya

dikelompokkan menjadi klaster 2, terdiri dari 10 kota atau kabupaten yang dapat dilihat pada Tabel 3 berikut.

Tabel 3. Hasil Medali per Kota atau Kabupaten Klaster 2

No	Kab/Kota	Jumlah Medali	Klaster
1	Kab. Bantul	123	2
2	Kab. Tangerang	144	2
3	Kota Bandung	141	2
4	Kota Jakarta Pusat	169	2
5	Kota Jakarta timur	211	2
6	Kota Jakarta Utara	131	2
7	Kota Malang	169	2
8	Kota Medan	134	2
9	Kota Pekanbaru	191	2
10	Kota Yogyakarta	171	2

Terdapat 15 kota atau kabupaten yang termasuk dalam klaster 3, wilayah tersebut dapat dilihat pada Tabel 4 berikut. Klaster 3 merupakan kelompok kabupaten dan kota yang mendapatkan medali terbanyak setelah cluster 4 dan 2.

Tabel 4. Hasil Perolehan Medali per Kota atau Kabupaten Klaster 3

No	Kab/Kota	Jumlah Medali	Klaster
1	Kab. Banyumas	64	3
2	Kab. Bone Bolango	69	3
3	Kab. Magelang	87	3
4	Kab. Sragen	64	3
5	Kota Bandar Lampung	65	3
6	Kota Bekasi	80	3
7	Kota Bogor	85	3
8	Kota Cirebon	68	3
9	Kota Denpasar	110	3
10	Kota Depok	104	3
11	Kota Jambi	58	3
12	Kota Makassar	90	3
13	Kota Palembang	73	3
14	Kota Pontianak	91	3
15	Kota Surakarta	67	3

Klaster 5 merupakan kelompok kabupaten dengan perolehan medali diurutkan ke-4. Terdapat 24 kabupaten atau kota yang termasuk dalam klaster 5 yang dapat dilihat pada Tabel 5.

Tabel 5. Hasil Medali per Kota atau Kabupaten Klaster 5

No	Kab/Kota	Jumlah Medali	Klaster
1	Kab. Banjarnegara	29	5
2	Kab. Banyuwangi	31	5
3	Kab. Bekasi	29	5
4	Kab. Bogor	38	5
5	Kab. Boyolali	25	5
6	Kab. Buleleng	37	5
7	Kab. Gresik	24	5
8	Kab. Jember	36	5
9	Kab. Jombang	29	5
10	Kab. Kebumen	29	5
11	Kab. Kudus	25	5
12	Kab. Pati	27	5
13	Kab. Sleman	21	5

14	Kota Banda Aceh	35	5
15	Kota Batam	41	5
16	Kota Bukittinggi	21	5
17	Kota Magelang	22	5
18	Kota Manado	27	5
19	Kota Mataram	25	5
20	Kota Metro	27	5
21	Kota Padang	32	5
22	Kota Padang Panjang	38	5
23	Kota Palu	32	5
24	Kota Tangerang	21	5

Tabel 6 memperlihatkan daftar kabupaten atau kota yang termasuk dalam klaster 6. Terdapat 35 kabupaten atau kota yang termasuk klaster 6, klaster 6 merupakan kelompok daerah yang memperoleh medali terbanyak urutan ke-5.

Tabel 6. Hasil Medali per Kota atau Kabupaten Klaster 6

No	Kab/Kota	Jumlah Medali	Klaster
1	Kab. Aceh Besar	16	6
2	Kab. Badung	14	6
3	Kab. Bandung	7	6
4	Kab. Bandung Barat	10	6
5	Kab. Banjar	14	6
6	Kab. Belitung Timur	8	6
7	Kab. Bengkulu	12	6
8	Kab. Bojonegoro	19	6
9	Kab. Bone	7	6
10	Kab. Brebes	12	6
11	Kab. Cilacap	9	6
12	Kab. Deli Serdang	9	6
13	Kab. Gianyar	12	6
14	Kab. Gowa	7	6
15	Kab. Gunung Kidul	8	6
16	Kab. Jembrana	14	6
17	Kab. Jepara	11	6
18	Kab. Kendal	7	6
19	Kab. Kuningan	9	6
20	Kab. Lampung Timur	9	6
21	Kab. Malang	11	6
22	Kab. Muaro Jambi	17	6
23	Kab. Nganjuk	9	6
24	Kab. Ogan Komering Ilir	9	6
25	Kab. Padang Pariaman	12	6
26	Kab. Pamekasan	10	6
27	Kab. Ponorogo	10	6
28	Kab. Pringsewu	13	6
29	Kab. Purbalingga	9	6
30	Kab. Purworejo	19	6
31	Kab. Rembang	12	6
32	Kab. Sidoarjo	9	6
33	Kab. Sukabumi	13	6
34	Kab. Sukoharjo	18	6
35	Kab. Tabanan	8	6
36	Kab. Tapanuli Selatan	7	6
37	Kab. Tapanuli Tengah	13	6
38	Kab. Temanggung	19	6
39	Kab. Toba	16	6
40	Kab. Trenggalek	7	6

41	Kab. Tuban	8	6
42	Kab. Tulungagung	7	6
43	Kab. Wonosobo	12	6
44	Kota Balikpapan	17	6
45	Kota Bengkulu	9	6
46	Kota Blitar	12	6
47	Kota Bontang	13	6
48	Kota Kediri	15	6
49	Kota Madiun	8	6
50	Kota Pangkalpinang	10	6
51	Kota Payakumbuh	10	6
52	Kota Pekalongan	20	6
53	Kota Probolinggo	11	6
54	Kota Salatiga	10	6
55	Kota Samarinda	17	6
56	Kota Serang	9	6
57	Kota Tomohon	10	6

Segmentasi daerah dengan urutan perolehan medali paling sedikit masuk pada klaster 1. Terdapat 167 kabupaten dan kota yang masuk dalam klaster 1. Dimana klaster 1 merupakan kabupaten atau kota yang memperoleh medali paling sedikit diantara daerah lainnya yang dapat dilihat pada Tabel 7.

Tabel 7. Hasil Perolehan Medali per Kota atau Kabupaten Klaster 1

No	Kab/Kota	Jumlah Medali	Klaster
1	Kab. Aceh Selatan	2	1
2	Kab. Aceh Singkil	2	1
3	Kab. Aceh Timur	1	1
4	Kab. Agam	1	1
5	Kab. Asahan	2	1
..	.....	...	..
163	Kota Tegal	2	1
164	Kota Ternate	1	1
165	Kota Tidore Kepulauan	1	1
166	Malaysia	2	1
167	Mesir	1	1

Klaster 1 merupakan daerah yang perlu mendapatkan perhatian khusus tentang pembinaan OSN dari Pusat Prestasi Nasional (Puspresnas) untuk pemerataan kesempatan bagi seluruh peserta didik Indonesia untuk berprestasi.

#### 4. Kesimpulan

Dalam penelitian ini, peneliti berhasil mengoptimalkan analisis data OSN. Dari dataset OSN, didapat 3 kelompok provinsi yang dilambangkan dengan klaster 1, 2, dan 3. Dimana klaster 3 merupakan provinsi dengan perolehan medali OSN terbanyak yaitu provinsi DKI Jakarta dan Jawa Tengah. Sedangkan klaster 2 merupakan provinsi dengan mendapatkan medali menengah yaitu Banten, Jawa Barat, dan Jawa Timur. Sisanya, 33 provinsi lainnya masuk dalam klaster 1. Klaster 1 merupakan segmen yang terdiri dari provinsi dengan perolehan medali paling sedikit. Dengan demikian, apabila Pusat Prestasi Nasional

(Puspresnas) berminat untuk menambah pembinaan OSN untuk daerah tertinggal maka dapat memprioritaskan klaster 1 sebagai sasaran utama. Sedangkan pengelompokan hasil OSN per kabupaten atau kota dibagi menjadi 6 kelompok atau klaster. Dari *clustering* per kota atau kabupaten diperoleh hasil, klaster 1 terdiri dari 167 kabupaten atau kota yang paling sedikit mendapatkan medali OSN. Hanya 1-2 medali yang didapat selama tahun 2009 sampai 2024. Sedangkan klaster 2 merupakan kelompok daerah yang mendapatkan medali 123 sampai 211 yang terdiri dari 10 kota atau kabupaten. Selanjutnya untuk klaster 3, 4, 5, dan 6 merupakan kelompok daerah yang masing-masing mendapatkan medali 64-110, 245-306, 21-38, dan 7-17. Nilai  $k=3$  dan  $k=6$  pada *clustering* K-Means data OSN pada pengelompokan berdasarkan provinsi ataupun kota menunjukkan bahwa pengelompokan pada masing-masing klaster memiliki tingkat kemiripan yang tinggi. Hasil dari *clustering* per provinsi ataupun kabupaten didapat bahwa klaster 1 merupakan daerah yang memerlukan perhatian khusus dari Puspresnas, karena daerah pada klaster ini mendapatkan medali paling sedikit. Hasil dari penelitian ini dapat menjadi acuan untuk Puspresnas dalam menegakkan prinsip OSN yaitu pemerataan kesempatan bagi seluruh peserta didik Indonesia

#### Daftar Rujukan

- [1] Kementerian Pendidikan dan Kebudayaan. (2024). *Olimpiade Sains Nasional*. Diakses dari <https://osn-onmipa-bpti.kemdikbud.go.id/osn.html>
- [2] Detik.com. (2023). *Olimpiade Sains Nasional (OSN): Tujuan, Bidang Lomba, dan Jadwalnya*. Diakses dari <https://www.detik.com/edu/sekolah/d-7428572/olimpiade-sains-nasional-osn-tujuan-bidang-lomba-dan-jadwalnya>.
- [3] Pusat Prestasi Nasional. (2024). *Olimpiade Sains Nasional 2024*. Diakses dari <https://pusatprestasinasional.kemdikbud.go.id/event/riset-dan-inovasi/smp/olimpiade-sains-nasional-2024-2024-smp>.
- [4] Pusat Prestasi Nasional. (2024). *Olimpiade Sains Nasional 2024*. Diakses dari [https://pusatprestasinasional.kemdikbud.go.id/uploads/lampiran/Pdoman%20OSN%20SMA\\_MA%202024%20fix.pdf](https://pusatprestasinasional.kemdikbud.go.id/uploads/lampiran/Pdoman%20OSN%20SMA_MA%202024%20fix.pdf)
- [5] Liu, R. (2022). "Data Analysis of Educational Evaluation Using K-Means Clustering Method." *Computational Intelligence and Neuroscience*. <https://doi.org/10.1155/2022/3762431>.
- [6] Hasanah, M., Defit, S., & Nurcahyo, G. W. (2021). *Implementasi Algoritma K-Means untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA*. Diakses dari <http://www.jsisfotek.org/index.php/JSisfotek/article/view/6/6>.
- [7] Khamidah, N. (2023). "Penerapan Multi-Clustering dalam Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Berdasarkan Indeks Desa Membangun."
- [8] Jain, A.K. (2010). "Data Clustering: 50 Years Beyond K-Means." *Pattern Recognition Letters*, 31, 651-666. <https://doi.org/10.1016/J.PATREC.2009.09.011>.
- [9] Iklima, T., Pujiyanta, A., 2023. Perbandingan Metode K-Means Clustering Dan Metode Ward Dalam Mengelompokkan Peelanggan Mall. *Jurnal Fasilkom*, 13 (3), pp. 349-357.
- [10] M. R. Maarif, A. R. Saleh, M. Habibi, N. L. Fitriyani, and M. Syafrudin, "Energy Usage Forecasting Model Based on Long Short-Term Memory (LSTM) and eXplainable Artificial Intelligence (XAI)," *MPDI*, vol. 14, no. 265, 2023.
- [11] C. Hayat, "Enhanced K-Means Clustering Approach for Diagnosis Types of Acne," in *2021 2nd International Conference on Innovative and Creative Information Technology, ICITech 2021*, Institute of Electrical and Electronics Engineers

- 
- Inc., Sep. 2021, pp. 39–43. doi: 10.1109/ICITech50181.2021.9590124.
- [12] Cetinkaya, A., Peker, S. and Kuvvetli, Ü. (2024), "Analysis of countries' performances in individual Olympic Games using cluster analysis and decision trees: the case of Tokyo 2020"., *Sport, Business and Management*, Vol. 14 No. 5/6, pp. 648-666. <https://doi.org/10.1108/SBM-12-2023-0151>
- [13] Rachman, F.I., Mujadilah, S., Wahyuni, T., Anas, L., 2023. Prediksi Tingkat Kelulusan Menggunakan K-Means Pada Program Studi Informatika Unismuh Makassar. *Jurnal Fasilkom*, 13 (2), pp. 504-510.
- [14] G. Zeng, Y. Fang and X. Huang, "Research on Olympic Games Hosting Strategy Based on Machine Learning Algorithm," *2023 International Conference on Electronics and Devices, Computational Science (ICEDCS)*, Marseille, France, 2023, pp. 452-458, doi: 10.1109/ICEDCS60513.2023.00087.
- [15] Prastyabudi, W.A., Alifah, A.N., Nurdin, A. "Segmenting the Higher Education Market: An Analysis of Admissions Data Using K-Means Clustering," *Procedia Computer Science*, volume 234, 2024, Pages 96-105, <https://doi.org/10.1016/j.procs.2024.02.15>