

## Analisis Perbandingan Metode DBSCAN dan Mean Shift Dalam Mengelompokkan Data IPM Kabupaten/Kota se-Indonesia

Muhamad Rizki Fasihullisan Damopolii<sup>1</sup>, Lisna Zahrotun<sup>2</sup>, Anna Hendri Soleliza Jones

<sup>1,2</sup>Program Studi Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan

<sup>1</sup>muhamad2100018164@webmail.uad.ac.id, <sup>2</sup>lisna.zahrotun@tif.uad.ac.id, <sup>3</sup>annahendri@tif.uad.ac.id

### Abstract

*This study extends previous research that clustered the 2019 Human Development Index (HDI) data of regencies and cities in Indonesia using K-Means, K-Medoids, and Agglomerative Hierarchical Clustering (AHC). HDI is an important indicator for describing the level of regional development; therefore, clustering analysis of HDI data is needed to support more targeted development policy formulation. However, these conventional clustering methods have limitations, including the requirement to predefine the number of clusters and their limited ability to handle noise. Therefore, this study applies and compares two density-based clustering algorithms, namely DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and Mean Shift, which are capable of forming clusters automatically without specifying the number of clusters in advance and can effectively handle noise. The determination of optimal parameters for each method is conducted using the Sw/Sb Ratio metric, which measures the ratio between within-cluster and between-cluster standard deviations. The results show that Mean Shift with an optimal bandwidth parameter of 1 achieves an Sw/Sb Ratio value of 0.3609, which is better than DBSCAN with a value of 0.3739, and also outperforms the clustering methods used in previous studies, which produced a value of 0.51. These findings indicate that density-based clustering algorithms, particularly Mean Shift, provide more representative clustering results for HDI data and may serve as a more effective alternative method for analyzing human development data in Indonesia.*

*Keywords: clustering, DBSCAN, human development index, mean shift, SW/SB ratio*

### Abstrak

Penelitian ini mengembangkan studi sebelumnya yang mengelompokkan data Indeks Pembangunan Manusia (IPM) Kabupaten/Kota se-Indonesia tahun 2019 menggunakan *K-Means*, *K-Medoids*, dan *Agglomerative Hierarchical Clustering* (AHC). IPM merupakan indikator penting untuk menggambarkan tingkat pembangunan wilayah, sehingga analisis pengelompokan data IPM diperlukan guna mendukung perumusan kebijakan pembangunan yang lebih tepat sasaran. Namun, metode klasterisasi konvensional tersebut memiliki keterbatasan, antara lain memerlukan penentuan jumlah cluster di awal dan kurang mampu menangani data noise. Oleh karena itu, penelitian ini menerapkan dan membandingkan dua algoritma klasterisasi berbasis kepadatan, yaitu DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) dan *Mean Shift*, yang dapat membentuk cluster secara otomatis tanpa menentukan jumlah cluster sebelumnya serta mampu mengabaikan noise. Penentuan parameter optimal pada masing-masing metode dilakukan menggunakan metrik Rasio Sw/Sb, yang mengukur perbandingan simpangan baku dalam dan antar cluster. Hasil penelitian menunjukkan bahwa *Mean Shift* dengan parameter optimal *bandwidth* = 1 menghasilkan nilai Rasio Sw/Sb sebesar 0,3609, lebih baik dibandingkan DBSCAN dengan nilai 0,3739, serta lebih unggul dibandingkan metode klasterisasi pada penelitian sebelumnya yang menghasilkan nilai 0,51. Temuan ini mengindikasikan bahwa algoritma klasterisasi berbasis kepadatan, khususnya *Mean Shift*, mampu memberikan hasil pengelompokan yang lebih representatif pada data IPM dan berpotensi menjadi alternatif metode yang lebih efektif dalam analisis pembangunan manusia di Indonesia.

Kata kunci: DBSCAN, indeks pembangunan manusia, klasterisasi, mean shift, rasio SW/SB

### 1. Pendahuluan

Dalam era digital saat ini, pemanfaatan teknologi informasi memegang peranan penting dalam berbagai sektor, termasuk dalam bidang pembangunan manusia [1]. Salah satu indikator utama dalam menilai pembangunan suatu wilayah adalah Indeks Pembangunan Manusia (IPM), yang mencakup aspek pendidikan, kesehatan, dan standar hidup. Analisis terhadap data IPM secara menyeluruh dapat membantu mengidentifikasi kondisi sosial ekonomi suatu daerah

serta mendukung perencanaan kebijakan yang lebih efektif [2].

Metode analisis data yang banyak digunakan dalam mengkaji sektor sosial adalah klasterisasi. Beberapa penelitian sebelumnya banyak menerapkan klasterisasi untuk memetakan karakteristik pembangunan wilayah, dan untuk mengidentifikasi ketimpangan pendidikan, Kesejahteraan, maupun kesehatan, menggunakan klasterisasi dan regresi logistik untuk menemukan hubungan antara pembangunan manusia, ketimpangan ekonomi, kebahagiaan, dan kemiskinan di 98 negara

[3]. Analisis hubungan antara biaya kesehatan, tingkat kematian, dan pembangunan manusia di negara-negara anggota OECD pernah dilakukan menggunakan beberapa metode *machine learning* salah satunya klusterisasi [4]. Penggunaan *unsupervised learning* seperti PCA dan klusterisasi juga dilakukan untuk menganalisis berbagai macam variabel sosial, ekonomi, dan pembangunan di beberapa negara untuk melihat pola pembangunan yang berkelanjutan [5]. Penerapan klusterisasi dan klasifikasi dalam mengelompokkan dan memprediksi kebahagiaan global di 156 negara berbeda [6]. Selain beberapa penelitian tersebut, beberapa penelitian lain juga memanfaatkan *clustering* dalam analisis sosial ekonomi [7], [8], [9], [10].

Metode klusterisasi menjadi pendekatan yang umum digunakan untuk mengelompokkan wilayah berdasarkan kemiripan nilai IPM. Dengan teknik ini, kita dapat mengelompokkan kabupaten/kota berdasarkan karakteristik pembangunan manusianya [2]. Penelitian sebelumnya telah menerapkan metode seperti *K-Means*, *K-Medoids*, dan *Agglomerative Hierarchical Clustering* (AHC) untuk mengelompokkan data IPM Kabupaten/Kota se-Indonesia [2]. Namun, metode-metode tersebut mengharuskan pengguna untuk menentukan jumlah cluster di awal serta cenderung sensitif terhadap data *outlier* atau *noise*, sehingga hasilnya bisa kurang optimal.

Dalam konteks Indonesia, berbagai studi menunjukkan bahwa pembangunan manusia antar wilayah berlangsung secara tidak merata dan bersifat persisten dalam jangka waktu yang panjang, yang tercermin dari adanya perbedaan nilai IPM yang signifikan antara wilayah dengan tingkat pembangunan yang tinggi dan wilayah yang relatif tertinggal [11], [12], [13], [14]. Kondisi ini dipengaruhi oleh faktor struktural seperti sentralisasi pembangunan, perbedaan akses terhadap infrastruktur, layanan pendidikan dan kesehatan, serta karakteristik geografis wilayah. Perbedaan yang mencolok tersebut berimplikasi pada munculnya nilai IPM yang ekstrem antar kabupaten/kota, yang secara statistik berpotensi teridentifikasi sebagai data menyimpang (*outlier*) atau *noise* dalam proses klusterisasi, termasuk pada data IPM tahun 2019. Pemilihan data IPM tahun 2019 dalam penelitian ini dilakukan untuk menjaga konsistensi dan keterbandingan dengan penelitian sebelumnya yang menggunakan data pada tahun yang sama, sehingga perbedaan hasil pengelompokan yang diperoleh dapat secara langsung dikaitkan dengan perbedaan metode klusterisasi yang digunakan, bukan dengan perbedaan karakteristik data antar tahun.

Untuk mengatasi keterbatasan tersebut, algoritma dengan pendekatan berbasis kepadatan seperti DBSCAN dan *Mean Shift* mulai dilirik sebagai alternatif. Kedua algoritma ini memiliki keunggulan dalam mendeteksi jumlah cluster secara otomatis serta mampu mengidentifikasi dan mengabaikan *noise*

dalam data [15], [16]. Dengan demikian, pendekatan ini berpotensi memberikan hasil pengelompokan yang lebih representatif dan adaptif terhadap karakteristik data IPM yang cenderung kompleks dan tersebar tidak merata [15], [16], [17].

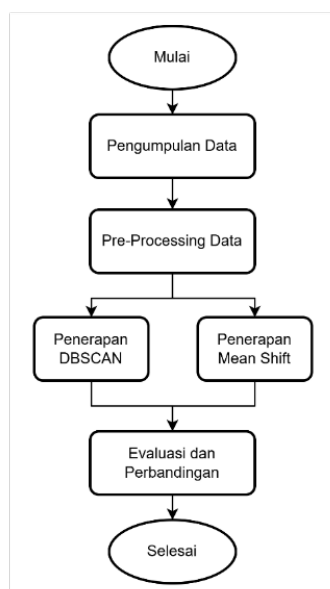
Relevansi penggunaan algoritma berbasis kepadatan tersebut juga didukung oleh penelitian lain yang menganalisis kinerja algoritma berbasis kepadatan dan membandingkannya dengan algoritma lain. Perbandingan algoritma *K-Means*, *Mean Shift*, dan DBSCAN untuk mengelompokkan data ancaman jaringan menggunakan metrik *Silhouette Score*, dan menemukan bahwa *Mean Shift* memiliki performa terbaik [15]. Penelitian lainya juga menemukan bahwa *Mean Shift* menghasilkan nilai *Silhouette Score* yang sangat tinggi dalam pengelompokan data calon penerima bantuan sosial, menunjukkan kualitas cluster yang baik [18]. Kajian literatur mengenai DBSCAN, menyimpulkan bahwa DBSCAN lebih unggul dibandingkan algoritma *non-density-based* dalam menangani bentuk data yang kompleks dan baik dalam mengidentifikasi *noise* [19]. Sementara itu, perbandingan performa *K-Means* dengan *Mean Shift* dan menunjukkan bahwa meskipun *K-Means* lebih cepat, *Mean Shift* lebih efektif untuk data yang tidak beraturan serta mampu memperkirakan jumlah *cluster* secara otomatis. Ketika estimasi jumlah *cluster* tidak diketahui [20]. Algoritma berbasis kepadatan secara umum lebih baik dalam mendeteksi *noise* dan mengelompokkan data yang kompleks [21]. Secara keseluruhan, penelitian-penelitian tersebut menunjukkan bahwa DBSCAN dan *Mean Shift* memiliki performa yang kompetitif dalam berbagai jenis data, namun penerapannya pada data IPM secara langsung masih sangat terbatas.

Oleh karena itu, penelitian ini bertujuan untuk menerapkan dan membandingkan algoritma DBSCAN dan *Mean Shift* dalam pengelompokan data IPM Kabupaten/Kota se-Indonesia tahun 2019, serta mengevaluasi keunggulan masing-masing algoritma menggunakan metrik rasio simpangan baku dalam kelompok dan antar kelompok (Sw/Sb). Metode klusterisasi berbasis densitas dipandang relevan dalam konteks pemetaan ketimpangan pembangunan wilayah karena mampu mengelompokkan wilayah berdasarkan kepadatan karakteristik pembangunan manusia tanpa memaksakan seluruh wilayah masuk ke dalam cluster tertentu, serta memungkinkan wilayah dengan karakteristik ekstrem tetap teridentifikasi sebagai *noise*. Dalam konteks data IPM Indonesia yang bersifat heterogen dan mengandung perbedaan ekstrem antar wilayah, penanganan *noise* menjadi aspek penting agar hasil pengelompokan yang diperoleh tidak bias dan lebih merepresentasikan kondisi pembangunan yang sebenarnya. Oleh karena itu, penelitian ini menekankan evaluasi performa algoritma klusterisasi berbasis densitas dalam menangani *noise* melalui perbandingan nilai Rasio Sw/Sb yang dihasilkan oleh masing-masing metode.

Selain itu, kontribusi penelitian ini terletak pada penilaian komparatif yang dilakukan secara konsisten, yaitu dengan membandingkan hasil klusterisasi DBSCAN dan *Mean Shift* tidak hanya satu sama lain, tetapi juga terhadap metode klusterisasi konvensional yang digunakan pada penelitian sebelumnya (*K-Means*, *K-Medoids*, dan AHC), sehingga diperoleh pemahaman yang lebih komprehensif mengenai efektivitas relatif masing-masing metode dalam mengelompokkan data IPM pada konteks pembangunan wilayah di Indonesia.

## 2. Metode Penelitian

Ada beberapa tahapan penelitian yang akan dilakukan dalam penelitian ini, mulai dari Pengumpulan Data, *Pre-processing Data*, Penerapan DBSCAN dan Penerapan *Mean Shift*, dan Evaluasi dan Perbandingan. Aplikasi atau *software* yang akan digunakan adalah Bahasa pemrograman R dan RStudio.



Gambar 1. Alur Tahapan penelitian

Berdasarkan gambar 1 metode penelitian ini terdiri atas beberapa tahapan, yaitu pengumpulan data, pra-pemrosesan data, penerapan algoritma DBSCAN dan *Mean Shift*, serta evaluasi dan perbandingan hasil klusterisasi. Alur tahapan penelitian ini disajikan dalam bentuk diagram alir untuk memudahkan pemahaman proses penelitian secara keseluruhan.

### 2.1. Perangkat Lunak dan Bahasa Pemrograman

Penelitian ini menggunakan bahasa pemrograman R sebagai alat utama dalam pengolahan dan analisis data. Bahasa pemrograman R dipilih karena memiliki dukungan yang kuat terhadap analisis statistik dan data mining, serta menyediakan berbagai paket yang relevan untuk analisis klusterisasi. Proses pengembangan dan eksekusi kode dilakukan menggunakan RStudio sebagai *integrated development environment* (IDE) untuk memudahkan

pengelolaan skrip, visualisasi hasil, serta reproduktibilitas analisis.

Penerapan algoritma DBSCAN dan *Mean Shift*, serta proses pra-pemrosesan dan evaluasi klusterisasi, dilakukan dengan memanfaatkan paket-paket yang tersedia dalam ekosistem R yang mendukung analisis klusterisasi dan pengolahan data statistik.

### 2.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari situs resmi Badan Pusat Statistik (BPS) yaitu [bps.go.id](http://bps.go.id). Adapun data yang digunakan adalah data Indeks Pembangunan Manusia (IPM) kabupaten/kota di Indonesia tahun 2019. Pemilihan tahun 2019 dilakukan untuk menjaga konsistensi dan kesinambungan analisis dengan penelitian terdahulu yang menggunakan periode data yang sama, sehingga perbedaan hasil yang diperoleh dalam penelitian ini dapat secara langsung dikaitkan dengan perbedaan pendekatan metode klusterisasi yang digunakan.

Indeks Pembangunan Manusia atau yang biasa disingkat IPM merupakan suatu indeks yang menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. IPM merupakan indikator penting untuk mengukur kualitas hidup manusia. IPM terdiri dari empat variabel, yaitu umur harapan hidup, harapan lama sekolah, pengeluaran per kapita disesuaikan, dan rata-rata lama sekolah, di mana keempat variabel tersebut mewakili tiga dimensi dasar, yaitu umur panjang dan hidup sehat, pengetahuan, serta standar hidup layak. *Dataset* yang digunakan dalam penelitian ini adalah data empat variabel IPM, yaitu harapan lama sekolah (HLS), pengeluaran per kapita disesuaikan (PPK), rata-rata lama sekolah (RLS), dan umur harapan hidup (UHH) pada tingkat kabupaten/kota di Indonesia tahun 2019. *Dataset* tersebut terdiri dari 515 baris dan lima kolom, yaitu Kabupaten/Kota, HLS, PPK, RLS, dan UHH.

### 2.3. Pre-Processing Data

Sebelum dilakukan proses klusterisasi, data terlebih dahulu melalui tahap pra-pemrosesan untuk memastikan kualitas dan kesesuaian data dengan metode yang digunakan. Langkah-langkah pra-pemrosesan yang dilakukan adalah sebagai berikut:

#### 2.3.1. Pemeriksaan *Missing Values*

Pemeriksaan *missing value* ini dilakukan karena data diambil dari Badan Pusat Statistik. Pada data set awal perlu dilakukan pengecekan terhadap adanya nilai kosong (*missing values*) pada *dataframe* yang ada. Data resmi dari pusat statistik biasanya terdapat 3 sampai 4 baris yang berisi keterangan tentang *dataframe* saja, bukan berisi data sehingga baris tersebut perlu dihapus.

Setelah proses identifikasi *missing value* dilakukan, tahap selanjutnya adalah pembersihan data dengan

menghapus baris dan kolom yang tidak digunakan dalam penelitian. Dalam penelitian ini, baris data yang merepresentasikan tingkat provinsi dihapus karena unit analisis yang digunakan adalah kabupaten/kota. Selain itu, hanya kolom yang relevan dengan penelitian yang dipertahankan, yaitu kolom yang menunjukkan identitas wilayah (provinsi/kabupaten/kota) dan variabel IPM.

### 2.3.2. Integrasi dan Standarisasi Data

Setelah setiap *dataset* yaitu *dataset* HLS, PPK, RLS, dan UHH dihapus bagian-bagian yang tidak digunakan, keempat *dataset* ini digabungkan dan akan di Standarisasi. Mengingat variabel-variabel memiliki satuan dan skala yang berbeda yaitu HLS, RLS, dan UHH menggunakan satuan tahun sedangkan PPK menggunakan satuan Ribu Rupiah/, maka dilakukan standarisasi. Standarisasi bekerja dengan mentransformasikan data sehingga memiliki nilai rata-rata (*mean*) sebesar 0 dan standar deviasi sebesar 1. Standarisasi ini bertujuan agar setiap variabel memiliki kontribusi yang setara dalam proses klusterisasi dan mencegah dominasi variabel tertentu akibat perbedaan skala. Sampel data yang siap digunakan dapat dilihat pada Tabel 1.

Tabel 1. Sampel Data IPM Kabupaten/Kota se-Indonesia

No.	Kabupaten/Kota	HLS	PPK	RLS	UHH
1	Simeulue	0.463	-1.193	0.524	-1.199
2	Aceh Singkil	1.053	-0.638	0.184	-0.585
3	Aceh Selatan	1.135	-0.833	0.227	-1.472
4	Aceh Tenggara	0.821	-0.877	0.870	-0.390
5	Aceh Timur	0.098	-0.680	-0.216	-0.209
...	...	...	...	...	...
514	Kota Jayapura	1.573	1.745	2.021	0.282

### 2.4. Penerapan Clustering

Proses klusterisasi dilakukan menggunakan dua metode berbeda, yaitu *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) dan *Mean Shift Clustering*. Setiap metode akan diterapkan dengan variasi parameter tertentu untuk memperoleh hasil klusterisasi yang optimal.

#### 2.4.1. Penerapan DBSCAN

Data yang telah di *pre-processing* selanjutnya akan diterapkan DBSCAN untuk dikelompokkan berdasarkan kemiripan antar data atau yang biasa disebut klusterisasi. DBSCAN atau *Density-Based Spatial Clustering Algorithm with Noise*, adalah suatu algoritma clustering yang digunakan dalam *data mining* atau *machine learning* [16], [22]. DBSCAN bekerja dengan cara mengidentifikasi kerapatan data untuk membentuk *cluster*. Sehingga *cluster* yang dihasilkan oleh DBSCAN dianggap sebagai kumpulan data yang sangat dekat atau rapat satu sama lain dan antara *cluster* yang 1 dengan lainnya dianggap tidak rapat atau dekat.

Tahapan Algoritma DBSCAN umumnya seperti berikut [16], [22]: Menentukan parameter *Epsilon*(Eps) dan minimum *points* (MinPts). Lalu, Menetapkan titik awal(p) secara random. Setiap titik nantinya akan ditentukan apakah termasuk inti (*core*), batas(*border*), atau *noise*. Untuk tiap titik, hitung jumlah titik dalam radius (Eps) dari titik tersebut (termasuk titik itu sendiri). Akan gunakan *Euclidean Distance* untuk menghitung jarak, misalkan Eps = 5 maka jarak yang <= 5 termasuk dalam radius.

Pada persamaan (1) yaitu Rumus *Euclidean Distance*,  $d(x, y)$  adalah jarak *Euclidean* antara titik x dan y,  $y_1$  dan  $x_1$  adalah variabel x dan y ke-1, sedangkan  $y_i$  dan  $x_i$  adalah variabel x dan y ke-i.

$$d(x, y) = \sqrt{(y_1 - x_1)^2 + \dots + (y_i - x_i)^2} \quad (1)$$

Jika jumlah titik dalam radius (Eps) >= MinPts, maka titik tersebut adalah titik inti(*core*). Jika jumlah titik dalam radius (Eps) < MinPts tetap titik tersebut berada dalam radius dari *core point* lainnya, maka titik tersebut adalah *border point*. Jika jumlah titik dalam radius (Eps) < MinPts dan tidak berada dalam radius dari *core point* lainnya, maka titik tersebut adalah *noise point*.

Ambil *core point* sembarang untuk memulai *cluster* baru, untuk semua titik yang dapat dijangkau dari *core point* tersebut dalam radiusnya adalah bagian dari *clusternya*. Semua *core point* yang dapat dijangkau dari titik awal cluster juga dimasukkan ke dalam *clusternya*. Dua *core points* adalah *density-reachable* jika mereka dapat dihubungkan oleh serangkaian *core points* lain di mana setiap *core point* dalam seri tersebut berada dalam radius (Eps) dari *core point* berikutnya.

Ulangi langkah 5 sampai semua *core point* telah dikunjungi. Setiap kali ada *core point* yang tidak termasuk dalam cluster maka akan jadi cluster baru.

#### 2.4.2. Penerapan Mean Shift

Selain DBSCAN, metode klusterisasi lain yang akan digunakan adalah *Mean Shift*. *Mean Shift* adalah suatu algoritma *clustering* yang mengidentifikasi cluster berdasarkan pusat kepadatan data, dimana jumlah titik data di sekitar pusat tersebut maksimum dan akan mengelompokkan data yang ada di sekitarnya [23]. Algoritma ini memiliki satu parameter utama yang perlu ditentukan, yaitu *bandwidth*, yang berfungsi sebagai lebar kernel dalam proses estimasi kepadatan. Pemilihan nilai *bandwidth* yang tidak tepat dapat secara langsung memengaruhi jumlah kluster yang terbentuk, di mana nilai *bandwidth* yang terlalu kecil cenderung menghasilkan jumlah kluster yang terlalu banyak (*over-clustering*), sedangkan nilai *bandwidth* yang terlalu besar dapat menyebabkan penggabungan beberapa kluster menjadi satu sehingga jumlah kluster menjadi terlalu sedikit (*under-clustering*).

Langkah-langkah algoritma nya kurang lebih seperti berikut [24], [25]: Menentukan kernel apa yang akan

digunakan (umumnya gaussian kernel) dan *bandwidth* yang merupakan jari-jari radius. Pada persamaan (2) yaitu kernel Gaussian,  $\|x_i - x_j\|^2$  adalah jarak *Euclidean* antara titik  $X_i$  dan  $X_j$ , sedangkan  $\sigma$  adalah *bandwith*.

$$Kernel(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

Selanjutnya, menginisialisasi pusat-pusat *centroid* awal, lalu hitung pusat massa baru menggunakan rumus *Mean-Shift* pada setiap *centroid* awal. Pada persamaan (3) untuk mencari pusat massa yang baru,  $m(x)$  adalah pusat massa yang baru,  $x$  adalah pusat massa saat ini,  $x_i$  adalah titik data dalam radius, dan  $K$  adalah fungsi kernel. Berikut rumusnya:

$$m(x) = \frac{\sum_{K(x-x_i)} x_i}{\sum_{K(x-x_i)} 1} \quad (3)$$

Ulangi langkah 3 sampai *centroid* stabil atau telah memenuhi kondisi berhenti tertentu. Akhirnya setelah iterasi selesai dan telah konvergensi, setiap *centroid* akan mewakili pusat cluster data

### 2.5. Evaluasi dan Perbandingan

Evaluasi kluster dilakukan dengan menggunakan metrik Rasio SW/SB. Rasio SW/SB atau lebih tepatnya adalah Rasio perbandingan rata-rata simpangan baku dalam kelompok (SW) dibagi dengan simpangan baku antar kelompok (SB) bisa digunakan untuk menguji dan mengevaluasi seberapa baik model kluster yang dihasilkan. Hasil kluster yang baik adalah yang memiliki nilai  $Sw$  yang kecil dan nilai  $Sb$  yang besar [26], [27]. Untuk persamaan (4) yaitu mencari nilai rata-rata simpangan baku dalam kelompok,  $K$  adalah banyaknya cluster yang terbentuk, dan  $S_k$  adalah Simpangan baku cluster ke- $k$ .

$$Sw = K^{(-1)} \sum_{k=1}^K S_k \quad (4)$$

Untuk persamaan (5) yaitu mencari simpangan baku kelompok ke- $k$  atau  $S_k$ ,  $x_i$  adalah rata-rata objek pada setiap variabel,  $\bar{x}$  adalah rata-rata keseluruhan cluster ke- $k$ , dan  $n$  adalah jumlah anggota setiap cluster.

$$S_k = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_k)^2} \quad (5)$$

Persamaan (6) untuk mencari simpangan baku antar kelompok,  $\bar{X}_k$  adalah rata-rata *cluster* ke- $k$ , dan  $\bar{X}$  adalah rata-rata seluruh kelompok.

$$Sb = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2} \quad (6)$$

Setelah Rasio SW/SB dihitung maka akan dilakukan perbandingan untuk setiap parameter yang diuji mencari mana yang lebih baik. Mengingat kedua metode secara otomatis mengidentifikasi *noise* dan memilih jumlah *cluster* yang optimal, hasil kluster berpotensi menghasilkan jumlah kluster dan *noise* yang sangat banyak sehingga sulit diinterpretasikan, sehingga nantinya hasil kluster yang dianggap optimal tidak hanya Rasio SW/SB yang rendah tetapi juga

menghasilkan jumlah kluster dan *noise* yang rendah agar interpretatif.

### 3. Hasil dan Pembahasan

Penelitian ini menggunakan data Indeks Pembangunan Manusia (IPM) tahun 2019 dari 514 Kabupaten/Kota se-Indonesia, yang mencakup empat indikator utama: harapan lama sekolah (HLS), pengeluaran per-kapita disesuaikan (PPK), rata-rata lama sekolah (RLS), dan umur harapan hidup (UHH). Keempat indikator ini merepresentasikan tiga dimensi utama IPM: kesehatan, pendidikan, dan standar hidup. Tabel 2 menampilkan statistik deskriptif dari data IPM yang digunakan.

Tabel 2. Statistik Deskriptif Data IPM 2019

Variabel	Min	Max	Mean	Standar Deviasi
HLS	3.29	17.39	12.89	1.34
PPK	4181	23851	10443	2712
RLS	0.97	12.64	8.22	1.65
UHH	55.12	77.55	69.4	3.49

Tabel 2 menyajikan statistik deskriptif data Indeks Pembangunan Manusia (IPM) Kabupaten/Kota di Indonesia tahun 2019 yang meliputi nilai minimum, maksimum, rata-rata, dan standar deviasi dari setiap variabel penyusun IPM. Variabel Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Umur Harapan Hidup (UHH) dinyatakan dalam satuan tahun, sedangkan Pengeluaran Per Kapita Disesuaikan (PPK) dinyatakan dalam satuan ribu rupiah per orang per tahun. Berdasarkan tabel tersebut, terlihat adanya rentang nilai yang cukup lebar pada seluruh variabel, khususnya pada variabel PPK dan RLS, yang mengindikasikan adanya perbedaan kondisi pembangunan manusia yang signifikan antar daerah. Perbedaan nilai minimum dan maksimum yang ekstrem, terutama pada PPK dan HLS, menunjukkan bahwa sebagian daerah masih berada pada tingkat pembangunan yang relatif rendah, sementara daerah lain telah mencapai capaian pembangunan yang jauh lebih tinggi. Kondisi ini mengindikasikan adanya ketimpangan pembangunan manusia yang cukup ekstrem antar kabupaten/kota di Indonesia, sehingga diperlukan pendekatan analisis yang mampu menangkap variasi dan anomali data tersebut secara lebih adaptif.

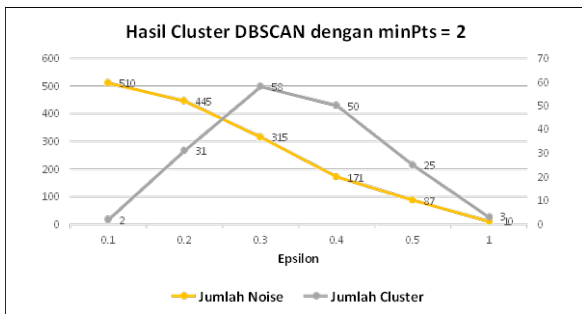
Tahap pra-pemrosesan melibatkan penghapusan nilai kosong, integrasi keempat dataset menjadi satu dataframe, serta standarisasi untuk menyamakan skala antar variabel sebelum digunakan dalam proses klusterisasi.

#### 3.1. Penerapan DBSCAN

DBSCAN diterapkan dengan berbagai kombinasi parameter epsilon (0.1–0.5, 1) dan minPts (2–5). Evaluasi dilakukan menggunakan metrik rasio SW/SB untuk menilai kualitas kluster (kompak dan terpisah),

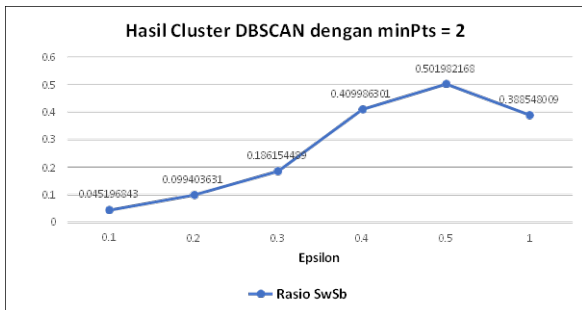
serta jumlah kluster dan proporsi data yang teridentifikasi sebagai *noise*.

Hasil awal dengan parameter epsilon = 0.1 dan minPts = 2 menghasilkan rasio Sw/Sb paling kecil (0,0452), namun 510 dari 514 data dikategorikan sebagai *noise*, yang membuat hasilnya sulit diinterpretasi. Parameter optimal dipilih pada *epsilon* = 1 dan minPts = 3, karena menghasilkan jumlah kluster yang wajar (2-3 kluster) dan data *noise* yang minimal serta rasio SW/SB terendah dari parameter-parameter lain yang menghasilkan jumlah *cluster* dan *noise* rendah. Gambar 2 sampai 9 menampilkan grafik jumlah *cluster*, jumlah *noise*, dan nilai rasio SW/SB hasil penerapan DBSCAN.



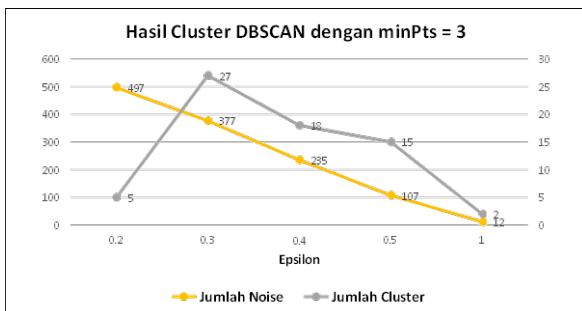
Gambar 2. Jumlah Cluster dan Noise hasil DBSCAN dengan minPts = 2

Dari gambar 2 Terlihat bahwa pada nilai eps kecil, model cenderung menganggap hampir semua data sebagai *noise*. Jumlah kluster meningkat dan kemudian menurun lagi seiring bertambahnya eps.



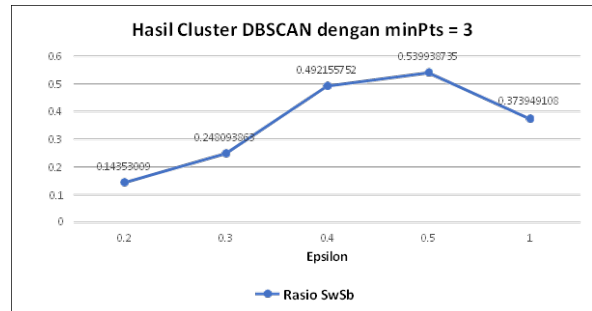
Gambar 3. Rasio Sw/Sb hasil DBSCAN dengan minPts = 2

Dari gambar 3 Rasio Sw/Sb yang rendah tidak selalu menjamin hasil kluster yang baik secara visual, terutama jika sebagian besar data tergolong *noise*.



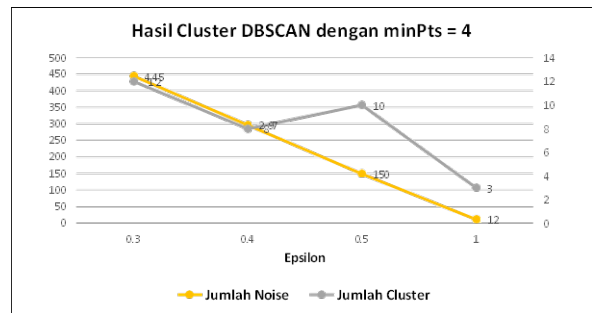
Gambar 4. Jumlah Cluster dan Noise hasil DBSCAN dengan minPts = 3

Gambar 4 menunjukkan jumlah *noise* sangat tinggi pada eps rendah karena radius pencarian terlalu sempit. Kluster baru mulai terbentuk ketika eps mencapai nilai tertentu.



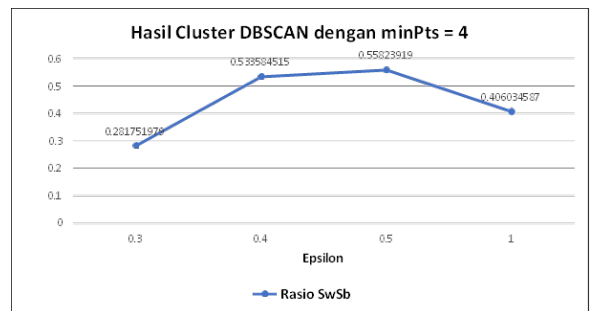
Gambar 5. Rasio Sw/Sb hasil DBSCAN dengan minPts = 3

Gambar 5 menunjukkan Interpretasi hasil *clustering* tidak bisa bergantung hanya pada rasio Sw/Sb, karena nilai kecil juga bisa muncul akibat minimnya data yang benar-benar tergolong dalam kluster.



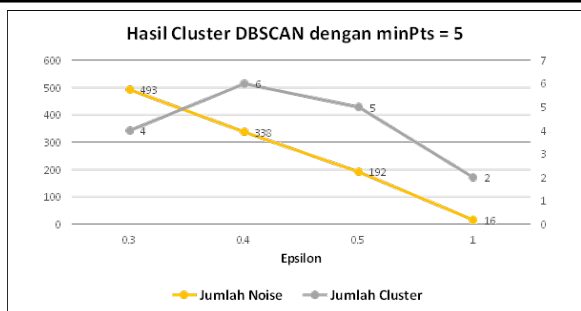
Gambar 6. Jumlah Cluster dan Noise hasil DBSCAN dengan minPts = 4

Dari gambar 6 jumlah *noise* yang tinggi dan kluster yang sedikit mengindikasikan bahwa kombinasi eps dan minPts yang digunakan masih kurang optimal.



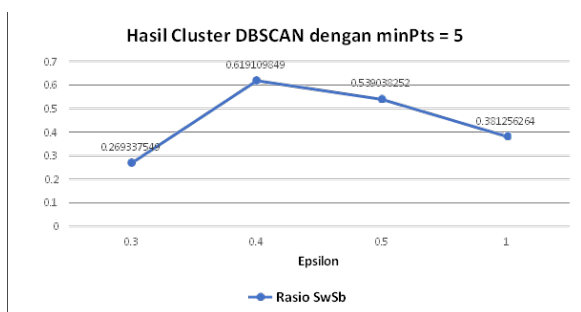
Gambar 7. Rasio Sw/Sb hasil DBSCAN dengan minPts = 4

Gambar 7 menunjukkan rasio Sw/Sb yang rendah muncul karena sebagian besar data dianggap *noise*, sehingga variansi dalam-kluster menjadi sangat kecil, bukan karena kluster terbentuk dengan baik.



Gambar 8. Rasio Sw/Sb hasil DBSCAN dengan minPts = 5

Sama seperti pada grafik-grafik diatas, dari gambar menunjukkan jumlah *noise* sangat tinggi pada eps rendah karena radius pencarian terlalu sempit.



Gambar 9. Rasio Sw/Sb hasil DBSCAN dengan minPts = 5

Sama seperti grafik sebelumnya juga, gambar 9 menunjukkan Meskipun nilai Sw/Sb mengecil, perlu dilihat kembali sebaran klaster. Nilai terlalu rendah kadang muncul akibat dominasi data *noise*.

Gambar 2 sampai 9 diatas memberikan gambaran bahwa Rasio Sw/Sb yang kecil belum tentu hasil clusternya baik, jumlah *cluster* dan *noise* juga perlu diperhatikan agar hasil *cluster* bisa di interpretasikan dengan baik. Tabel 2 menampilkan hasil penerapan DBSCAN pada berbagai nilai parameter epsilon dan minPts, yang menampilkan jumlah klaster, jumlah noise, dan nilai rasio Sw/Sb beserta parameter epsilon dan minPts yang menghasilkan masing-masing nilai tersebut.

Tidak seluruh kombinasi parameter hasil klasterisasi ditampilkan dalam bentuk visualisasi. Hal ini disebabkan oleh beberapa kombinasi parameter yang menghasilkan hanya satu klaster atau seluruh data terklasifikasi sebagai *noise*. Pada kondisi tersebut, nilai rasio simpangan baku dalam dan antar klaster (Sw/Sb) tidak dapat dihitung karena tidak terbentuk pemisahan antar klaster yang bermakna.

Tabel 3. Hasil Penerapan DBSCAN

Eps	MinPts	Jumlah Klaster	Jumlah Noise	Rasio Sw/Sb
0.1	2	2	510	0.045197
0.2	2	31	445	0.099404
0.3	2	58	315	0.186154
0.4	2	50	171	0.409986
0.5	2	25	87	0.501982
1	2	3	10	0.388548

0.2	3	5	497	0.14353
0.3	3	27	377	0.248094
0.4	3	18	235	0.492156
0.5	3	15	107	0.539939
1	3	2	12	0.373949
0.3	4	12	445	0.281752
0.4	4	8	297	0.533585
0.5	4	10	150	0.558239
1	4	3	12	0.406035
0.3	5	4	493	0.269338
0.4	5	6	338	0.61911
0.5	5	5	192	0.539038
1	5	2	16	0.381256

Tabel 3 menyajikan rekapitulasi numerik dari hasil penerapan algoritma DBSCAN yang sebelumnya telah ditampilkan dalam bentuk grafik untuk berbagai kombinasi parameter *epsilon* dan MinPts. Tabel ini merangkum jumlah klaster yang terbentuk, jumlah data yang teridentifikasi sebagai *noise*, serta nilai Rasio Sw/Sb pada setiap kombinasi parameter yang diuji. Penyajian dalam bentuk tabel dimaksudkan untuk mempermudah perbandingan antar kombinasi parameter serta mendukung penentuan parameter optimal berdasarkan nilai Rasio Sw/Sb.

### 3.2. Penerapan Mean Shift

Mean Shift diterapkan dengan variasi nilai bandwidth dari 0,1 hingga 1. Tidak seperti DBSCAN, algoritma *Mean Shift* tidak secara eksplisit mendeteksi *noise*; oleh karena itu, dalam penelitian ini klaster yang hanya terdiri dari satu data dianggap sebagai representasi *noise*. Penentuan rentang *bandwidth* dilakukan secara bertahap dengan mengamati perubahan jumlah klaster, jumlah data yang teridentifikasi sebagai *noise*, serta nilai rasio Sw/Sb. Hasil pengujian menunjukkan bahwa mulai pada nilai *bandwidth* sebesar 0,6, struktur klaster yang terbentuk cenderung stabil, di mana jumlah klaster, jumlah *noise*, serta nilai rasio Sw/Sb tidak mengalami perubahan yang signifikan hingga *bandwidth* = 1. Pengujian tidak dilanjutkan pada nilai *bandwidth* di atas 1 karena peningkatan nilai *bandwidth* yang lebih besar cenderung menggabungkan seluruh data ke dalam satu klaster tunggal, sehingga struktur klaster menjadi tidak informatif dan metrik Rasio Sw/Sb tidak dapat dievaluasi secara bermakna. Hasil optimal diperoleh pada *bandwidth* = 1, yang menghasilkan dua klaster dan satu data *noise* dengan nilai Rasio Sw/Sb sebesar 0,3610. Gambar 10 dan 11 menampilkan grafik jumlah klaster, jumlah noise, dan nilai Rasio Sw/Sb hasil penerapan *Mean Shift*.

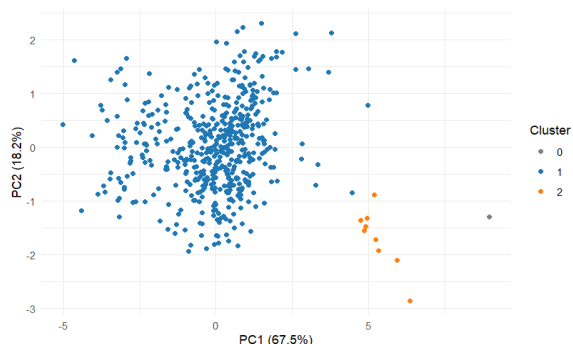


Dari gambar 12 diatas, terlihat bahwa *Mean Shift* secara konsisten menghasilkan nilai Rasio Sw/Sb yang lebih rendah dibandingkan DBSCAN, terutama pada parameter *bandwidth* = 1. Hal ini memperkuat kesimpulan bahwa algoritma *Mean Shift* lebih optimal dalam membentuk klaster yang kompak dan mudah diinterpretasikan.

### 3.4. Analisis Hasil Cluster Terbaik

Berdasarkan hasil evaluasi menggunakan Rasio Sw/Sb, *Mean Shift* dengan *bandwidth* = 1 ditetapkan sebagai metode paling optimal dalam pengelompokan data IPM Kabupaten/Kota se-Indonesia. Pada konfigurasi ini, *Mean Shift* menghasilkan dua cluster utama serta satu data yang teridentifikasi sebagai *noise*. Pemilihan metode ini menunjukkan bahwa pendekatan berbasis kepadatan mampu membentuk struktur cluster yang lebih kompak dibandingkan metode lainnya, sehingga analisis berikut difokuskan pada konfigurasi terbaik tersebut.

Untuk menampilkan hasil klasterisasi dalam bentuk visual, digunakan *Principal Component Analysis* (PCA) guna mereduksi empat variabel IPM menjadi dua komponen utama. Dua komponen pertama (PC1 dan PC2) dipilih karena mampu menjelaskan sekitar 86% variansi total, sehingga cukup representatif untuk menggambarkan struktur sebaran data dalam ruang dua dimensi. Perlu dicatat bahwa PCA hanya digunakan sebagai alat bantu visualisasi dan tidak mempengaruhi proses pembentukan cluster oleh algoritma. Hasil klasterisasi divisualisasikan dengan scatter plot pada Gambar 3 berikut.



Gambar 13. Hasil Klasterisasi Metode Terbaik *Mean Shift* dengan Bandwidth = 1

Scatter plot hasil proyeksi PCA menunjukkan pemisahan yang jelas antara dua cluster yang terbentuk. Cluster pertama yang diisi sebagian besar kabupaten/kota di Indonesia tampak lebih padat dan mencerminkan wilayah dengan karakteristik IPM yang relatif seragam, sedangkan cluster kedua yang diisi kabupaten/kota di wilayah timur Indonesia berada pada posisi yang lebih terpisah, menandakan adanya perbedaan pola IPM yang lebih rendah dibandingkan cluster pertama. Satu data terlihat terletak jauh dari kedua cluster dan dikategorikan sebagai *noise* yaitu kabupaten Nduga di papua yang merupakan

kabupaten/kota dengan IPM terendah. Visualisasi ini mengonfirmasi bahwa *Mean Shift* mampu mengelompokkan wilayah berdasarkan pola IPM secara lebih representatif dan konsisten dengan hasil evaluasi numerik sebelumnya. Hal ini sejalan dengan penelitian-penelitian internasional yang juga memanfaatkan klasterisasi untuk menganalisis indikator-indikator pembangunan manusia yang menunjukkan pemisahan klaster berdasarkan disparitas pembangunan [3], [4], [28].

### 4. Kesimpulan

Berdasarkan hasil penelitian, algoritma DBSCAN dan *Mean Shift* berhasil diterapkan dalam pengelompokan data IPM Kabupaten/Kota se-Indonesia. Kedua algoritma menunjukkan sensitivitas terhadap pemilihan parameter, sehingga proses evaluasi yang cermat sangat diperlukan untuk memperoleh hasil klasterisasi yang optimal. Hasil penelitian menunjukkan bahwa algoritma *Mean Shift* menghasilkan pengelompokan yang lebih baik dibandingkan DBSCAN, yang ditunjukkan oleh nilai rasio Sw/Sb yang lebih kecil, yaitu sebesar 0,360996, serta struktur klaster yang proporsional dan mudah diinterpretasikan, dengan terbentuknya dua klaster utama dan satu data *noise*.

Penggunaan algoritma *Mean Shift* memungkinkan proses pengelompokan wilayah dilakukan secara objektif berdasarkan pola kepadatan data IPM tanpa memerlukan penentuan jumlah klaster di awal, sehingga dapat meminimalkan bias subjektif dalam proses analisis. Dengan demikian, hasil klasterisasi yang diperoleh berpotensi membantu pemerintah dan pemangku kebijakan dalam memetakan kondisi pembangunan manusia antar wilayah secara lebih adil dan berbasis data, terutama dalam mengidentifikasi wilayah dengan karakteristik pembangunan yang berbeda atau menyimpang. Jika dibandingkan dengan metode sebelumnya seperti K-Means, K-Medoids, dan Hierarchical Clustering, pendekatan berbasis kepadatan dalam penelitian ini menunjukkan performa yang lebih unggul, baik dari sisi kualitas klasterisasi maupun kemampuannya dalam mengidentifikasi data *noise*.

Untuk penelitian selanjutnya, disarankan agar eksplorasi nilai parameter pada DBSCAN dan *Mean Shift* dilakukan lebih terarah, dengan mempersempit rentang nilai berdasarkan temuan optimal dalam penelitian ini. Selain itu, penggunaan data dari tahun yang berbeda maupun penambahan indikator pembangunan lain juga dianjurkan guna memperoleh hasil klasterisasi yang lebih menyeluruh dan representatif terhadap dinamika pembangunan wilayah.

### Daftar Rujukan

[1] H. T. L. Emilia Khristina Kiha, Sirilius Seran, "PENGARUH JUMLAH PENDUDUK, PENGANGGURAN, DAN KEMISKINAN TERHADAP INDEKS PEMBANGUNAN MANUSIA (IPM) DI KABUPATEN BELU," *Intelektiva J.*

- Ekon. Sos. Hum.*, vol. 2, no. 07, pp. 60–84, 2021.
- [2] E. Luthfi and A. W. Wijayanto, “Analisis perbandingan metode hirarchical, k-means, dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia,” *Inovasi*, vol. 17, no. 4, pp. 761–773, 2021, doi: 10.30872/jinv.v17i4.10106.
- [3] I. GEORGESCU, J. KINNUNEN, A. ANDRONICEANU, and A.-M. ANDRONICEANU, “A Computational Approach to Economic Inequality, Happiness and Human Development,” *Inform. Econ.*, vol. 24, no. 4/2020, pp. 16–28, 2020, doi: 10.24818/issn14531305/24.4.2020.02.
- [4] kartal Nazan, “Evaluation of Oecd Countries in Terms of Human Development Index , Mortality Rates and Health OecÜlkeleriniİnsaniGelişmeEndeksi , Ölüm,” *Hacettepe Sağlık İdaresi Dergisi*, 2025; 28(1), 65-80, vol. 28, no. 1, p. 16, 2025.
- [5] J. Salinas and J. Zhang, “Unveiling structural change determinants: A machine learning approach to long-term dynamics,” *Socioecon. Plann. Sci.*, vol. 101, pp. 1–17, 2025, doi: 10.1016/j.seps.2025.102290.
- [6] B. Yang and X. Xie, “Analyzing and predicting global happiness index via integrated multilayer clustering and machine learning models,” *PLoS One*, vol. 20, no. 4 April, pp. 1–20, 2025, doi: 10.1371/journal.pone.0322287.
- [7] I. Fahmiyah and R. A. Ningrum, “Human Development Clustering in Indonesia,” *J. Adv. Technol. Multidiscip.*, vol. 02, no. 01, pp. 27–33, 2023.
- [8] M. F. J. Muttaqin, “Cluster Analysis Using K-Means Method to Classify Sumatera Regency and City Based on Human Development Index Indicator,” *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 967–976, 2022, doi: 10.34123/semnasoffstat.v2022i1.1299.
- [9] A. D. Pratama and U. Ciptawaty, “Economic Spatial Patterns and Human Development Index Districts and Cities in Five Southern Sumatera Provinces,” *J. Pembang. Wil. dan Kota*, vol. 18, no. 2, pp. 192–208, 2022, doi: 10.14710/pwk.v18i2.36430.
- [10] A. Rahmat, H. Hardi, F. A. Syam, Z. Zamzami, B. Febriadi, and A. P. Windarto, “Utilization of the field of data mining in mapping the area of the Human Development Index (HDI) in Indonesia,” *J. Phys. Conf. Ser.*, vol. 1783, no. 1, 2021, doi: 10.1088/1742-6596/1783/1/012035.
- [11] J. S. R *et al.*, “Mengikis Jawa-sentris dengan Kolaborasi Berbagai Instansi,” 2022.
- [12] S. R. Azkiyah, G. Aryola, and B. D. Lukitoaji, “EDUCREATIVA : Isu Kesenjangan Pendidikan di Daerah Terpencil: Solusi untuk Mewujudkan Pendidikan yang Merata,” vol. 1, no. 1, pp. 121–129, 2025.
- [13] E. P. Ningrum, S. E. Nursyamsi, N. Siregar, U. Bhayangkara, and J. Raya, “Faktor Terkait Kesenjangan Ekonomi dan Kesejahteraan,” vol. 7, no. September, pp. 116–126, 2024.
- [14] C. F. Nisa, A. R. Y. Taher, P. Ananta, and I. W. Suparta, “ANALISIS KETIMPANGAN PEMBANGUNAN DAERAH DI 34 PROVINSI INDONESIA,” vol. 3, 2025, doi: 10.23960/efebe.v3i2.254.
- [15] T. A. Cinderatama, R. Z. Alhamri, and Y. Yunhasnawa, “Implementasi Metode K-Means, Dbscan, Dan Meanshift Untuk Analisis Jenis Ancaman Jaringan Pada Intrusion Detection System,” *INOVTEK Polbeng - Seri Inform.*, vol. 7, no. 1, p. 169, 2022, doi: 10.35314/isi.v7i1.2336.
- [16] L. Qadrini, “Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017,” *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 13, no. 2, pp. 5–11, 2020, doi: 10.36456/jstat.vol13.no2.a2886.
- [17] R. Reliovani, N. N. Syafitri Husein, K. Z. Abdurrafi, C. Rafqi, A. Husni, and M. A. Khowarizmi, “Algoritma Mean Shift untuk Menentukan Segmentasi Pelanggan pada Penjualan Toko Online,” *Gunung Djati Conf. Ser.*, vol. 3, pp. 93–98, 2021.
- [18] R. Rizuan, E. Haerani, J. Jasril, and L. Oktavia, “Penerapan Algoritma Mean-Shift Pada Clustering Penerimaan Bantuan Pangan Non Tunai,” *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 1019–1027, 2023, doi: 10.47065/josyc.v4i4.3876.
- [19] H. V. Singh, A. Girdhar, and S. Dahiya, “A Literature survey based on DBSCAN algorithms,” *Proc. - 2022 6th Int. Conf. Intell. Comput. Control Syst. ICICCS 2022*, no. Iccics, pp. 751–758, 2022, doi: 10.1109/ICICCS53718.2022.9788440.
- [20] M. Nigar Shumaila, “A Comparison of K-Means and Mean Shift Algorithms,” *Int. J. Theor. Appl. Math.*, vol. 7, no. 5, p. 76, 2021, doi: 10.11648/j.ijtam.20210705.12.
- [21] P. Bhattacharjee and P. Mitra, “A survey of density based clustering algorithms,” *Front. Comput. Sci.*, vol. 15, no. 1, 2021, doi: 10.1007/s11704-019-9059-3.
- [22] A. Rifqi and R. T. Aldisa, “Penerapan Data Mining Untuk Clustering Kualitas Udara,” *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, pp. 289–297, 2023, doi: 10.30865/json.v5i2.7145.
- [23] D. E. P. Lature and A. N. Sebayang, “Identifikasi Perilaku Pelanggan Berdasarkan Penilaian Tingkat Kepuasan Menggunakan Mean Shift Clustering,” *J. Inform. Press*, vol. 1, no. 2, pp. 9–14, 2024.
- [24] A. N. Azizah, T. Widiarihari, and A. R. Hakim, “Kernel K-Means Clustering untuk Pengelompokan Sungai di Kota Semarang Berdasarkan Faktor Pencemaran Air,” *J. Gaussian*, vol. 11, no. 2, pp. 228–236, 2022, doi: 10.14710/j.gauss.v11i2.35470.
- [25] R. F. Utari, F. Insani, S. Agustian, and L. Afriyanti, “Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mean Shift,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. July, pp. 1015–1023, 2024.
- [26] U. Rahardja, Q. Aini, and M. Iqbal, “Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward,” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 5, no. 1, pp. 40–43, 2020, [Online]. Available: <https://doi.org/10.30743/infotekjar.v5i1.2464%0Ahttps://doi.org/10.30743/infotekjar.v5i1.1997>
- [27] H. Sofyan, M. Iqbal, M. Marzuki, and M. Muhammad, “The comparison of k-modes clustering and ROCK clustering to the poverty indicator in Samadua Subdistrict, South Aceh,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1087, no. 1, p. 012085, 2021, doi: 10.1088/1757-899x/1087/1/012085.
- [28] J. L. M. Rocha, M. A. C. Zela, N. I. V. Torres, and G. S. Medina, “Analogy of the Application of Clustering and K-Means Techniques for the Approximation of Values of Human Development Indicators,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 9, pp. 526–532, 2021, doi: 10.14569/IJACSA.2021.0120959.