

Perbandingan Kinerja Agglomerative *Clustering* Pada Data Stunting Untuk Segmentasi Wilayah

Hardian Oktavianto^{1,*}, Budi Satria Bakti², Adi Fajaryanto Cobantoro³

¹ Teknik Informatika; Universitas Muhammadiyah Jember; Jl. Karimata 49 Sumbersari Jember Jawa Timur, (0331) 336728; e-mail: hardian@unmuhjember.ac.id,

² Manajemen Informatika; Universitas Muhammadiyah Jember; Jl. Karimata 49 Sumbersari Jember Jawa Timur ; (0331) 336728; e-mail: satrio93@gmail.com.

³ Teknik Informatika; Universitas Muhammadiyah Ponorogo; Jl. Budi Utomo No.10 Ponorogo, Jawa Timur, (0352) 481124, 487662 (psw 2211); e-mail: adifajar@umpo.ac.id.

* Korespondensi: e-mail: hardian@unmuhjember.ac.id

Diterima:06 November 2023; Review: 07 November 2023; Disetujui: 04 Desember 2023

Cara sitasi: Oktavianto H, Bakti BS, Cobantoro AF. 2023. Perbandingan Kinerja Agglomerative *Clustering* Pada Data Stunting Untuk Segmentasi Wilayah. Bina Insani ICT Journal. Vol 10(2): 145 - 153.

Abstrak: Penelitian ini melakukan segmentasi wilayah berdasarkan faktor – faktor yang mempengaruhi stunting melalui teknik data mining yaitu agglomerative *clustering*. Hasil uji coba menunjukkan ketika menggunakan *single linkage*, maka *cluster* optimal terjadi ketika terbentuk 10 *cluster*, sedangkan ketika menggunakan *average linkage* dan *complete linkage* maka *cluster* optimal terjadi ketika terbentuk 3 *cluster*. Hasil analisis yang dapat diambil dari terbentuknya 3 *cluster* adalah, untuk Kabupaten/Kota yang menjadi anggota *cluster* 1 merupakan daerah yang mempunyai tingkat kepadatan penduduk tinggi, sedangkan Kabupaten/Kota yang menjadi anggota *cluster* 3 adalah daerah yang mempunyai tingkat kepadatan penduduk rendah, sedangkan anggota *cluster* 2 adalah daerah yang mempunyai tingkat kepadatan penduduk diantara anggota *cluster* 1 dan anggota *cluster* 3. Berkaitan dengan data stunting, maka kabupaten kotamadya yang termasuk dalam *cluster* 1 memiliki tingkat stunting rendah, sedangkan *cluster* 2 memiliki tingkat stunting sedang, dan *cluster* 3 adalah kabupaten kotamadya yang termasuk ke dalam daerah yang memiliki tingkat stunting tinggi.

Kata kunci: agglomerative *clustering*, segmentasi, stunting

Abstract: This research segments regions based on factors that influence stunting using data mining techniques, namely agglomerative clustering. The test results show that when using single linkage, the optimal cluster occurs when 10 clusters are formed, whereas when using average linkage and complete linkage, the optimal cluster occurs when 3 clusters are formed. The results of the analysis that can be taken from the formation of 3 clusters are that the regencies/cities that are members of cluster 1 are areas that have a high level of population density, while the regencies/cities that are members of cluster 3 are areas that have a low level of population density, while the cluster members 2 is an area that has a population density level between members of cluster 1 and members of cluster 3. Regarding stunting data, the municipal districts included in cluster 1 have low levels of stunting, while cluster 2 has moderate levels of stunting, and cluster 3 is a municipal district that including areas that have high levels of stunting.

Keywords: agglomerative clustering, segmentation, stunting

1. Pendahuluan

Stunting adalah sebuah kondisi dimana terjadi kekurangan gizi. Seorang wanita atau ibu yang mengalami stunting saat masih anak – anak lebih mempunyai resiko melahirkan anak

stunting juga. Hal ini kemudian menyebabkan lingkaran stunting antargenerasi yang sulit untuk diselesaikan [1]. Stunting juga dikaitkan dengan kondisi ekonomi suatu negara. Namun, hal ini tampaknya tidak relevan di Indonesia, di mana kasus stunting tetap tinggi meskipun ada kemajuan ekonomi yang nyata [2]. Pada tahun 2018, terdapat bukti bahwa lebih dari separuh anak stunting di bawah usia lima tahun tinggal di Asia, terhitung 55% dari semua anak yang terkena dampak tersebut [3]. Stunting merupakan pertanda atau indikator malnutrisi yang berlangsung secara terus menerus dalam waktu yang lama yang disebabkan oleh kekurangan pangan dan gizi jangka panjang [4]. Postur tubuh atau bentuk tubuh dari ibu, jarak kehamilan yang satu dengan kehamilan selanjutnya yang terlalu dekat, usia ibu yang masih tergolong remaja, serta pertumbuhan janin merupakan faktor-faktor yang mempengaruhi perkembangan janin dan risiko stunting [5]. Indonesia menempati peringkat ketiga di kawasan Asia Tenggara untuk prevalensi kejadian tertinggi, dengan rata-rata prevalensi 36,4% dari tahun 2005 hingga 2017 [6].

Nutrisi merupakan salah satu faktor yang mempengaruhi stunting, dimana nutrisi adalah komponen penting bagi kelangsungan hidup, tumbuh kembang anak. Kondisi kekurangan gizi atau malnutrisi merupakan masalah global dalam dunia kesehatan. Korban utama dari fenomena malnutrisi global ini adalah anak-anak di bawah usia lima tahun. Istilah malnutrisi dapat digunakan untuk segala bentuk gizi buruk; terutama terdiri dari kekurangan gizi atau kelebihan gizi. Kurang gizi banyak terjadi pada anak usia prasekolah. Kita dapat mendefinisikan kekurangan gizi sebagai kekurangan jumlah protein dan nutrisi penting yang merupakan kebutuhan utama untuk pertumbuhan yang baik [4].

Data mining adalah proses menemukan informasi berharga dari kumpulan data yang sangat besar. *Data mining* menggunakan berbagai metode untuk menganalisis data dan memperoleh informasi berharga. [7]. *Clustering* adalah suatu metode analisis data dengan cara mengelompokkan objek-objek data yang sejenis satu sama lain dalam satu *cluster* yang sama dan berbeda dengan objek-objek pada *cluster* lainnya [8]. *Clustering* adalah teknik data mining yang bertujuan untuk menetapkan atau mengelompokkan objek-objek ke dalam sekumpulan kelompok yang disebut *cluster*, sehingga objek-objek yang tergabung dalam sebuah *cluster* yang sama memiliki karakteristik yang lebih mirip satu sama lain dibandingkan objek-objek yang menjadi anggota dari *cluster* yang berbeda. [9].

Beberapa penelitian dengan topik *clustering* dan menggunakan data stunting telah dilakukan, diantaranya adalah aplikasi cerdas untuk mengidentifikasi kondisi gizi balita di Indonesia dibangun dengan menggunakan teknik *clustering* K-Means [10], pengelompokan Kabupaten/Kota Di Jawa Timur Berdasarkan Kasus Stunting Balita Menggunakan Algoritma *Fuzzy Particle Swarm Optimization-Fuzzy C-Means* [11], dan penelitian tentang Optimasi *Cluster* Pada Data Stunting menggunakan Teknik Evaluasi *Cluster Sum of Square Error* dan *Davies Bouldin Index* [12]. Semua penelitian tersebut menggunakan algoritma *clustering* berbasis partisi.

Penelitian kali ini melakukan pengelompokan dengan dataset faktor – faktor yang mempengaruhi stunting pada Kabupaten Jawa Timur melalui penerapan teknik *data mining* yaitu *clustering* yang membandingkan *algoritma agglomerative hierarchical clustering* yaitu *single linkage, average linkage, dan complete linkage*, selanjutnya hasil *cluster* akan divalidasi. *Cluster* yang terbentuk nantinya dapat digunakan untuk menentukan apakah langkah-langkah dan inisiatif yang diambil telah dilaksanakan di bidang yang relevan, sehingga memungkinkan penerapan langkah-langkah untuk mengatasi keterbelakangan pertumbuhan dengan lebih tepat sasaran. Lebih lanjut, hasil pengelompokan di Provinsi Jawa Timur dapat dijadikan acuan pemerintah dalam memerangi epidemi anak stunting di tingkat nasional.

2. Metode Penelitian

Fase atau langkah penelitian yang dilakukan umumnya terdiri dari empat fase. Diawali dengan tinjauan pustaka, dilanjutkan dengan pengumpulan dan pengolahan data, dilanjutkan dengan melakukan *clustering*, dan terakhir penarikan kesimpulan atau analisis hasil. Tahapan atau langkah umum penelitian ini ditunjukkan pada gambar 1.



Sumber: Hasil Penelitian (2023)

Gambar 1. Tahapan Penelitian

Studi literatur meliputi pemeriksaan referensi berupa jurnal penelitian, artikel, dan referensi lain yang berkaitan dengan penelitian untuk melengkapi pengetahuan awal dan memahami teori yang dapat digunakan untuk menunjang penelitian.

Dataset yang digunakan adalah data rata – rata lama pemberian ASI, data persentase anak batuta berdasarkan jenis makanan yang diterima, dan data persentase balita yang mendapatkan imunisasi. Data ini berasal dari situs resmi BPS [13].

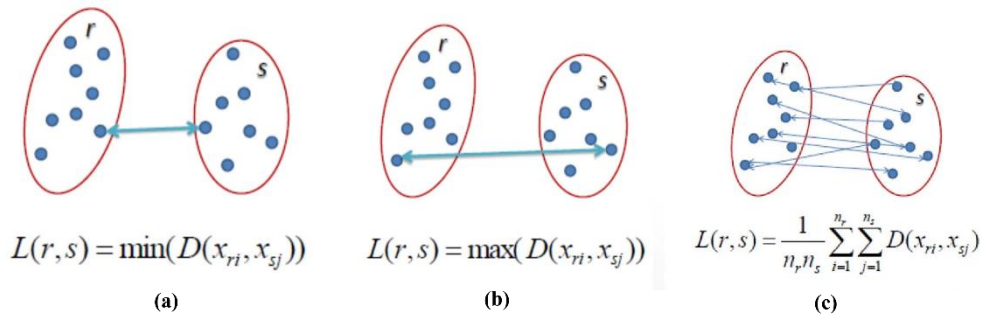
Proses *clustering* menggunakan *agglomerative hierarchical clustering*, dan *sillhouette index* digunakan untuk melakukan validasi *cluster*. Algoritma *hierarchical agglomerative clustering* mengelompokkan elemen – elemen yang mempunyai kemiripan ke dalam sebuah *cluster*. Adapun tahapan atau langkah – langkah *hierarchical agglomerative clustering* yang bekerja secara *bottom-up* [14], adalah sebagai berikut :

1. Setiap obyek dikelompokkan sendiri – sendiri atau independen (*leaf*).
2. Dua buah obyek yang memiliki ciri yang mirip berdasarkan kedekatan jarak dimasukkan atau dikelompokkan dalam sebuah *cluster*.
3. Langkah kedua tersebut terus diulang sampai hanya terbentuk sebuah *cluster* besar (*root*)

Dendrogram dapat digunakan untuk menjelaskan proses pengelompokan dan menentukan jumlah *cluster* yang optimal. Setelah dendrogram dibuat, jumlah *cluster* optimal dapat ditentukan sebagai berikut:

1. Tentukan jarak vertikal terjauh yang terbentuk pada dendrogram dan tidak memotong *cluster* lainnya.
2. Tariklah garis mendatar antara jarak vertikal titik 1.
3. Jumlah *cluster* optimal dapat diketahui dari banyaknya garis vertikal yang dipotong oleh suatu garis horizontal.

Pada *hierarchical agglomerative clustering* terdapat istilah *linkage*, yaitu kriteria yang digunakan acuan terhadap bagaimana jarak antar *cluster* dihitung. Secara umum terdapat 3 buah *linkage*, yaitu *Single linkage*, *Complete linkage*, dan *Average linkage*.



Sumber: Hasil Penelitian (2023)

Gambar 2 (a) *Single linkage* (b) *Complete linkage* (c) *Average linkage*

Single linkage → Jarak antara dua *cluster* adalah jarak terpendek antara dua titik pada setiap *cluster*

Complete linkage → Jarak antara dua *cluster* adalah jarak terjauh antara dua titik pada setiap *cluster*

Average linkage → Jarak antar *cluster* adalah jarak rata-rata antara setiap titik dalam satu *cluster* dengan setiap titik di *cluster* lainnya

Sedangkan untuk mencari jarak atau menghitung jarak antar elemen akan digunakan rumus *euclidean* $\sqrt{(a - c)^2 + (b - d)^2}$ jika $x = (a, b)$ dan $y = (c, d)$.

Analisis hasil adalah tahap memberikan penjelasan tentang *cluster* yang terbentuk dan disebut juga sebagai *cluster profiling*. Hasil yang diharapkan adalah diperolehnya informasi mengenai kasus stunting berdasarkan atribut yang digunakan.

Silhouette index menetapkan metrik kualitas yang disebut *silhouette-width* ke setiap titik data di kluster C_k . *silhouette-width* pada *cluster* C_k merupakan indikator keyakinan pada keanggotaan titik data ke- i . Metode ini membantu menghitung *silhouette-width* untuk setiap titik data, rata-rata *silhouette index* untuk setiap *cluster*, dan rata-rata *silhouette index* keseluruhan. *Silhouette index* ditentukan menggunakan rumus di bawah ini.

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

dimana $a(i)$ adalah jarak rata-rata antara titik data ke- i dan semua titik data dalam *cluster* yang sama C_k . $b(i)$ adalah jarak rata-rata terpendek antara titik data ke- i dan semua titik data lain yang tidak berada dalam *cluster* yang sama. Nilai $s(i)$ berkisar dari -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa titik data i ditugaskan ke *cluster* yang benar, sedangkan nilai yang mendekati -1 menunjukkan bahwa titik data salah klasifikasi. Angka yang mendekati nol menunjukkan bahwa titik data yang terkandung dalam satu *cluster* sama jauhnya dari *cluster* lain dan mungkin terkandung di dalamnya. C_k adalah rata-rata lebar siluet yang menunjukkan heterogenitas *cluster* tertentu.

$$S_j = \frac{1}{n} \sum_{i=1}^n s(i)$$

dimana n adalah jumlah elemen pada $s(i)$. Nilai global silhouette-width (GS) dihitung menggunakan

$$GS = \frac{1}{K} \sum_{j=1}^K S_j$$

Formasi *cluster* yang memiliki nilai GS tertinggi selanjutnya dipilih atau ditetapkan sebagai jumlah *cluster* yang optimal [15].

3. Hasil dan Pembahasan

Dataset yang digunakan pada penelitian ini data rata – rata lama pemberian ASI, data persentase anak batuta berdasarkan jenis makanan yang diterima, dan data persentase balita yang mendapatkan imunisasi. Data tersebut mengambil dari website resmi BPS, dengan 3 tabel berbeda yaitu Tabel 1. Tabel Rata-Rata Lama Pemberian ASI Kepada Anak Usia 0-23 Bulan di Jawa Timur Menurut Kabupaten/Kota dan Jenis Kelamin, Tabel 2. Tabel Persentase Balita di Jawa Timur Dirinci Menurut Kabupaten/Kota, Jenis Kelamin dan jenis Imunisasi yang Didapatkan, dan Tabel 3. Tabel Persentase Anak Usia 0-23 Bulan (Baduta) di Jawa Timur Dirinci Menurut Kabupaten/Kota, Jenis Kelamin dan Cairan/Makanan yang Diterima dalam 24 Jam Terakhir [13].

Tabel 1. Lama Pemberian ASI Anak Usia 0-23 Bulan

Kabupaten/Kota <i>Region</i>	Pemberian ASI	Tanpa Makanan Pendamping	Dengan Makanan Pendamping
Pacitan	10,50	4,30	6,19
Ponorogo	11,39	4,30	7,09
Trenggalek	11,60	4,05	7,55
Tulungagung	10,41	4,34	6,07
Blitar	10,21	3,82	6,39
Kediri	10,55	4,47	6,08
Malang	10,05	3,76	6,28
...
...
Probolinggo	11,03	3,52	7,51
Pasuruan	9,14	2,51	6,63
Mojokerto	10,01	4,41	5,60
Madiun	10,25	4,13	6,12
Surabaya	10,44	4,79	5,65
Batu	9,77	2,95	6,82

Sumber: Hasil Penelitian (2023)

Pada tabel 1 menunjukkan sebagian data dimana memuat 3 buah atribut yaitu Pemberian ASI, Tanpa Makanan Pendamping, Dengan Makanan Pendamping. Data ini diperoleh dari BPS Propinsi Jawa Timur. 3 buah atribut menunjukkan rata – rata pemberian ASI oleh ibu kepada bayi, baik yang dibarengi dengan makanan pendamping atau tidak dibarengi dengan makanan pendamping.

Tabel 2 Jenis Imunisasi yang Didapatkan

Kabupaten/Kota	BCG	DPT	Polio	Campak	Hepatitis B
Pacitan	95,55	94,11	96,20	82,66	94,15

Ponorogo	97,36	92,89	96,64	76,30	91,39
Trenggalek	97,27	94,72	95,94	82,37	97,68
Tulungagung	96,21	87,97	96,62	71,01	87,05
Blitar	94,60	89,10	92,54	80,18	91,05
Kediri	94,92	92,38	93,57	75,70	94,05
Malang	95,09	90,71	92,87	80,13	90,40
...
...
Sidoarjo	95,65	86,82	91,49	77,35	85,63
Jombang	96,00	92,43	94,30	82,01	94,73
Nganjuk	98,77	95,78	96,78	87,62	97,39
Pasuruan	97,40	91,28	94,51	76,23	93,84
Mojokerto	97,02	95,53	96,75	83,50	97,02
Madiun	94,57	96,03	94,07	88,17	97,51
Surabaya	93,79	84,12	89,38	70,07	83,02
Batu	94,07	88,64	91,82	79,27	89,66

Sumber: Hasil Penelitian (2023)

Tabel 2 memuat 5 atribut yaitu BCG, DPT, Polio, Campak, Hepatitis B. Tabel ini menunjukkan sebagian data kabupaten atau kota di Jawa Timur yang menyajikan jumlah persentase dari total bayi usia tiga tahun yang telah mendapatkan imunisasi, baik imunisasi BCG, imunisasi DPT, imunisasi Polio, imunisasi Campak, dan imunisasi Hepatitis B.

Tabel 3 Jenis Makanan/Cairan Yang Diterima

Kabupaten/Kota	Air Putih	Air Tajin/Madu/The/Air Gula	Bubur/Nasi/Roti/Mie/Jagung	Kacang-kacangan	Susu selain ASI, Keju	Daging/Hati/Jeroan/Ikan	Telur	Sayuran	Buah-buahan	Lainnya (Kue)
Pacitan	62,39	27,20	69,44	16,45	29,84	22,90	30,92	46,09	53,81	12,96
Ponorogo	81,99	40,77	77,24	25,82	40,39	37,12	54,92	67,66	58,28	44,04
Trenggalek	78,29	45,16	70,00	25,57	28,00	40,89	54,58	63,60	61,03	44,62
Tulungagung	66,28	38,29	65,68	26,12	45,51	26,92	40,62	47,86	48,11	15,58
Lumajang	69,07	29,15	61,55	21,82	38,31	30,10	46,13	63,26	38,35	23,08
Banyuwangi	68,46	32,00	66,31	23,15	37,64	34,63	37,94	54,46	48,41	30,10
Bondowoso	77,36	36,66	72,49	20,10	23,24	23,73	38,58	50,68	43,80	22,78
Situbondo	80,57	35,81	73,59	26,82	39,81	39,31	45,00	47,56	35,78	52,13
Batu	71,88	58,41	6,62	31,19	58,71	45,94	54,53	59,17	64,98	33,54

Sumber: Hasil Penelitian (2023)

Tabel 3 memuat informasi sebagian data Cairan/Makanan yang diterima dalam 24 Jam Terakhir, menyajikan data tentang makanan baik berupa makanan padat atau makanan cair yang diberikan kepada bayi di setiap kota atau kabupaten di Jawa Timur. Tabel 3 memuat 10 atribut yaitu, Air Putih, Air Tajin/Madu/The/Air Gula, Bubur/Nasi/ Roti/Mie/ Jagung, Kacang-kacangan, Susu selain ASI atau Keju, Daging/Hati/Jeroan/Ikan, Telur, Sayuran, Buah-buahan, Lainnya (Kue).

Dari ketiga buah tabel yaitu tabel lama pemberian ASI, tabel imunisasi, dan tabel jenis makanan/cairan, kemudian digabungkan menjadi sebuah tabel dataset utuh. Penggabungan dilakukan berdasarkan kesesuaian atribut kabupaten/kota. Potongan dataset dapat dilihat pada tabel 4.

Tabel 4. Potongan Dataset

Kab/Kota	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Pacitan	10,5	4,3	6,1	62,3	27,2	69,4	16,4	29,8	22,9	30,9	46,0	53,8	12,9	95,5	94,1	96,2	82,6	94,1
Ponorogo	11,3	4,3	7,0	81,9	40,7	77,2	25,8	40,3	37,1	54,9	67,3	58,2	44,0	97,3	92,8	96,6	76,3	91,3

Trenggalek	11,6	4,05	7,5	78,2	45,1	70,5	25,7	28	40,8	54,5	63,6	61,0	44,6	97,2	94,7	95,9	82,3	97,6
Tulungagung	10,4	4,3	6,0	66,2	38,2	65,6	26,1	45,5	26,9	40,6	47,8	48,1	15,5	96,2	87,9	96,6	71,0	87,0
Blitar	10,2	3,8	6,3	68,5	39,0	59,0	19,4	41,7	31,4	29,6	50,1	34,6	27,3	94,6	89,1	92,5	80,4	91,0
Kediri	10,5	4,4	6,0	80,1	40,0	68,0	20,4	22,1	23,9	44,8	63,0	51,8	21,3	94,9	92,3	93,5	75,7	94,0
...
...
...
Malang	10,0	3,7	6,2	74,1	44,6	74,7	21,7	37,3	23,3	46,5	56,8	39,8	30,2	95,0	90,7	92,8	80,1	90,4
Lumajang	10,1	4,2	5,9	69,0	29,1	61,5	21,8	38,3	30,1	46,1	63,2	38,3	23,0	95,2	89,8	94,1	77,2	88,8
Jember	10,2	5,0	5,1	78,9	37,6	68,5	25,6	28,4	25,8	42,7	53,0	45,7	34,2	95,4	91,1	95,6	81,8	89,8
Banyuwangi	10,1	4,4	5,7	68,4	32,3	66,1	23,6	37,6	34,6	37,9	54,4	48,4	30,1	94,2	87,4	89,4	75,5	89,3
	7	2	5	6	1	5	4	3	4	6	1	1	4	1	4	1	1	9

Sumber: Hasil Penelitian (2023)

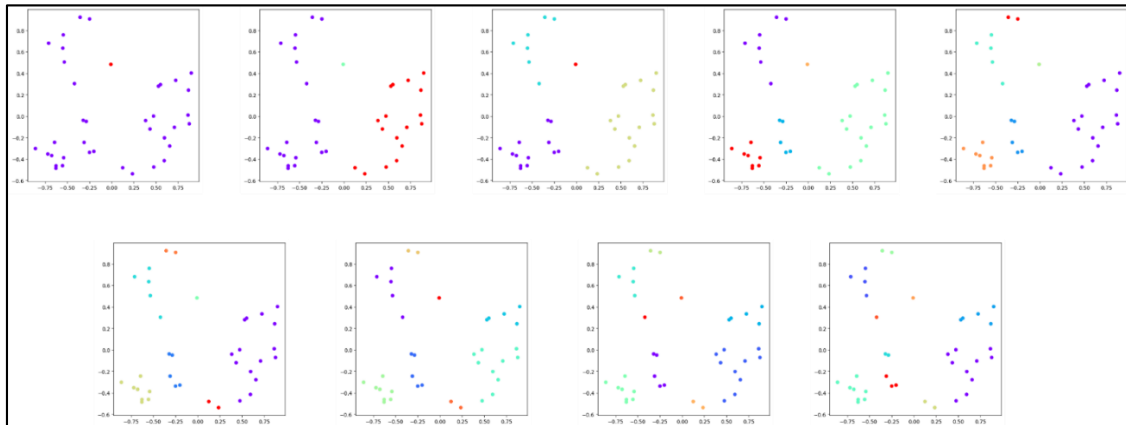
Pada tabel 4 terdapat 18 kolom yang merupakan gabungan dari 3 tabel. Nama atribut juga dirubah menggunakan alfabet, 3 atribut pertama adalah atribut dari tabel lama pemberian ASI, 5 atribut berikutnya yaitu berasal dari tabel imunisasi, dan 10 atribut terakhir adalah berasal dari tabel makanan/cairan yang diterima. Dataset inilah yang kemudian digunakan pada proses *clustering*. Rincian perubahan nama atribut adalah sebagai berikut, atribut A, B, C, secara berurutan menggantikan nama atribut 'Pemberian ASI', 'Tanpa Makanan Pendamping', 'Dengan Makanan Pendamping'. Atribut D, E, F, G, H, secara berurutan menggantikan 'BCG', 'DPT', 'Polio', 'Campak', 'Hepatitis B'. Atribut I sampai R secara berurutan menggantikan atribut 'Air Putih', 'Air Tajin/Madu/The/AirGula', 'Bubur/Nasi/Roti/Mie/Jagung', 'Kacang-kacangan', 'Susu selain ASI atau Keju', 'Daging/Hati/Jeroan/Ikan', 'Telur', 'Sayuran', 'Buah-buahan', 'Lainnya (Kue)'.

Proses *clustering* menggunakan perangkat lunak Jupyter yang berbasis python dengan utilisasi beberapa *library machine learning* dan visualisasi data. Jumlah *cluster* dilakukan uji coba mulai 2 *cluster* sampai dengan 10 *cluster*. Masing – masing *cluster* yang terbentuk kemudian dianalisis menggunakan *silhouette index*. Adapun tujuan untuk melakukan analisis menggunakan *silhouette index* adalah mengetahui *cluster* manakah yang memiliki nilai index paling tinggi sebagai parameter dalam menentukan *cluster* mana yang terbaik.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.cluster import AgglomerativeClustering
from sklearn.preprocessing import StandardScaler, normalize
from sklearn.metrics import silhouette_score
import scipy.cluster.hierarchy as shc
```

Sumber: Hasil Penelitian (2023)

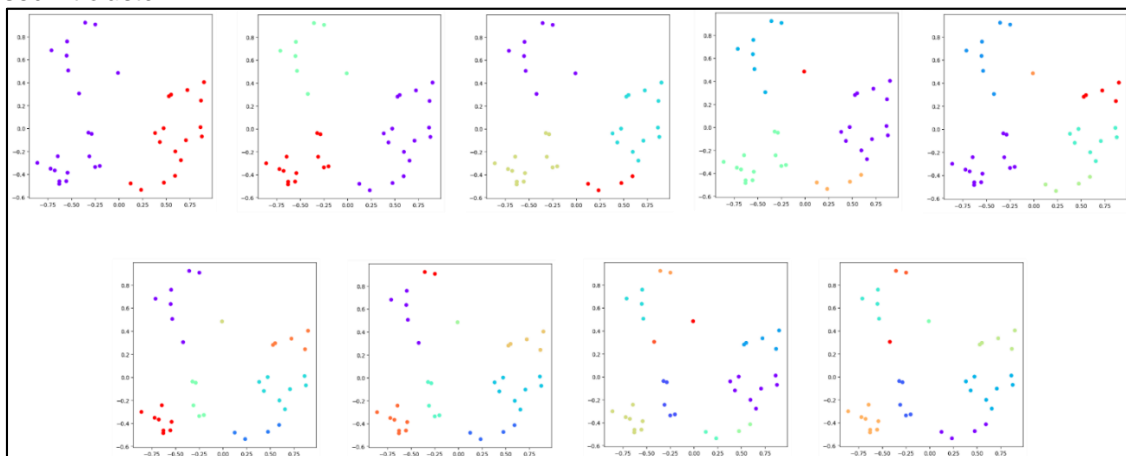
Gambar 3. *Library machine learning* yang digunakan



Sumber: Hasil Penelitian (2023)

Gambar 4. Visualisasi *Cluster Single Cluster*

Gambar 4 menunjukkan hasil *clustering* secara grafis berdasarkan distribusi masing – masing anggota dalam tiap *cluster*, uji dilakukan dengan membentuk 2 *cluster* sampai dengan 10 *cluster*. Pada *single linkage* cenderung menghasilkan *cluster* yang berbentuk memanjang, dimana anggota – anggota pada suatu *cluster* yang sama akan memiliki jarak berdekatan, hal ini disebabkan karena konsep *single linkage* adalah berdasarkan jarak terdekat diantara anggota. Pada beberapa model data, hal ini dapat menyebabkan kesulitan ketika menginginkan terbentuk sedikit *cluster*.

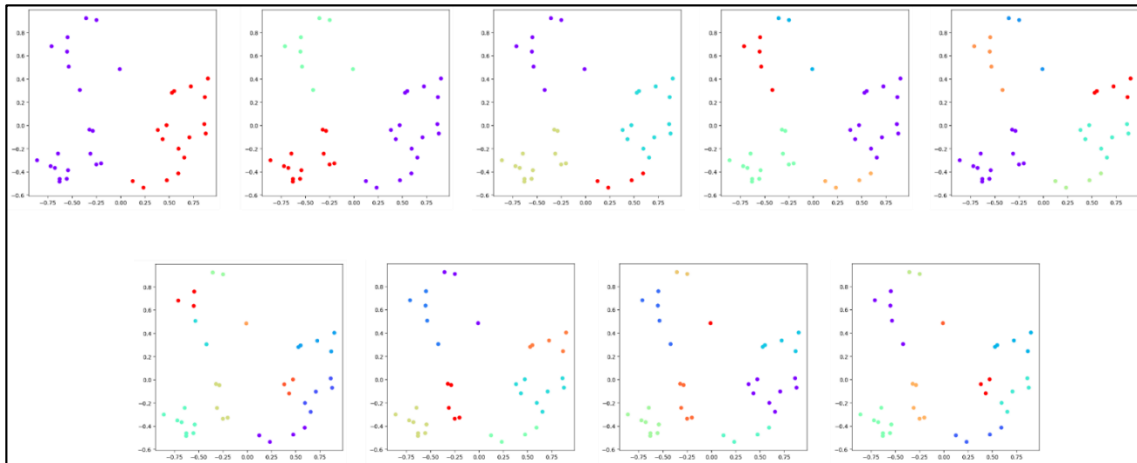


Sumber: Hasil Penelitian (2023)

Gambar 5. Visualisasi *Cluster Average linkage*

Gambar 5 menunjukkan hasil *cluster* yang terbentuk pada *Average linkage* yang dilakukan uji coba membentuk 2 *cluster* sampai dengan 10 *cluster*. Pada konsep *average linkage*, konsep jarak yang digunakan adalah jarak rata – rata antara setiap titik dalam satu *cluster* dengan setiap titik di *cluster* lainnya, kemudian dengan mengacu pada definisi konsep jarak antar *cluster* tersebut, pada setiap langkah proses tersebut maka kemudian dua *cluster* digabungkan berdasarkan jarak rata – rata yang terpendek.

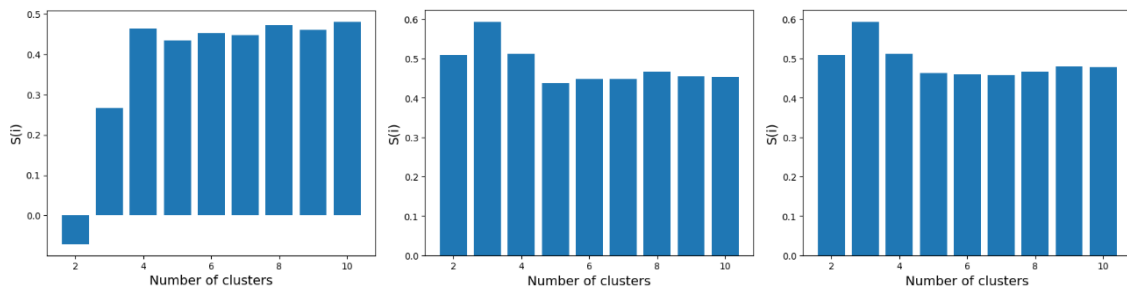
Gambar 6 menyajikan visualisasi sebaran *cluster* yang terbentuk ketika dilakukan uji 2 *kaslter* sampai dengan 10 *cluster* menggunakan *Complete linkage*. Pada *complete linkage*, konsep pengelompokkan menggunakan jarak terjauh, yang artinya bahwa anggota pada dua *cluster* adalah yang memiliki jarak terjauh, hal ini menjadi salah satu keunggulan *complete linkage* karena dapat mengelompokkan dengan baik ketika anggota – anggota di dalam data mempunyai noise, akan tetapi *complete linkage* mempunyai kelemahan pada pembentukan data karena cenderung membentuk banyak *cluster*.



Sumber: Hasil Penelitian (2023)

Gambar 6. Visualisasi Cluster Complete linkage

Hasil clustering masing – masing dari penerapan *single linkage*, *average linkage*, dan *complete linkage*, kemudian dilakukan analisis validasi cluster menggunakan *silhouette index*.



Sumber: Hasil Penelitian (2023)

Gambar 7. Grafik Hasil Silhouette index

Pada gambar 7 ditunjukkan hasil perolehan nilai *silhouette index*. Berdasarkan nilai *silhouette index* yang didapatkan, maka menunjukkan ketika menggunakan *single linkage*, maka cluster optimal terjadi ketika terbentuk 10 cluster, sedangkan ketika menggunakan *average linkage* dan *complete linkage* maka cluster optimal terjadi ketika terbentuk sama yaitu sejumlah 3 cluster.

Tabel 5. Hasil Cluster

KabKota	Cluster	KabKota	Cluster	KabKota	Cluster
Kotamadya Batu	1	Kabupaten Pacitan	2	Kabupaten Sumenep	3
Kabupaten Ponorogo	1	Kotamadya Pasuruan	2	Kabupaten Pamekasan	3
Kabupaten Trenggalek	1	Kabupaten Lamongan	2	Kabupaten Sampang	3
Kotamadya Surabaya	1	Kabupaten Tulungagung	2	Kabupaten Bangkalan	3
Kotamadya Madiun	1	Kabupaten Blitar	2	Kabupaten Pasuruan	3
Kotamadya Mojokerto	1	Kabupaten Kediri	2	Kabupaten Probolinggo	3
Kotamadya Probolinggo	1	Kabupaten Malang	2	Kabupaten Situbondo	3
Kotamadya Malang	1	Kabupaten Lumajang	2	Kabupaten Bondowoso	3
Kotamadya Blitar	1	Kabupaten Jember	2		
Kotamadya Kediri	1	Kabupaten Banyuwangi	2		
Kabupaten Gresik	1	Kabupaten Tuban	2		
Kabupaten Magetan	1	Kabupaten Bojonegoro	2		
Kabupaten Madiun	1	Kabupaten Ngawi	2		
Kabupaten Nganjuk	1				
Kabupaten Sidoarjo	1				
Kabupaten Mojokerto	1				
Kabupaten Jombang	1				

Sumber: Hasil Penelitian (2023)

Pada tabel 5 ditunjukkan bahwa hasil *cluster* yang diasumsikan optimal adalah 3 *cluster*. Hasil ini diambil berdasarkan nilai *silhouette index* yang diperoleh. *Cluster* 1 mempunyai 17 anggota, *cluster* 2 mempunyai 13 anggota, dan *cluster* 3 mempunyai 8 anggota.

Adapun analisis yang dapat diambil dari terbentuknya 3 *cluster* adalah, untuk Kabupaten/Kota yang menjadi anggota *cluster* 1 yang berwarna hijau, merupakan daerah yang mempunyai tingkat kepadatan penduduk tinggi, sedangkan Kabupaten/Kota yang menjadi anggota *cluster* 3 yang diwarnai biru, adalah daerah yang mempunyai tingkat kepadatan penduduk rendah, sedangkan anggota *cluster* 2 yang berwarna oranye, adalah daerah yang mempunyai tingkat kepadatan penduduk diantara anggota *cluster* 1 dan anggota *cluster* 3. Pada *cluster* 1 hampir seluruh kotamadya di propinsi Jawa Timur menjadi anggota, hanya 1 kotamadya saja yang menjadi anggota *cluster* 2. Pada *cluster* 3 terdapat hasil menarik yaitu bahwa seluruh kabupaten yang terletak di pulau madura menjadi anggota pada *cluster* ini, dan seluruh anggota pada *cluster* 3 merupakan kabupaten yang berada pada bagian timur propinsi Jawa Timur.

4. Kesimpulan

Hasil pengujian menunjukkan bahwa pada penggunaan *single linkage cluster* optimal terjadi ketika terbentuk 10 *cluster*, sedangkan pada penggunaan *average linkage* dan *complete linkage* menunjukkan bahwa *cluster* optimal terjadi ketika terbentuk 3 *cluster*.

Hasil analisis yang dapat diambil dari terbentuknya 3 *cluster* adalah, untuk Kabupaten/Kota yang menjadi anggota *cluster* 1 merupakan daerah yang mempunyai tingkat kepadatan penduduk tinggi, sedangkan Kabupaten/Kota yang menjadi anggota *cluster* 3 adalah daerah yang mempunyai tingkat kepadatan penduduk rendah, sedangkan anggota *cluster* 2 adalah daerah yang mempunyai tingkat kepadatan penduduk diantara anggota *cluster* 1 dan anggota *cluster* 3.

Berkaitan dengan data stunting, maka kabupaten kotamadya yang termasuk dalam *cluster* 1 memiliki tingkat stunting rendah, sedangkan *cluster* 2 memiliki tingkat stunting sedang, dan *cluster* 3 adalah kabupaten kotamadya yang termasuk ke dalam daerah yang memiliki tingkat stunting tinggi.

Saran terhadap penelitian ini adalah pada bagian dataset yang digunakan, bisa dikembangkan dengan menambahkan atribut lain terkait kasus stunting di Indonesia, misalkan atribut berat badan dan tinggi badan.

Ucapan Terima Kasih

Terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Muhammadiyah Jember yang telah memberikan dukungan berupa dana penelitian sehingga penelitian ini dapat berjalan dengan lancar sampai selesai. Terima kasih juga kami sampaikan kepada pihak yang telah membantu terlaksananya penelitian ini.

Referensi

- [1] A. J. Prendergast e J. H. Humphrey, "The stunting syndrome in developing countries," *Paediatr. Int. Child Health*, p. 250–265, 2014.
- [2] R. Rahutomo, I. Nurlaila, A. S. Perbangsa e B. Pardamean, "Database Management System Design with Time Series Modification for Child Growth and Malnutrition Monitoring in The Regency of Serdang Bedagai," em *2020 International Conference on Information Management and Technology (ICIMTech)*, Bandung, Indonesia, 2020.
- [3] "Joint child malnutrition estimates," 6 June 2019. [Online]. Available: <https://www.who.int/nutgrowthdb/estimates2018/en/>.
- [4] Z. Momand, P. Mongkolnam, P. Kositpanthavong e J. H. Chan, "Data Mining Based Prediction of Malnutrition in Afghan Children," em *2020 12th International Conference on Knowledge and Smart Technology (KST)*, Pattaya, Thailand, 2020.
- [5] F. Oktaviana, M. N. Widyawati, K. Kurnianingsih e N. Kubota, "Early Detection of the Risk of Stunting in Pregnant Women and Its Recommendations," em *2020 International Symposium on Community-centric Systems (CcS)*, Tokyo, Japan, 2020.
- [6] BAPPENAS, "Tujuan Pembangunan Berkelanjutan," 2022. [Online]. Available: <https://sdgs.bappenas.go.id/>.

- [7] J. Han, M. Kamber e J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann Publishers, 2011.
- [8] P. Tan, M. Steinbach e V. Kumar, *Cluster Analysis: Basic Concepts and Algorithms*. In: *Introduction to Data Mining*, Boston: Addison-Wesley, 2005.
- [9] F. Gorunescu, *Data Mining Concepts, Models and Techniques*, Heidelberg: Springer Science dan Business Media, 2011.
- [10] S. Winiarti, H. Yuliansyah e A. A. Purnama, "Identification of Toddlers' Nutritional Status using Data Mining Approach," *International Journal of Advanced Computer Science and Applications (IJACSA)*, pp. 164-169, 2018.
- [11] S. Wulandari e R. Kurniawan, "Pengelompokan Kabupaten/Kota Di Jawa Timur Berdasarkan Kasus Stunting Balita Menggunakan Algoritme Fuzzy Particle Swarm Optimization-Fuzzy C-Means," *Statistika, Vol. 7, No. 1*, 2019.
- [12] D. Jollyta, S. Efendi, M. Zarlis e H. Mawengkang, "Optimasi *Cluster* Pada Data Stunting Teknik Evaluasi *Cluster* Sum of Square Error dan Davies Bouldin Index," em *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2019.
- [13] BPS, "BPS Propinsi Jawa Timur," [Online]. Available: <https://jatim.bps.go.id/subject/30/kesehatan.html#subjekViewTab3>.
- [14] P. Praveen, R. Kumar, M. A. Shaik, R. Ravikumar e R. Kiran, "The Comparative Study On Agglomerative Hierarchical *Clustering* Using Numerical Data," em *ICRAEM 2020*.
- [15] A. A. Munshi, "*Clustering* of Wind Power Patterns Based on Partitional and Swarm Algorithms," *IEEE Access (Volume: 8)*, pp. 111913 - 111930, 2020.