

Klasterisasi Kota dan Kabupaten di Jawa Timur Berdasarkan Angka Stunting Menggunakan Metode K-Means

Steven Andre Gonassis^{*,1,2,3,4}, M. Arya Putra Darmawan⁵, Peres Hendri Virgiawan⁶, Bernadus Anggo Seno Aji⁶

^{1,2,3,4} Teknologi Informasi, Fakultas Teknologi Informasi dan Bisnis, Telkom University Surabaya,
Jl. Ketintang No.156, Ketintang, Kec. Gayungan, 60231, Kota Surabaya, Indonesia
⁵ stevengonassis@student.ittelkom-sby.ac.id, ⁶ muhammadaryapd21@student.ittelkom-sby.ac.id,
peres.hendri@student.ittelkom-sby.ac.id, bernadus.seno@ittelkom-sby.ac.id

ARTICLE INFO

Article history:

Keywords:

Klasterisasi, prevalensi balita stunting, indeks daya beli, indeks kesehatan, indeks pendidikan, *k-means*.

ABSTRACT

Stunting adalah kondisi gagal tumbuh pada balita akibat kekurangan gizi kronis. *Stunting* disebabkan oleh berbagai faktor, seperti angka pernikahan dini yang tinggi dan angka pernikahan dini dipengaruhi oleh indeks daya beli, indeks kesehatan, dan indeks pendidikan. Jawa Timur merupakan provinsi dengan ekonomi terbesar kedua di Indonesia. Namun, prevalensi balita *stunting* di Jawa Timur masih belum memenuhi target nasional. Penelitian ini bertujuan untuk mengelompokkan kota/kabupaten di Jawa Timur berdasarkan angka *stunting*, indeks daya beli, indeks kesehatan, dan indeks pendidikan menggunakan metode *K-Means*. Data yang digunakan adalah data sekunder dari BPS Jawa Timur tahun 2022. Dari hasil analisis, diperoleh tiga klaster. Klaster pertama beranggotakan 8 kota/kabupaten dengan prevalensi stunting rendah, indeks daya beli tinggi, indeks kesehatan tinggi, dan indeks pendidikan tinggi. Klaster kedua beranggotakan 12 kota/kabupaten dengan prevalensi stunting tinggi, indeks daya beli rendah, indeks kesehatan rendah, dan indeks pendidikan rendah. Klaster ketiga beranggotakan 18 kota/kabupaten dengan prevalensi stunting sedang, indeks daya beli sedang, indeks kesehatan sedang, dan indeks pendidikan sedang.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Pendahuluan

Stunting adalah kondisi gagal tumbuh pada anak balita akibat kekurangan gizi kronis yang berlangsung lama[1]. *Stunting* ditandai dengan tinggi badan anak yang lebih pendek dari standar usianya. *Stunting* disebabkan oleh kekurangan gizi, antara lain: kalori, protein, zat besi, yodium, vitamin A, dan vitamin D. Kekurangan gizi kronis dapat disebabkan oleh beberapa faktor, antara lain: asupan makanan yang kurang, infeksi, cara asuh anak yang salah, dan faktor lingkungan. *Stunting* merupakan salah satu indeks pembangunan manusia. Studi WHO di Indonesia menyebutkan salah satu penyebab *stunting* di Indonesia adalah tingginya angka pernikahan dini. Angka pernikahan dini yang tinggi dipengaruhi oleh indeks daya beli, indeks kesehatan, dan indeks pendidikan[2].

Jawa Timur merupakan salah satu provinsi penopang ekonomi Indonesia. Jawa Timur menyumbang 15 persen produk domestik bruto nasional, terbesar kedua setelah DKI Jakarta [3]. Namun, Berdasarkan Survei Status Gizi Indonesia (SSGI), angka stunting di Jawa Timur pada tahun 2022 adalah 19,2 persen. Sedangkan target nasional adalah 14 persen. Selain itu, masih terdapat beberapa

kota/kabupaten di Jawa Timur yang angka prevalensi balita stuntingnya masih sangat tinggi, di atas 30 persen [4].

Berdasarkan kondisi tersebut, peneliti tertarik untuk melakukan penelitian dengan mengelompokkan kota/kabupaten di Jawa Timur berdasarkan prevalensi balita stunting, indeks daya beli, persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun, indeks kesehatan, dan indeks pendidikan. Pengelompokkan dilakukan untuk mengetahui kota/kabupaten mana saja di Jawa Timur yang membutuhkan perhatian dan pembangunan lebih dari pemerintah, khususnya dalam masalah *stunting*. Pengelompokkan dilakukan dengan menggunakan metode *clustering*.

Clustering atau klasterisasi adalah tugas mengelompokkan sekumpulan objek sedemikian rupa sehingga objek-objek dalam kelompok yang sama (klaster yang sama) lebih mirip satu sama lain dibandingkan dengan objek-objek dalam kelompok lain. Klasterisasi termasuk dalam kategori *unsupervised learning* yang berfungsi untuk mempartisi data tanpa label ke dalam satu kelompok [5]. Klasterisasi menggunakan algoritma *K-Means* adalah dengan membagi data menjadi *K cluster* dengan cara meminimalkan jumlah *euclidean distance* antara setiap titik data dan *centroid* klasternya [6]. Pengelompokkan dilakukan dengan menggunakan metode *K-Means Clustering* karena algoritma *K-Means* cukup efisien secara komputasional dan *K-Means* cukup efektif untuk melakukan eksplorasi data yang cepat [7]. Selain itu hasil *clustering* menggunakan metode *K-Means* mudah untuk divisualisasikan, sehingga memudahkan interpretasi visual dari struktur data dan hubungan antar titik data.

2. Metode Penelitian

Penelitian ini dilakukan dengan menggunakan metode *K-Means Clustering* untuk mengelompokkan kota/kabupaten di Jawa Timur berdasarkan angka stunting, tinggi angka pernikahan dini, dan indeks yang mempengaruhi tingginya angka pernikahan dini. Penelitian ini menggunakan data dari BPS Jatim (Badan Pusat Statistik Jawa Timur) tahun 2022. Set data terdiri dari 5 variabel, yaitu prevalensi balita stunting, indeks daya beli, indeks kesehatan, indeks pendidikan, dan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun [8]–[12].

Selain itu, penelitian ini juga menggunakan metode analisis deskriptif untuk melihat secara umum kota/kabupaten di Jawa Timur berdasarkan variabel-variabel penelitian yang digunakan. Penelitian ini menggunakan perangkat lunak *Google Colab* untuk mengolah data dan *Google Maps* untuk memvisualisasikan hasil *clustering*.

2.1. Data

Penelitian ini menggunakan data sekunder yang diakses dari <https://jatim.bps.go.id>. Dataset penelitian ini menggunakan 5 variabel di 38 kota dan kabupaten di Jawa Timur, yaitu:

- x_1 : prevalensi balita stunting
- x_2 : indeks daya beli
- x_3 : persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun.
- x_4 : indeks kesehatan
- x_5 : indeks pendidikan

2.2. Penentuan Jumlah Cluster Optimum

Penentuan jumlah *cluster* optimum yang digunakan dalam penelitian ini adalah *Elbow Method*. *Elbow Method* bekerja dengan cara memplot ukuran seberapa baik data dikelompokkan ke dalam klaster sebagai fungsi dari jumlah klaster. *Elbow method* menghitung menggunakan beberapa metrik, seperti *Within-Cluster Sum of Squares (WCSS)* [13]. *WCSS* adalah jumlah kuadrat jarak antara setiap titik data dengan *centroid* klasternya. Pada *Elbow Method*, titik dimana kurva *WCSS* mulai mendatar, atau titik patahan yang menyerupai siku disebut dengan *elbow*. Titik *Elbow* adalah titik dimana jumlah klaster paling optimum.

2.3. K-Means

K-Means adalah salah satu metode pengelompokan yang sangat umum digunakan. *K-Means* menggunakan nilai rata-rata sebagai pusat *centroid*. Langkah-langkah metode klasterisasi *K-Means* adalah:

1. Input data
2. Menentukan jumlah klaster yang diinginkan
3. Menentukan titik *centroid* (titik pusat klaster) dengan cara mengambil data secara acak sebagai *centroid* awal.
4. Hitung jarak data ke pusat klaster dengan menggunakan *Euclidean Distance*, yaitu:

$$dist = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \quad (1)$$

5. Data diklasifikasikan berdasarkan kedekatannya dengan *centroid*. Data dikelompokkan ke dalam klaster dengan nilai yang terdekat. Kemudian, pusat klaster baru dihitung berdasarkan rata-rata anggota yang ada pada klaster. Proses diulangi sampai hasil perulangan bernilai sama dengan perulangan sebelumnya.

2.4. Langkah Penelitian

Langkah-langkah penelitian dalam penelitian ini adalah sebagai berikut:

1. Menginputkan data yang digunakan
2. Menentukan variabel penelitian dan melihat variasi data.
3. Melakukan analisis deskriptif.
4. Melakukan normalisasi data.
5. Mencari jumlah klaster paling optimum menggunakan *Elbow Method*.
6. Melakukan klasterisasi data menggunakan *K-Means*.
7. Menganalisis hasil klasterisasi.
8. Menarik kesimpulan dari hasil klasterisasi.

3. Hasil dan Pembahasan

3.1. Menginputkan data yang digunakan

Melakukan penginputan data ke dalam sistem untuk melihat variasi data dan menentukan variabel penelitian. Variabel yang digunakan adalah x untuk prevalensi balita stunting, x untuk indeks daya beli, x untuk persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun, x untuk indeks kesehatan, dan x untuk indeks pendidikan.

Table 1. Data yang Digunakan untuk Klasterisasi

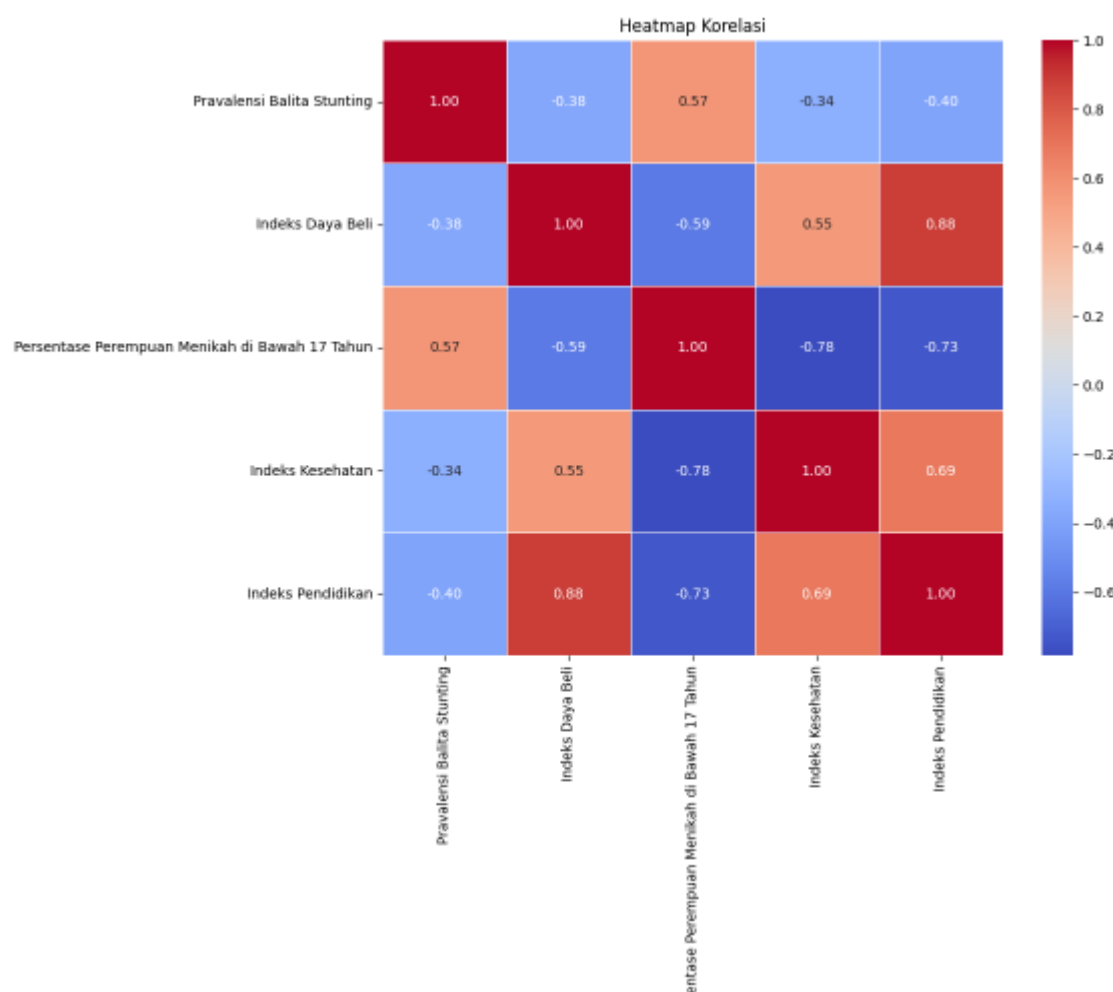
Data Ke-i	Kota/Kabupaten	x1	x2	x3	x4	x5
1	Kabupaten Pacitan	20.6	0.68	14.99	0.81	0.61
2	Kabupaten Ponorogo	14.2	0.71	20.07	0.82	0.63
3	Kabupaten Trenggalek	19.5	0.7	20.01	0.83	0.6
4	Kabupaten Tulungagung	17.3	0.73	15.42	0.84	0.65
5	Kabupaten Blitar	14.3	0.73	15.4	0.83	0.6
	...					
34	Kota Pasuruan	21.1	0.8	10.75	0.8	0.69
35	Kota Mojokerto	11.6	0.81	8.16	0.83	0.74
36	Kota Madiun	9.7	0.85	4.14	0.82	0.78
37	Kota Surabaya	4.8	0.89	6.82	0.84	0.76
38	Kota Batu	25.2	0.78	15.2	0.81	0.7

Pada tabel 1 ditampilkan data yang dari kota/kabupaten di Jawa Timur pada tahun 2022.

3.2. Analisis Deskriptif

Pada tahun 2022, Kota/kabupaten dengan prevalensi balita stunting adalah Kabupaten Jember (34.9) dan prevalensi terendah adalah Kota Surabaya (4.8). Kota dengan indeks daya beli tertinggi adalah Kota Surabaya (0.89), sedangkan indeks daya beli terendah adalah Kabupaten Sampang, Kabupaten Pamekasan, dan Kabupaten Bangkalan (0.67). Kota dengan indeks kesehatan tertinggi adalah Kabupaten Tulungagung, Kota Surabaya, Kota Kediri, dan Kabupaten Sidoarjo (0.84), sedangkan indeks kesehatan terendah adalah Kabupaten Bondowoso (0.73).

Kota/kabupaten dengan indeks pendidikan tertinggi adalah Kota Madiun dan Kota Malang (0.78) sedangkan indeks pendidikan terendah adalah Kabupaten Sampang (0.51). Kota/kabupaten dengan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun tertinggi adalah Kabupaten Bondowoso (45.83%) dan terendah adalah kota Madiun (4.14%).



Gambar 1. Heatmap Korelasi Antar Variabel.

Berdasarkan Gambar 1., angka yang melebihi positif 0.5 menunjukkan hubungan berbanding lurus. Sedangkan angka yang melebihi negatif 0.5 menunjukkan hubungan berbanding terbalik. Prevalensi balita *stunting* memiliki nilai korelasi 0.57 dengan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun, artinya semakin tinggi angka perempuan menikah di bawah 17 tahun di suatu daerah, semakin tinggi juga prevalensi balita *stunting* di daerah tersebut, begitu juga

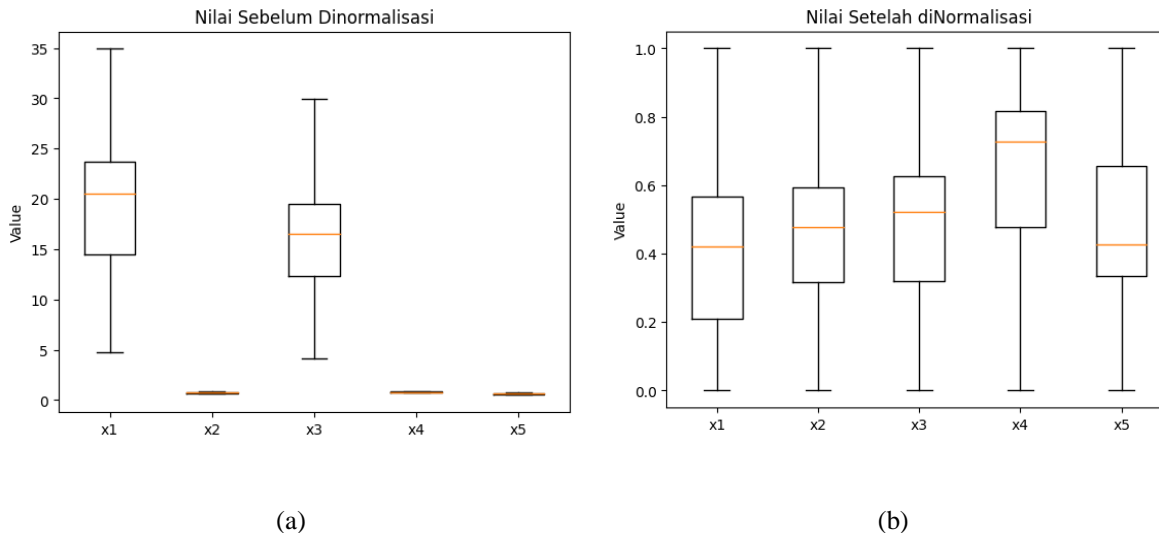
sebaliknya. Selain itu indeks daya beli memiliki nilai korelasi -0.59 dengan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun, artinya semakin tinggi indeks daya beli suatu daerah, semakin rendah angka perempuan menikah di bawah 17 tahun di daerah tersebut, begitu juga sebaliknya.

Indeks daya beli juga memiliki nilai korelasi 0.55 dengan indeks kesehatan, artinya semakin tinggi indeks daya beli semakin tinggi juga indeks kesehatannya, begitu juga sebaliknya. Indeks daya beli juga memiliki nilai korelasi 0.88 dengan indeks pendidikan. Jadi, semakin tinggi indeks daya beli, semakin tinggi indeks pendidikan suatu daerah, begitu juga sebaliknya.

Indeks kesehatan memiliki nilai korelasi -0.78 dengan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun, artinya semakin tinggi indeks kesehatan, semakin rendah angka perempuan menikah di bawah 17 tahun di daerah tersebut. Indeks pendidikan memiliki nilai korelasi -0.73 dengan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun, artinya semakin tinggi indeks pendidikan, semakin rendah angka perempuan menikah di bawah 17 tahun di daerah tersebut, begitu juga sebaliknya.

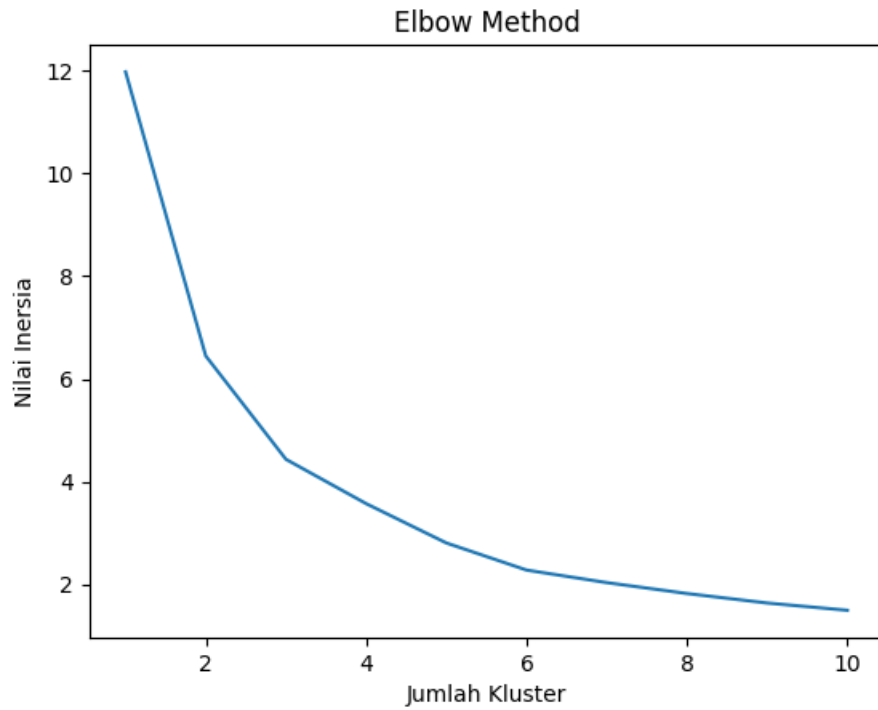
3.3. Penentuan Jumlah Cluster Optimum

Penentuan jumlah *cluster* optimum (K) dilakukan menggunakan *Elbow Method*. Penentuan kluster menggunakan metode tersebut dapat dilakukan berdasarkan plot pada Gambar 3. Dari gambar tersebut bisa dilihat bahwa titik *elbow* ada di saat K = 3. Maka dari itu, pada penelitian ini ditetapkan nilai K adalah 3. Hasil dari *K-Means* dari data yang sudah dinormalisasi diperoleh *centroid* dari tiga kluster, yaitu: kluster ke-1 pada [$x_1 = 0.27782392$; $x_2 = 0.73026316$; $x_3 = 0.10139319$; $x_4 = 0.92045455$; $x_5 = 0.90277778$], kluster ke-2 pada [$x_1 = 0.60852713$; $x_2 = 0.15789474$; $x_3 = 0.62661249$; $x_4 = 0.34090909$; $x_5 = 0.2191358$], dan kluster ke-3 pada [$x_1 = 0.50572167$; $x_2 = 0.40350877$; $x_3 = 0.46759976$; $x_4 = 0.72727273$; $x_5 = 0.47942387$]. Selanjutnya, anggota tiap kluster ditentukan berdasarkan jarak terdekat dengan *centroid* masing-masing kluster. Gambar 4 dan Gambar 5 menunjukkan visualisasi dari keanggotaan tiap kluster.

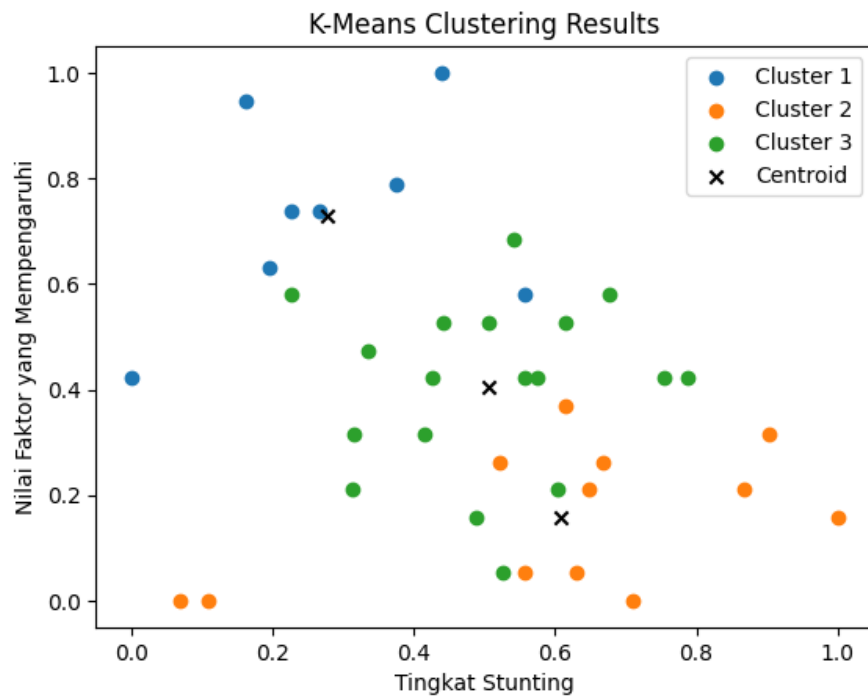


Gambar 2. *Boxplot* prevalensi balita *stunting*, indeks daya beli, indeks pendidikan, indeks kesehatan, dan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun di kota/kabupaten di Jawa Timur tahun 2022

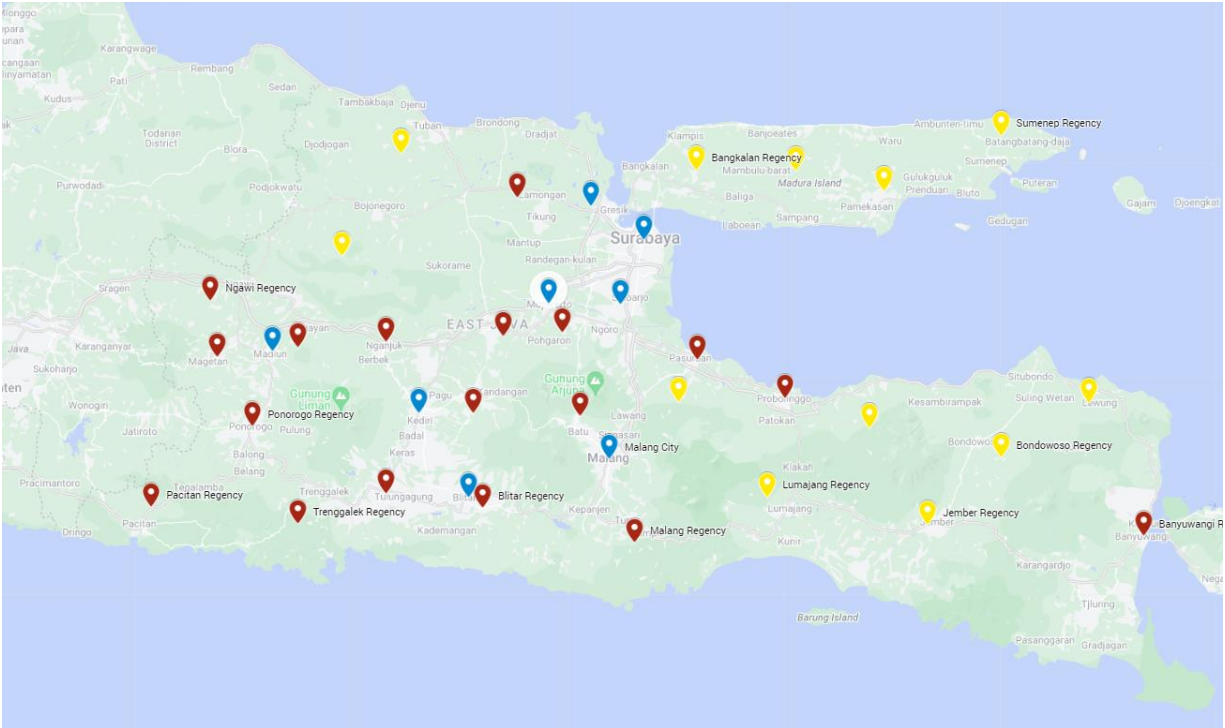
Menggunakan (a) Data Aktual; dan (b) Data yang Sudah Distandarisasi



Gambar 3. Penentuan Jumlah Kluster Optimum Menggunakan *Elbow Method*



Gambar 4. Visualisasi Keanggotaan Kluster *K-Means* ($K = 3$)



Gambar 5. Visualisasi Klasterisasi Menggunakan *Google Maps*

Klaster ke-1 direpresentasikan dengan warna biru terdiri atas 8 kota/kabupaten di Jawa Timur, yaitu Kabupaten Sidoarjo, Kabupaten Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Mojokerto, Kota Madiun, dan Kota Surabaya. Klaster ke-2 direpresentasikan dengan warna kuning terdiri atas 12 kota/kabupaten di Jawa Timur, yaitu Kabupaten Lumajang, Kabupaten Jember, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan, Kabupaten Bojonegoro, Kabupaten Tuban, Kabupaten Bangkalan, Kabupaten Sampang, Kabupaten Pamekasan, Kabupaten Sumenep. Klaster ke-3 direpresentasikan dengan warna merah terdiri atas 18 kota/kabupaten di Jawa Timur, yaitu Kabupaten Pacitan, Kabupaten Ponorogo, Kabupaten Trenggalek, Kabupaten Tulungagung, Kabupaten Blitar, Kabupaten Kediri, Kabupaten Malang, Kabupaten Banyuwangi, Kabupaten Mojokerto, Kabupaten Jombang, Kabupaten Nganjuk, Kabupaten Madiun, Kabupaten Magetan, Kabupaten Ngawi, Kabupaten Lamongan, Kota Probolinggo, Kota Pasuruan, Kota Batu.

Tiga klaster yang dihasilkan memiliki karakteristik yang berbeda. Klaster ke-1 memiliki prevalensi balita *stunting* rendah, indeks daya beli tinggi, indeks pendidikan tinggi, indeks kesehatan tinggi, dan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun rendah. Klaster ke-2 memiliki prevalensi balita *stunting* tinggi, indeks daya beli rendah, indeks pendidikan rendah, indeks kesehatan rendah, dan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun tinggi. Klaster ke-3 memiliki prevalensi balita *stunting* sedang, indeks daya beli sedang, indeks pendidikan sedang, indeks kesehatan sedang, dan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun sedang.

4. Kesimpulan

Pada penelitian ini dilakukan klasterisasi menggunakan *K-Means* dengan jumlah klaster optimum $K = 3$. Klaster ke-1 beranggotakan 21 kota/kabupaten di Jawa Timur yang butuh perhatian lebih dari pemerintah. Klaster ke-2 beranggotakan 8 kota/kabupaten di Jawa Timur yang sangat butuh perhatian lebih

ke pemerintah. Klaster ke-3 beranggotakan 9 kota/kabupaten di Jawa Timur yang sudah mendapat cukup perhatian dari pemerintah.

Maka dari itu, penelitian ini menyarankan agar pemerintah lebih menitikberatkan pembangunan pada kota/kabupaten di klaster ke-2 dan klaster ke-3. Terutama klaster ke-2 yang sangat membutuhkan perhatian lebih dari pemerintah. Anggaran untuk pembangunan daerah di klaster ke-2 harus paling banyak. Strategi yang dapat dipertimbangkan adalah peningkatan akses dan kualitas pendidikan dengan membangun sekolah-sekolah baru, memberikan beasiswa, biaya sekolah yang rendah, dan meningkatkan kualitas guru. Kemudian pemerintah bisa meningkatkan akses dan kualitas kesehatan dengan membangun puskesmas dan rumah sakit baru, bantuan biaya kesehatan, meningkatkan kualitas tenaga kesehatan, dan meningkatkan sanitasi. Pemerintah juga perlu meningkatkan ekonomi masyarakat dengan mendukung UMKM, memberikan pelatihan kepada masyarakat, dan menciptakan lapangan kerja baru. Sehingga dengan diterapkannya strategi-strategi itu, diharapkan bisa menurunkan prevalensi balita *stunting*, meningkatkan indeks daya beli, meningkatkan indeks pendidikan, meningkatkan indeks kesehatan, dan menurunkan persentase penduduk perempuan usia 10 tahun ke atas dengan usia kawin pertama di bawah 17 tahun.

Referensi

- [1] "Malnutrition." Diakses: 7 Januari 2024. [Daring]. Tersedia pada: https://www.who.int/health-topics/malnutrition#tab=tab_1
- [2] "Genbest." Diakses: 10 Januari 2024. [Daring]. Tersedia pada: <https://genbest.id/articles/bahaya-pernikahan-dini-sebagai-penyebab-stunting>
- [3] "Provinsi Jawa Timur: Perekonomian Terbesar Kedua di Jawa – Kompaspedia." Diakses: 10 Januari 2024. [Daring]. Tersedia pada: <https://kompaspedia.kompas.id/baca/profil/daerah/provinsi-jawa-timur>
- [4] "Meski Turun Tajam, Jatim Jadi Perhatian Utama Stunting dan Kemiskinan Ekstrem | Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan." Diakses: 10 Januari 2024. [Daring]. Tersedia pada: <https://www.kemenkopmk.go.id/meski-turun-tajam-jatim-jadi-perhatian-utama-stunting-dan-kemiskinan-ekstrem>
- [5] S. Suryadi, "PENERAPAN METODE CLUSTERING K-MEANS UNTUK PENGELOMPOKAN KELULUSAN MAHASISWA BERBASIS KOMPETENSI Oleh," 2018.
- [6] "Algoritma K-Means Clustering. Dalam data mining, terdapat beberapa... | by Resti Noor Fahmi | Medium." Diakses: 10 Januari 2024. [Daring]. Tersedia pada: <https://restinf21.medium.com/algoritma-k-means-clustering-70625a84eb5>
- [7] "Kelebihan dan Kekurangan Jenis Clustering dalam Data Science." Diakses: 10 Januari 2024. [Daring]. Tersedia pada: <https://blog.algorit.ma/jenis-clustering/#1-k-means-clustering>
- [8] "Badan Pusat Statistik." Diakses: 7 Januari 2024. [Daring]. Tersedia pada: <https://malangkota.bps.go.id/indicator/26/515/1/indeks-pendidikan-menurut-kabupaten-kota-di-jawa-timur.html>
- [9] "Badan Pusat Statistik." Diakses: 7 Januari 2024. [Daring]. Tersedia pada: <https://malangkota.bps.go.id/indicator/26/514/1/indeks-kesehatan-menurut-kabupaten-kota-di-jawa-timur.html>
- [10] "BPS Provinsi Jawa Timur." Diakses: 7 Januari 2024. [Daring]. Tersedia pada: <https://jatim.bps.go.id/statictable/2023/06/12/2851/persentase-penduduk-perempuan-usia-10-tahun-ke-atas-di-jawa-timur-dengan-usia-kawin-pertama-di-bawah-17-tahun-dirinci-menurut-kabupaten-kota-2020-2022.html>

- [11] "Badan Pusat Statistik." Diakses: 7 Januari 2024. [Daring]. Tersedia pada:
<https://malangkota.bps.go.id/indicator/26/516/1/indeks-daya-beli-pengeluaran-menurut-kabupaten-kota-di-jawa-timur.html>
- [12] "Ini Rincian Angka Balita Stunting di Wilayah Jawa Timur pada 2022, Kabupaten Jember Terbesar." Diakses: 7 Januari 2024. [Daring]. Tersedia pada:
<https://databoks.katadata.co.id/datapublish/2023/02/14/ini-rincian-angka-balita-stunting-di-wilayah-jawa-timur-pada-2022-kabupaten-jember-terbesar>
- [13] "Machine Learning: K-Means Clustering - Page 2 of 4 - Artificial intelligence and data science." Diakses: 10 Januari 2024. [Daring]. Tersedia pada:
<https://www.megabagus.id/k-means-clustering/2/>