

Perbandingan *Agglomerative Hierarchical* dan *K-Means* dalam Pengelompokan Provinsi Berdasarkan Pelayanan Kesehatan Maternal

Comparison of Agglomerative Hierarchical and K-Means in Grouping Provinces Based on Maternal Health Services

¹Alya Azzahra*, ²Arie Wahyu Wijayanto
Politeknik Statistika STIS
Jl. Otto Iskandardinata No.64C Jakarta 13330
*e-mail: 211810158@stis.ac.id

(*received*: 26 Desember 2021, *revised*: 26 Desember 2021, *accepted*: 26 Februari 2022)

Abstrak

Pada masa Covid-19 ini ditemukan adanya hambatan akses ibu hamil ke pelayanan kesehatan yang bisa mengganggu kesehatan maternal. Maka dari itu, perlunya diketahui capaian cakupan pelayanan kesehatan maternal di Indonesia pada masa Covid-19 di tahun 2020 ini utamanya pada level provinsi sehingga mampu membantu pemerintah untuk menentukan prioritas daerah untuk pemenuhan pelayanan kesehatan maternal yang lebih memadai. Penentuan prioritas provinsi untuk pemenuhan pelayanan kesehatan maternal dapat dicapai dengan melakukan pengelompokan daerah sesuai dengan karakteristik pelayanan kesehatan maternal di provinsi setempat. Analisis kluster mampu mengelompokkan objek berupa provinsi ke dalam satu kluster. Metode *clustering* yang akan digunakan adalah *agglomerative hierarchical clustering* dan *k-means clustering*. Hasil *clustering* kedua metode tersebut kemudian akan dibandingkan dengan validasi internal berupa indeks *dunn*, indeks *connectivity*, dan indeks *silhouette*. Hasil *clustering* terbaik diperoleh dengan menggunakan algoritma *agglomerative hierarchical clustering* dengan menggunakan fungsi kemiripan *average linkage* dengan jumlah *cluster* yang dihasilkan sebanyak lima *cluster*. Hasil identifikasi karakteristik kluster memberikan hasil bahwa kluster 1 dengan anggota 14 provinsi dikategorikan dengan provinsi dengan cakupan pelayanan kesehatan maternal yang baik. Kluster 2 yang beranggotakan 15 provinsi dikategorikan sangat baik. Kluster 3 dengan anggota NTT dan Maluku dikategorikan buruk. Kluster 4 yang hanya terisi oleh Kalimantan Timur sudah dikategorikan cukup. Sementara itu, kluster 5 dengan anggota Papua dan Papua Barat masih memprihatinkan dengan status sangat buruk.

Kata kunci: pengelompokan, *agglomerative hierarchical*, *k-means*, kesehatan maternal

Abstract

During the Covid-19 period, there were barriers to access for pregnant women to health services that could interfere with maternal health. Therefore, it is necessary to know the achievement of maternal health service coverage in Indonesia during the Covid-19 period in 2020, especially at the provincial level so that it can help the government to determine regional priorities for the fulfillment of more adequate maternal health services. Determination of provincial priorities for the fulfillment of maternal health services can be achieved by grouping the regions according to the characteristics of maternal health services in the local province. Cluster analysis is able to group objects in the form of provinces into one cluster. The clustering methods that will be used are *agglomerative hierarchical clustering* and *k-means clustering*. The results of the clustering of the two methods will be compared with internal validation in the form of *dunn index*, *connectivity index*, and *silhouette index*. The best clustering results are obtained by using *agglomerative hierarchical clustering* algorithm using the complete linkage similarity function with the resulting five clusters. The results of the identification of cluster

characteristics show that cluster 1 with 14 members is categorized as provinces with good coverage of maternal services. Cluster 2 which consists of 15 provinces is categorized as best coverage. Cluster 3 which member are NTT and Maluku is categorized as bad. Cluster 4 which member is East Kalimantan is categorized as sufficient coverage. Meanwhile cluster 5 which member are Papua and West Papua is still on concern because its categorized as worst coverage.

Keywords: *clustering, agglomerative hierarchical, k-means, maternal health*

1 Pendahuluan

Kesehatan maternal mencakup segala bentuk upaya kesehatan bagi ibu meliputi pencegahan kematian akibat kehamilan serta peningkatan kualitas hidup selama rentang waktu reproduksi [1]. Pelayanan kesehatan maternal harus diupayakan agar setiap ibu hamil mampu memperoleh keselamatan kehamilan maupun persalinannya. Isu kesehatan maternal di dunia sudah lama disoroti oleh lembaga kesehatan dunia atau *World Health Organization* (WHO) melalui inisiasi pembahasan di beberapa konferensi seperti *World Health Conference on Woman* di Beijing tahun 1995 dan *Safe Motherhood Technical Consultation* di Srilanka tahun 1997. Kedua konferensi tersebut sepakat untuk menurunkan Angka Kematian Ibu (AKI) di dunia menjadi 50% pada tahun 2000. Menindaklanjuti konferensi tersebut, pada tahun 1999 badan internasional seperti UNFPA, UNICEF, dan Bank Dunia bersama dengan WHO sepakat untuk bersama-sama menurunkan AKI di dunia dan kemudian menerbitkan program *Making Pregnancy Safe (MPS)*. Program MPS menerbitkan tiga poin penting yang harus dipatuhi oleh pemerintah di setiap negara di dunia. Pertama, menjadikan *safe motherhood* menjadi prioritas teratas dalam pembangunan. Kedua, menyusun dasar dan standar pelayanan kesehatan maternal dan neonatal. Ketiga, melakukan penjaminan standar yang sudah disusun melalui pengembangan sistem [2].

Data termutakhir dari survei kesehatan dunia yang diselenggarakan oleh WHO yaitu *Demographic and Health Survey* (DHS) tahun 2017 menunjukkan bahwa AKI Indonesia berada pada urutan ketiga teratas di Asia Tenggara dengan perolehan AKI sebesar 177 atau terdapat 177 kematian ibu per 100.000 kelahiran hidup [3]. Kondisi ini mengindikasikan bahwa AKI di Indonesia masih tergolong sangat tinggi jika dibandingkan dengan negara lain di Asia Tenggara. Capaian AKI Indonesia juga masih sangat jauh dengan target 3.1 Tujuan Pembangunan Berkelanjutan (TPB) yang bertujuan mengurangi AKI hingga 70 per 100.000 kelahiran hidup pada tahun 2030 [4].

Medis merupakan penyebab langsung dari meningkatnya AKI yang bisa dicegah melalui manajemen pelayanan kesehatan yang memadai [5]. Namun, semenjak mewabahnya pandemi Covid-19 ; negara-negara menghadapi tantangan besar untuk mempertahankan penyediaan layanan kesehatan ibu serta bayi baru lahir yang memadai. Tantangan ini muncul akibat ibu hamil dan ibu dengan bayi baru lahir mengalami kesulitan mengakses layanan kesehatan karena gangguan transportasi, *lockdown*, ataupun enggan mengunjungi fasilitas kesehatan karena takut tertular wabah [4]. Pernyataan tersebut didukung dengan hasil penelitian di Indonesia yang menyatakan bahwa pada masa Covid-19 ini ditemukan adanya hambatan akses ibu hamil ke pelayanan kesehatan yang bisa mengganggu kesehatan maternal [7].

Maka dari itu, perlunya diketahui capaian cakupan pelayanan kesehatan maternal di Indonesia pada masa Covid-19 di tahun 2020. Penentuan prioritas provinsi untuk pemenuhan pelayanan kesehatan maternal dapat dicapai dengan melakukan pengelompokan daerah sesuai dengan karakteristik pelayanan kesehatan maternal di provinsi setempat. Analisis kluster mampu mengelompokkan objek berupa provinsi ke dalam satu kluster dan nantinya provinsi berkarakteristik beda ke kluster yang lain melalui cara memaksimalkan kesamaan antarobjek pada satu kluster serta meminimalkan kemiripan antarkluster.

Terdapat berbagai pilihan algoritma yang bisa digunakan untuk melakukan *clustering*. Secara garis besar, *clustering* dibedakan menjadi *hierarchical clustering* dan *partitional clustering*. Penelitian ini akan membandingkan performa kedua metode *clustering* tersebut yaitu antara *hierarchical* dan *partitional k-means clustering*. Validasi internal melalui indeks *dunn*, indeks *silhouette*, dan indeks *connectivity* akan dilakukan untuk menentukan kinerja algoritma terbaik dan algoritma yang terpilih nanti akan menjadi dasar klusterisasi.

Tujuan diselenggarakan penelitian ini adalah untuk menemukan algoritma *clustering* terbaik diantara *agglomerative hierarchical* dan *partitional k-means* yang akan ditentukan melalui nilai indikator pada validasi internal. Kemudian algoritma terbaik tersebut akan menjadi dasar dalam mengelompokkan provinsi di Indonesia berdasarkan kesamaan karakteristik cakupan pelayanan kesehatan maternal. Dengan terbentuknya kelompok tersebut akan bermanfaat dalam membantu memudahkan pemerintah untuk membentuk dan menyalurkan kebijakan serta fasilitasi layanan kesehatan maternal dengan tepat sasaran utamanya di era adaptasi kebiasaan baru ini yang membutuhkan penyesuaian kondisi di lapangan.

2 Tinjauan Literatur

Penelitian terkait *clustering* cakupan pelayanan kesehatan maternal ini pernah dikaji sebelumnya. Salah satunya adalah penelitian *clustering* berdasarkan tujuh indikator pelayanan kesehatan ibu mencakup pelayanan K1, pelayanan K4, kunjungan ibu nifas 3 kali, ibu hamil yang mendapat 90 tablet tambah darah, ibu bersalin dengan pertolongan tenaga kesehatan, ibu hamil dengan imunisasi toksoid, dan ibu bersalin dengan pertolongan tenaga kesehatan di fasilitas kesehatan. Penelitian ini bertujuan membandingkan metode *hierarchical*, *ensemble*, dan *k-means*. Performa klusterisasi dikaji melalui nilai *davis bouldin* dan *compactness* kemudian didapatkan bahwa algoritma *ensemble* memiliki performa terbaik dalam klusterisasi kasus tersebut [8].

Selain itu, *k-means clustering* juga pernah digunakan untuk melakukan pemetaan berbasis sistem informasi geografis pada faktor-faktor yang berpengaruh terhadap Angka Kematian Ibu (AKI) dan Angka Kematian Bayi (AKB) di Provinsi Bengkulu. Variabel dari faktor ibu yang diteliti meliputi pelayanan K1, K4, pemberian tablet Fe, persalinan ditolong tenaga kesehatan, persalinan tanpa ditolong tenaga kesehatan, komplikasi kebidanan, PHBS, dan pelayanan nifas. Sementara itu, variabel dari faktor anak meliputi bayi dengan berat badan lahir rendah, tetanus neonatorum, kelainan bawaan, sepsis, diare, pneumonia, kelainan pada saraf, kelainan pada saluran cerna, dan malaria. Algoritma *k-means* dalam penelitian ini menghasilkan tiga *cluster* yang kemudian dipetakan melalui sistem informasi geografis [9].

Penelitian lain melalui metode *hierarchical clustering* dengan *average linkage* juga digunakan untuk mengelompokkan daerah di Jawa Timur berdasarkan program layanan kesehatan ibu hamil. Penelitian ini hanya menggunakan empat variabel mencakup pertolongan persalinan, penanganan komplikasi melalui tenaga bidan, keluarga berencana, dan pelayanan K4. Penentuan jumlah *cluster* optimumnya ditentukan melalui nilai *pseudo-f* yang kemudian didapatkan sejumlah 2 *cluster* yang optimum [10].

Sementara itu, penelitian ini akan mengarah pada perbandingan dua algoritma yaitu *agglomerative hierarchical* dan *k-means*. Penelitian terdahulu terkait kajian perbandingan algoritma *clustering* yang serupa juga pernah dikaji. Pertama, penelitian terkait komparasi antara *soft* dan *hard clustering* pada tingkat kesejahteraan kota/kabupaten di Pulau Jawa [11]. Selain itu, penelitian terkait pengelompokkan kota di Pulau Jawa berdasarkan IPM 2018 juga pernah membandingkan performa *hierarcichal* dan *non-hierarcichal clustering* [12]. Studi yang sama terkait IPM 2019 juga pernah bertujuan membandingkan algoritma *partitioning* dan *hierarchical* [13]. Penelitian lain terkait perbandingan *partitioning* dan *hierarcichal clustering* juga diterapkan pada pengelompokkan provinsi berdasarkan informasi kemiskinan [14]. Bahkan algoritma *hierarchical* dan *k-means* juga pernah dibandingkan dengan *k-medoids* untuk kasus pengelompokkan kota/kabupaten menurut indikator IPM [15].

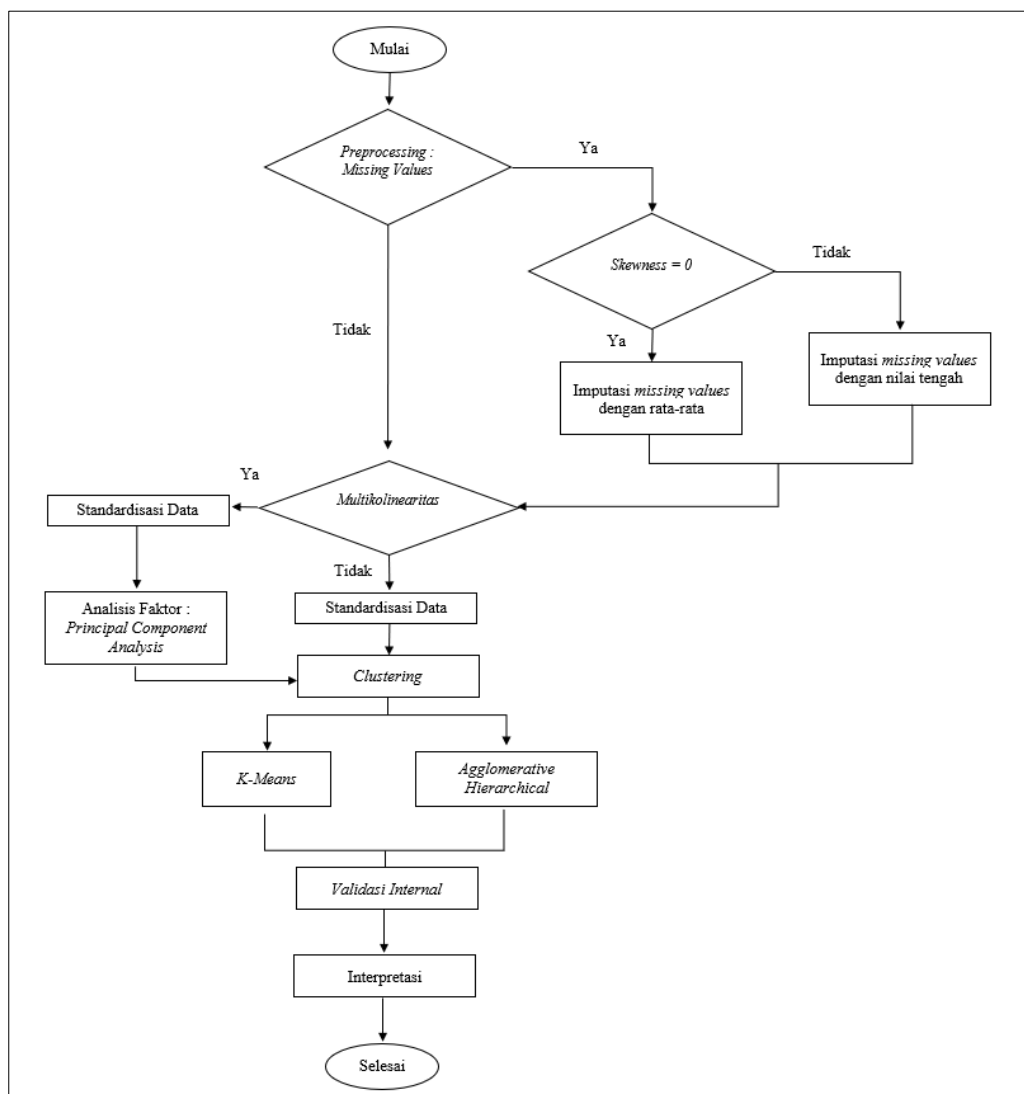
Berdasarkan hasil uraian tersebut, penelitian ini akan menerapkan pengelompokkan provinsi di Indonesia berdasarkan indikator cakupan pelayanan kesehatan maternal dengan menggunakan algoritma *agglomerative hierarchical clustering* dan *k-means*. Kedua algoritma tersebut akan dibandingkan performanya dengan menggunakan cara validasi internal untuk menemukan algoritma penghasil *cluster* terbaik. *Cluster* yang terbentuk dengan masing-masing karakteristiknya diharapkan mampu menjadi dasar bagi pemangku kebijakan agar tepat sasaran dalam implementasi kebijakan.

Meskipun penelitian serupa pernah dilakukan, namun penelitian ini akan berbeda karena akan menggunakan indikator kesehatan maternal yang lebih holistik jika dibandingkan penelitian sebelumnya yaitu akan digunakan indikator yang belum dicakup penelitian lain seperti persentase cakupan imunisasi td2+ pada ibu hamil, persentase puskesmas yang melaksanakan kelas ibu hamil, persentase ibu hamil

yang melaksanakan deteksi dini hepatitis B, persentase ibu nifas yang mendapatkan vitamin A, dan persentase ibu hamil yang diperiksa HIV. Selain itu, pembeda penelitian ini dengan penelitian lain juga dapat ditinjau dari aspek waktu penelitian yang lebih mutakhir. Penelitian ini menggunakan referensi waktu tahun 2020 saat situasi pandemi Covid-19. Pemilihan waktu di tahun 2020 dilakukan untuk melihat apakah krisis pandemi dapat memengaruhi cakupan pelayanan kesehatan maternal di Indonesia. Pembeda lainnya adalah penelitian ini hanya terbatas pada perbandingan kinerja algoritma *agglomerative hierarchical clustering* dan *partitional k-means*.

3 Metode Penelitian

Penelitian ini menerapkan metode kuantitatif dengan mengukur perbandingan dua algoritma *clustering* yaitu *agglomerative hierarchical clustering* dengan menggunakan pengukuran jarak meliputi *single linkage*, *average linkage*, dan *ward method* yang akan dibandingkan dengan *partitional clustering* melalui pendekatan *k-means*.



Gambar 1. Metode Penelitian

Berdasarkan Gambar 1, kerangka kerja penelitian ini meliputi tahapan berikut.

1) *Preprocessing*

Pada tahap ini dilakukan pendeteksian *missing values* pada dataset dengan rincian variabel yang berkaitan dengan pelayanan kesehatan maternal sebagaimana tertera pada Tabel 1. Penelitian ini menggunakan sumber data sekunder dari 34 provinsi di Indonesia yang diperoleh dari Publikasi Profil Kesehatan Indonesia Tahun 2020 oleh Kementerian Kesehatan Indonesia[16].

Tabel 1. Daftar Variabel

Nama Variabel	Keterangan Variabel	Tipe Data
Imunisasi Td2Plus	Persentase Cakupan Imunisasi Td2+ Pada Ibu Hamil	Numerik
TTD	Persentase Ibu Hamil Yang Mendapatkan Tablet Tambah Darah (TTD)	Numerik
KelasIbuHamil	Persentase Puskesmas Yang Melaksanakan Kelas Ibu Hamil	Numerik
DDHB	Persentase Ibu Hamil Yang Melaksanakan Deteksi Dini Hepatitis B	Numerik
K1	Persentase Ibu Hamil Yang Mendapatkan Pelayanan Antenatal Sesuai Standar Pertama Kali (Pada Trisemester Pertama)	Numerik
K4	Persentase Ibu Hamil Yang Mendapatkan Pelayanan Antenatal Sesuai Standar Yang Keempat Kali (Pada Trisemester Keempat)	Numerik
PersalinanNakes	Persentase Persalinan Ditolong Nakes	Numerik
PersalinanFasyankes	Persentase Persalinan di Fasilitas Pelayanan Kesehatan	Numerik
PersalinanKF1	Persentase Kunjungan Nifas Pertama	Numerik
PersalinanKFlengkap	Persentase Kunjungan Nifas Lengkap	Numerik
NifasVitA	Persentase Ibu Nifas Mendapat Vitamin A	Numerik
PeriksaHIV	Persentase Ibu Hamil diperiksa HIV	Numerik
KN1	Persentase Kunjungan Neonatal 1 kali	Numerik
KNLengkap	Persentase Kunjungan Neonatal Lengkap	Numerik
PuskesmasBidan	Persentase Puskesmas dengan Kecukupan Bidan	Numerik

Kemudian variabel yang terdeteksi memiliki *missing values* akan dideteksi *skewness-nya* guna mengetahui metode imputasi nilai yang terbaik. Apabila nilai *skewness* berada pada angka sekitar nol maka digunakan imputasi dengan rata-rata. Sedangkan nilai *skewness* yang menceng atau berada jauh dari nilai nol akan menggunakan imputasi *median*.

2) *Identifikasi Multikolinearitas*

Tahap ini diawali dengan pengujian korelasi yang berguna untuk mengukur adanya keterkaitan antarvariabel. Apabila signifikan berkorelasi akan terdapat dugaan multikolinearitas, sehingga akan dilanjutkan ke analisis faktor.

3) *Analisis Faktor dengan Principal Component Analysis (PCA)*

Sebelum melakukan analisis faktor akan diuji asumsi *Kaiser-Mayer-Olkin (KMO)* untuk mengukur kecukupan sampel dan uji *bartlett* untuk mengukur kelayakan variabel dalam analisis faktor. Apabila semua asumsi terpenuhi, maka reduksi variabel pun bisa dilakukan dengan menggunakan analisis faktor melalui pendekatan PCA.

4) *Modelling*

Pada tahap ini dilakukan *clustering* dengan algoritma *agglomerative hierarchical clustering* dan *k-means clustering*.

5) *Validation*

Pada tahap ini dilakukan validasi secara internal dengan menggunakan indeks *dunn*, indeks *silhouette*, dan indeks *connectivity*. Indeks-indeks tersebut kemudian akan dibandingkan satu sama lain untuk mendapatkan tipe *clustering* terbaik.

Analisis Faktor Dengan *Principal Component Analysis (PCA)*

Setelah mengetahui keberadaan korelasi antarvariabel maka akan muncul dugaan multikolinearitas. Kehadiran multikolinearitas antarvariabel akan memengaruhi ketepatan *clustering*. Salah satu metode untuk memperbaiki fenomena multikolinearitas adalah dengan menggunakan teknik multivariat berupa analisis faktor dengan metode *principal components analysis (PCA)* [17]. PCA bertujuan untuk mereduksi dimensi dalam rangka menyederhanakan variabel. Beberapa faktor atau komponen hasil PCA inilah yang akan menjadi variabel baru yang siap dianalisis menggunakan berbagai algoritma *clustering* [18].

Agglomerative Hierarchical Clustering

Agglomerative merupakan salah satu jenis metode *hierarchical clustering* dengan metode pengelompokan yang dimulai dari individu atau dikenal dengan metode *bottom-up*. Urutan proses *agglomerative hierarchical clustering* meliputi tahapan berikut [19].

- 1) Pada *cluster* di dalam N *cluster* terdapat satuan tunggal dan juga matriks $N \times N$ yang simetris dari jarak (dikenal dengan kemiripan).
- 2) Matriks jarak untuk tiap pasangan *cluster* paling dekat dicari. Contoh: jarak antar *cluster* U dan W disebut d_{UW} .
- 3) *Cluster* U dan W akan digabung dan dibentuk *cluster* baru yaitu UW. Sementara itu, entri di matriks jarak juga dimutakhirkan melalui cara berikut.
 - a. Menghapus kolom dan baris yang bersesuaian dengan *cluster* U dan W.
 - b. Menambahkan kolom dan baris yang menandai jarak antar *cluster* UW dengan *cluster* yang tersisa
- 4) Mengulangi poin nomor 2 dan 3 sebanyak $N-1$ kali dan semua objek nantinya akan memiliki tempat di *cluster* tunggal setelah algoritma berakhir.
- 5) Kemudian identitas *cluster* yang digabung beserta jarak kemiripannya diidentifikasi dimana gabungan tersebut ditempatkan.

Algoritma pada *agglomerative hierarchical clustering* sangat beragam diantaranya sebagai berikut [20].

1) *Single Linkage*

Algoritma ini mengelompokkan dengan menggunakan jarak terdekat antarobjek. Berikut adalah rumus untuk menentukan jarak antara UV ke *cluster* W.

$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \quad (1)$$

2) *Complete Linkage*

Algoritma ini mengelompokkan dengan menggunakan jarak terjauh antarobjek. Berikut adalah rumus untuk menentukan jarak antara UV ke *cluster* W.

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \quad (2)$$

3) *Average Linkage*

Algoritma ini mengelompokkan dengan menggunakan rata-rata jarak antarobjek. Berikut adalah rumus untuk menentukan jarak antara UV ke *cluster* W.

$$d_{(UV)W} = \frac{d_{(UW)} + d_{(VW)}}{n_{(UV)}n_W} \quad (3)$$

4) Metode Ward

Algoritma ini mengelompokkan dengan menggunakan rata-rata jarak antarobjek. Berikut adalah rumus untuk menentukan jarak antara UV ke cluster W.

$$d_{(UV)W} = \frac{[(n_W+n_U)d_{(UW)}+(n_W+n_V)d_{(VW)}]-n_W d_{(UV)}}{n_W+n_{(UV)}} \quad (4)$$

K-Means Clustering

Algoritma clustering dengan *k-means* akan mengelompokkan data menurut titik pusat cluster yang paling dekat dengan objek. Algoritma ini akan memaksimalkan kemiripan data melalui fungsi jarak pada satu cluster dan akan meminimalkan kemiripan antar cluster. Fungsi jarak yang digunakan salah satunya adalah jarak *euclidean* pada persamaan 6. Urutan proses algoritma *k-means* meliputi tahapan berikut.

- 1) Menentukan jumlah cluster atau k. Penentuan nilai k bisa menggunakan *elbow method* maupun *silhouette method*.
- 2) Penentuan titik pusat cluster (centroid) secara acak dari sebanyak k objek yang ada. Penentuan centroid selanjutnya menggunakan rumus pada persamaan 5.

$$v = \frac{\sum_{i=1}^n x_i}{n}, i = 1,2,3,\dots n \quad (5)$$

- 3) Menghitung jarak dari tiap objek ke centroid tiap cluster menggunakan fungsi jarak salah satunya adalah *euclidean* dengan rumus pada persamaan 6.

$$d_{(x_i, y_i)} = \sqrt{\sum_{i=1}^n (x_i, y_i)^2}, i = 1,2,3,\dots n \quad (6)$$

- 4) Menempatkan objek-objek pada centroid terdekat.
- 5) Melakukan iterasi serta menentukan posisi centroid yang baru.
- 6) Mengulangi poin 3 apabila posisi dari centroid yang baru berbeda.

Validasi Cluster

Koefisien Korelasi Cophenetic

Koefisien ini menunjukkan korelasi dari unit-unit asli dari matriks ketidakmiripan (*dissimilarity distance*) dengan matriks *cophenetic* yang diperoleh dari dendrogram. Nilai koefisien semakin mendekati satu menyatakan semakin baiknya proses clustering tersebut [21]. Rumus korelasi *cophenetic* ditunjukkan pada persamaan 7.

$$r_{coph} = \frac{\sum_{i<j}^n (d_{ij}-\bar{d})(d_{coph-ij}-\bar{d}_{coph})}{\sqrt{[\sum_{i<j}^n (d_{ij}-\bar{d})^2][\sum_{i<j}^n (d_{coph-ij}-\bar{d}_{coph})^2]}} \quad (7)$$

Indeks Silhouette

Optimalisasi dari suatu cluster dapat ditinjau melalui indeks ini dengan mengidentifikasi ketepatan data yang dikelompokkan. Rumus perhitungan indeks *silhouette* terdapat pada persamaan 8.

$$S_i = \frac{(b_i-a_i)}{\max(b_i-a_i)} \quad (8)$$

Dengan keterangan a_i merupakan rata-rata jarak pada objek ke-i dengan keseluruhan objek di kelompok yang sama. Sementara itu, b_i adalah rata-rata jarak pada objek ke-i dengan keseluruhan objek yang berbeda dan S_i adalah indeks *silhouette*. Hasil clustering akan semakin tepat apabila nilainya semakin mendekati satu[22].

Indeks Dunn

Indeks ini berfungsi mengkaji validitas cluster dengan dasar bahwa cluster yang terpisah umumnya mempunyai jarak antarcluster yang besar serta diameter dalam cluster yang kecil. Semakin besar nilai indeks *dunn*, maka hasil clusteringnya semakin baik. Rumus indeks *dunn* terdapat pada persamaan 9.

$$D = \min_{j = i + 1..n_c} \left(\min_{j = i + 1..n_c} \left(\frac{d(c_i, c_j)}{\max_{k=1..n_c} (diam(c_k))} \right) \right) \quad (9)$$

Dengan keterangan $d(c_i, c_j)$ adalah $\min_{x \in c_i, y \in c_j} (d(x, y))$ dan $(diam(c_k))$ adalah $\max_{x \in c_i, y \in c_j} (d(x, y))$.

Indeks Connectivity

Indeks *connectivity* digunakan untuk mengukur kepadatan untuk mengidentifikasi homogenitas *cluster* yang terlihat dari nilai varians intra-*cluster* atau dikenal dengan tetangga terdekat (*nearest neighbor*). Nilai indeks *connectivity* rendah akan mengindikasikan semakin baiknya proses *clustering*. Indeks *connectivity* dirumuskan pada persamaan 10.

$$CC = \sum_{i=1}^N \sum_{j=1}^L x_{i, nni(j)} \quad (10)$$

Dengan keterangan $x_{i, nni(j)}$ adalah nilai yang akan semakin mendekati nol ketika objek i dan j berada pada satu *cluster*, dengan $nni(j)$ adalah tetangga paling dekat dengan objek j dari objek di i dan L adalah parameter banyaknya jumlah tetangga.

4. Hasil dan Pembahasan

Preprocessing

Missing Values

Imputasi *missing values* dilakukan sebagai tahap awal atau *preprocessing* pada dataset agar dataset siap digunakan untuk *clustering*. Hasil identifikasi *missing values* pada dataset penelitian ini tertera pada Tabel 2.

Tabel 2. Daftar Variabel Yang Terdapat Missing Values

Nama Variabel	Jumlah Missing Values	Skewness	Metode Imputasi
Imunisasi Td2Plus	1	-0.167676	Mean
PersalinanNakes	1	-1.083231	Median
PersalinanKF1	1	-1.44156	Median
PersalinanKFlengkap	1	-1.055273	Median
NifasVitA	1	-1.493058	Median
KN1	2	-1.650599	Median
Total Missing Values	7		

Berdasarkan tabel 2 teridentifikasi terdapat total 7 *missing values*. Kemudian nilai *skewness* pada tiap variabel dicek untuk menentukan metode imputasi *missing value* yang sesuai. Variabel dengan kemencengan atau nilai *skewness*-nya jauh dari nol akan diimputasi menggunakan nilai tengah (*median*). Sementara itu, variabel yang tidak menceng atau nilai *skewness*-nya nol atau mendekati nol akan diimputasi dengan menggunakan metode rata-rata (*mean*).

Analisis Faktor

Uji Kaiser Meyer Olkin Measure of Sampling (KMO) dan Bartlett's Test of Sphericity

Uji ini dilakukan setelah teridentifikasi adanya korelasi antarvariabel yang signifikan. Kehadiran korelasi ini mengindikasikan adanya dugaan multikolinieritas. Maka, akan dilakukan uji asumsi untuk analisis faktor yaitu uji *Kaiser Meyer Olkin Measure of Sampling* (KMO) dan *Bartlett's Test of Sphericity*. Berdasarkan Tabel 3, diperoleh keputusan tolak H0 dari uji Bartlett dan nilai KMO yang lebih dari 0.5 menyimpulkan bahwa data layak untuk dilakukan analisis faktor dengan PCA.

Tabel 3. Uji Bartlett's Test of Sphericity

	KMO and Bartlett's Test	Nilai
Bartlett's Test of Sphericity	Approx Chi-Square	615.084
	Df.	105
	Sig.	p – value < 2.22×10^{-16}
Kaise-Meyer-Olkin Measure of Sampling Adequacy		0.7896634

Sementara itu, nilai MSA pada tiap variabel ditunjukkan pada tabel 4 di kolom MSA (1) dan terlihat bahwa masih terdapat variabel ImunisasiTd2Plus dan PuskesmasBidan yang nilai MSAny < 0.5. Maka, variabel ImunisasiTd2Plus harus dikeluarkan dari model, kemudian dilakukan pengukuran MSA ulang yang hasilnya ditunjukkan pada Tabel 4 di kolom MSA (2). Dengan demikian, hasilnya menunjukkan bahwa semua variabel kecuali variabel ImunisasiTd2Plus akan digunakan pada analisis faktor dengan PCA.

Tabel 4 . Nilai MSA

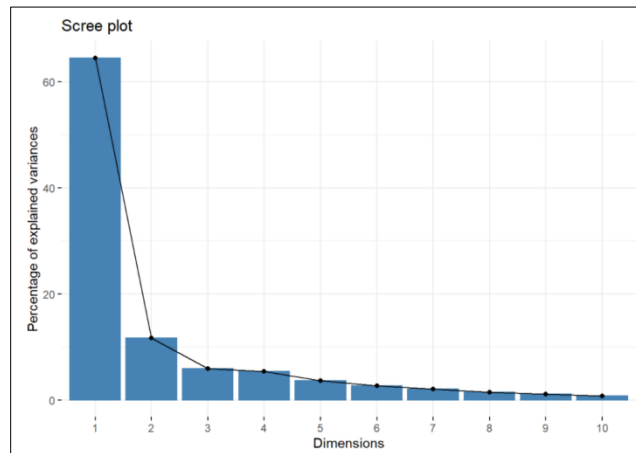
Variabel	MSA (1)	MSA (2)	Variabel	MSA(1)	MSA (2)
Imunisasi Td2Plus	0.4754	-	PersalinanKF1	0.7486	0.7976
TTD	0.8793	0.8908	PersalinanKFlengkap	0.8093	0.8710
KelasIbuHamil	0.8737	0.8629	NifasVitA	0.8320	0.8776
DDHB	0.8641	0.8566	PeriksaHIV	0.7275	0.7963
K1	0.6193	0.6158	KN1	0.9584	0.9621
K4	0.8291	0.8163	KNLengkap	0.8084	0.8046
PersalinanNakes	0.7482	0.7889	PuskesmasBidan	0.4840	0.7030
PersalinanFasyankes	0.7888	0.8031			

Principal Component Analysis (PCA)

Proses analisis faktor dilakukan dengan terlebih dahulu melakukan standarisasi pada dataset karena satuan tiap variabel masih beragam. Pemilihan jumlah komponen utama dilakukan dengan melihat nilai eigen >1 sebagaimana ditunjukkan pada Tabel 5 terdapat dua komponen utama yang bernilai eigen > 1 yang sudah mampu menjelaskan sebesar 76.17% keragaman pada data. Pemilihan dua komponen utama ini juga berdasarkan *scree plot* seperti pada Gambar 2. Maka, dua komponen utama inilah yang akan dijadikan dataset baru untuk selanjutnya diolah dengan metode *clustering*.

Tabel 5. Nilai Eigen

Faktor	Nilai Eigen	Kumulatif Varians
1	9.0324	64.52 %
2	1.6312	76.17 %



Gambar 2. Screen Plot

Modelling

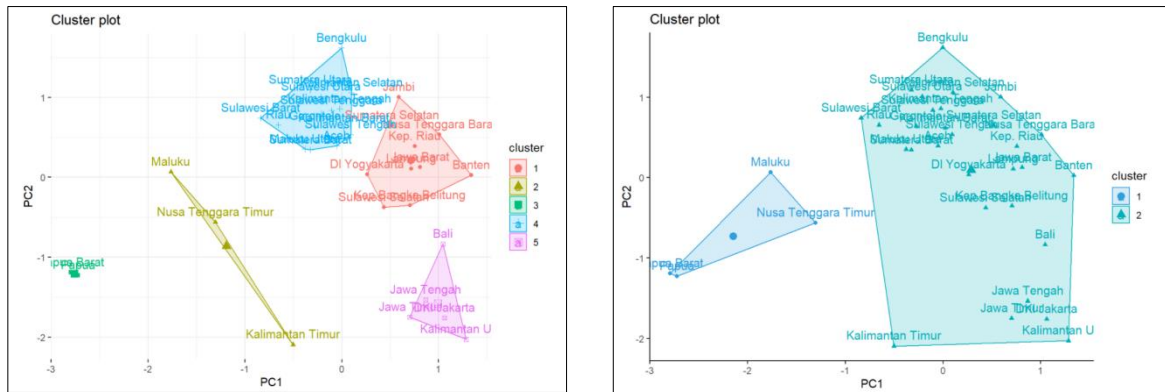
K-Means Clustering

Tahap *clustering* dengan pendekatan *k-means* diawali dengan penentuan nilai *k* yang paling optimal. Metode yang bisa digunakan untuk menentukan nilai *k* adalah metode *elbow* dan *silhouette*. Indikator yang bisa digunakan untuk melihat seberapa baik *cluster k-means* yang dihasilkan adalah dengan melihat nilai *Within Sum of Square (WSS)*. WSS merupakan perbandingan antara *between sum of squared* dan *total sum of squared* yang merepresentasikan seberapa berkumpulnya data pada tiap *centroid*. Nilai WSS yang semakin mendekati 100% mengindikasikan semakin baiknya *cluster* yang terbentuk. Hasil pembentukan *k-means clustering* dengan dua pendekatan tersebut diperlihatkan Tabel 6 yang memperlihatkan bahwa metode *elbow* menghasilkan jumlah *cluster* yang lebih baik dibandingkan dengan metode *silhouette*.

Tabel 6. K-Means Clustering

Metode Untuk Penentuan K-Optimal	Jumlah Cluster	Nilai WSS
<i>Elbow</i>	5	89.6 %
<i>Silhouette</i>	2	54.8%

Visualisasi hasil *clustering* dengan *k-means* menggunakan metode *elbow* dan *silhouette* dapat dilihat pada Gambar 3.



Gambar 3. Visualisasi Hasil K-Means Clustering dengan Elbow Method (Kiri) dan Silhouette Method (Kanan)

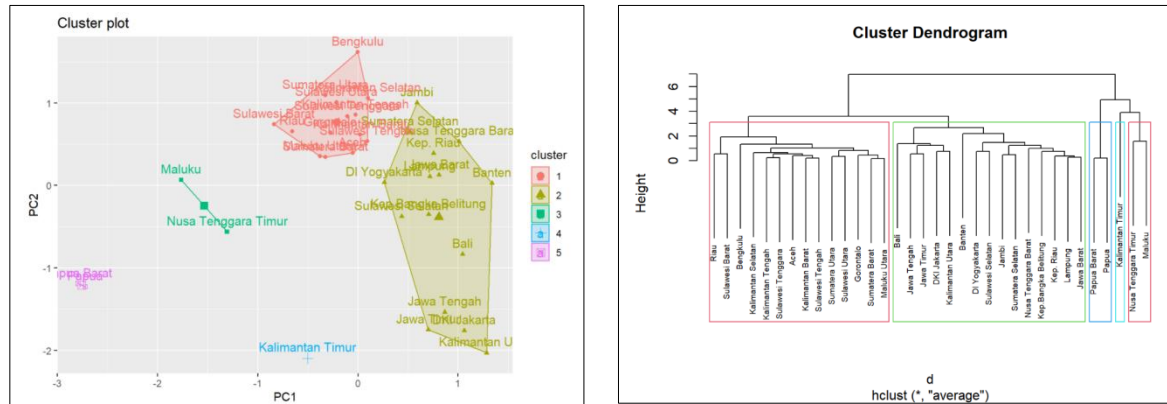
Agglomerative Hierarchical Clustering

Tahap *clustering* dengan pendekatan *hierarchical* akan diukur menggunakan *euclidean distance* dengan tingkat kemiripannya melalui metode *single linkage*, *ward method*, dan *average linkage*. Performa tingkat kemiripan pada *hierarchical clustering* ini selanjutnya akan dibandingkan dengan menghitung korelasi dari *cophenetic distance* dengan jarak aslinya yang ditunjukkan pada Tabel 7.

Tabel 7. K-Means Clustering

Jenis Tingkat Kemiripan	Korelasi Dari Cophenetic Distance Dengan Jarak Aslinya
Single Linkage	0.7973
Average Linkage	0.7981
Ward	0.7565

Penentuan jumlah *cluster* optimum juga dilakukan dengan menggunakan D indeks dan menghasilkan kesimpulan bahwa jumlah *cluster* terbaik untuk *hierarchical clustering* adalah 5 *cluster*. Maka, *hierarchical clustering* yang akan digunakan adalah yang memiliki tingkat kemiripan *average* (karena memiliki nilai korelasi terbesar sebagaimana pada Tabel 7. dengan jumlah *cluster* sebanyak 5. Maka, terbentuklah *cluster* sebagaimana ditunjukkan pada Gambar 4.



Gambar 4. Visualisasi Sebaran dan Dendrogram Hasil Hierarchical Clustering dengan Average Linkage dan Jumlah Cluster Optimal Adalah 5

Validasi

Validasi Internal

Verifikasi untuk melihat pengaruh perbedaan penggunaan metode dan pendekatan pada clustering provinsi berdasarkan indikator pelayanan kesehatan maternal di Indonesia dilakukan dengan membandingkan nilai dari indikator untuk validasi internal. Indikatornya meliputi indeks *connectivity*, indeks *dunn*, dan indeks *silhouette* yang hasilnya tertera pada Tabel 8.

Tabel 8. Validasi Internal

Jumlah Cluster (K)		2	3	4	5
<i>Connectivity</i>	<i>Hierarchical</i>	6.9544	9.4877	10.7877	13.9167
	<i>K-Means</i>	6.6587	11.6579	16.7940	18.1940
<i>Dunn</i>	<i>Hierarchical</i>	0.2987	0.2987	0.2987	0.1848
	<i>K-Means</i>	0.2987	0.1502	0.1204	0.1283
<i>Silhouette</i>	<i>Hierarchical</i>	0.5723	0.4777	0.3910	0.5130
	<i>K-Means</i>	0.6131	0.5002	0.4867	0.4889

Nilai indeks *connectivity* yang kecil akan menghasilkan cluster yang baik. Sebaliknya, indeks *dunn* dan *silhouette* yang semakin mendekati satu yang akan menghasilkan cluster terbaik. Berdasarkan Tabel 8, metode *hierarchical clustering* mayoritas lebih unggul. Pada saat $k = 5$, *hierarchical clustering* unggul di ketiga indeks validasi internal tersebut. Selain itu nilai $k=5$ juga sudah dijelaskan di bagian sebelumnya sebagai k paling optimal. Maka dari itu, untuk kasus penelitian ini metode clustering terbaik diraih oleh *hierarchical clustering* dengan jumlah cluster sebanyak 5.

Cluster Profiling

Cluster profiling digunakan untuk mengetahui karakteristik dari cluster yang terbentuk. Karakteristik cluster akan diukur melalui rata-rata untuk tiap variabel di cluster terkait.

Tabel 9. Rata-Rata Nilai Dari Tiap Variabel Menurut Cluster

Nama Variabel	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
TTD	75.53	89.67	55.2	74.4	27.55
KelasIbuHamil	87.04	83.05	49.4	46.00	7.15
DDHB	41.86	58.07	37.26	55.93	21.14
K1	80.71	97.67	73.6	89.5	58.05
K4	74.61	91.63	59.25	32.2	31.00
PersalinanNakes	80.26	96.28	55.35	86.7	49.00
PersalinanFasyankes	76.16	93.75	45.95	35.1	39.65
PersalinanKF1	81.11	95.92	57.4	87.3	37.75
PersalinanKFlengkap	78.82	94.93	53.45	80.00	31.10
NifasVitA	82.07	94.74	56.45	80.60	35.45
PeriksaHIV	25.44	56.62	21.99	45.35	31.81
KN1	79.79	99.5	65.05	90.5	40.30
KNLengkap	78.08	97.69	58.60	85.4	34.60
PuskesmasBidan	9.72	13.51	18.40	31.00	23.75

Berdasarkan Tabel 9 dapat disimpulkan bahwa capaian cakupan pelayanan kesehatan maternal paling baik ditunjukkan oleh *cluster* 2 karena rata-rata tertinggi tiap variabel didominasi oleh *cluster* 2 sehingga bisa dilabel sebagai *cluster* “sangat baik”. Sebaliknya, capaian rata-rata terendah didominasi oleh *cluster* 5 sehingga bisa dilabel sebagai *cluster* “sangat buruk”. Pada *cluster* 1, rata-rata capaiannya juga tinggi namun tak setinggi *cluster* 2 sehingga bisa dilabel sebagai *cluster* “baik”. Sementara itu, capaian *cluster* 4 masih berada di angka normal sehingga bisa dilabel sebagai *cluster* “cukup”. Pada *cluster* 3, capaian rata-ratanya cukup rendah tetapi tidak serendah capaian pada *cluster* 3 sehingga bisa dilabel sebagai *cluster* “buruk”. Maka, rincian daftar keanggotaan *cluster* ditunjukkan pada Tabel 10.

Tabel 10. Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Pelayanan Kesehatan Maternal Tahun 2020 Dengan Metode Hierarchical Clustering Dengan Average Linkage

Cluster	Nama Cluster	Provinsi	Jumlah Provinsi
1	Baik	Aceh, Sumatera Utara, Sumatera Barat, Riau, Bengkulu, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara	14
2	Sangat Baik	Jambi, Sumatera Selatan, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, NTB, Kalimantan Utara, Sulawesi Selatan	15
3	Buruk	Nusa Tenggara Timur dan Maluku	2
4	Cukup	Kalimantan Timur	1
5	Sangat Buruk	Papua dan Papua Barat	2

5. Kesimpulan

Berdasarkan data sekunder dari 34 provinsi di Indonesia yang diperoleh dari Publikasi Profil Kesehatan Indonesia Tahun 2020 oleh Kementerian Kesehatan Indonesia dengan atribut awal sebanyak 15 variabel diidentifikasi terdapat korelasi antarvariabel. Maka, proses reduksi variabel dengan analisis faktor melalui pendekatan *principal component analysis* (PCA) tetap dilakukan namun variabel persentase cakupan imunisasi td2+ pada ibu hamil tidak dapat diikuti dalam analisis karena nilai MSA

yang tidak memenuhi syarat. Kemudian analisis faktor menghasilkan dua faktor yang selanjutnya akan dijadikan data dasar untuk *clustering* dengan menggunakan algoritma *agglomerative hierarchical* dan *k-means*. Dengan menggunakan validasi intenat mencakup indeks *connectivity*, indeks *dunn*, dan indeks *silhouette* dapat disimpulkan bahwa metode *clustering* 34 provinsi di Indonesia yang paling baik diantara *hierarchical* dan *k-means clustering* adalah metode *hierarchical clustering* dengan pendekatan tingkat kemiripan berupa *average linkage* dan jumlah cluster sebanyak lima. Hasil identifikasi karakteristik klaster memberikan hasil bahwa klaster 1 dengan anggota 14 provinsi dikategorikan dengan provinsi dengan cakupan pelayanan kesehatan maternal yang baik. Klaster 2 yang beranggotakan 15 provinsi dikategorikan sangat baik. Klaster 3 dengan anggota NTT dan Maluku dikategorikan buruk. Klaster 4 yang hanya terisi oleh Kalimantan Timur sudah dikategorikan cukup. Sementara itu, klaster 5 dengan anggota Papua dan Papua Barat masih memprihatinkan dengan status sangat buruk. Dengan mengetahui karakteristik dari tiap klaster, diharapkan bisa bermanfaat bagi pemerintah sebagai langkah awal penetapan prioritas kebijakan terkait pengembangan cakupan pelayanan kesehatan maternal.

Mengacu pada penelitian yang telah dilakukan, saran bagi penelitian kedepannya antara lain perlu untuk dilakukan klasterisasi tidak hanya terbatas pada level provinsi tetapi juga level kabupaten/kota agar jika diimplementasikan untuk pembentukan kebijakan pemerintah bisa lebih tepat sasaran dan spesifik. Kemudian, penggunaan algoritma *clustering* sangat luas tidak hanya terbatas pada dua algoritma sebagaimana pada penelitian ini sehingga perbandingan untuk lebih banyak ragam algoritma *clustering* juga perlu untuk dilakukan.

Referensi

- [1] D. Hapsari, P. Sari, and L. Indrawati, "Indeks Kesehatan Maternal Sebagai Indikator Jumlah Kelahiran Hidup," *Jurnal Ekologi Kesehatan*, vol. 14, no. 3, pp. 259–272, Sep. 2015.
- [2] M. A. Naga, "Kesehatan Ibu dan Anak," Jakarta, 2009.
- [3] "World Health Statistics 2019 : Monitoring For The SDGs," Switzerland, 2019. Accessed: Dec. 04, 2021. [Online]. Available: <https://apps.who.int/iris/bitstream/handle/10665/324835/9789241565707-eng.pdf>
- [4] Kementerian Perencanaan dan Pembangunan Nasional/Bappenas, *Terjemahan Tujuan dan Target Global Tujuan Pembangunan Berkelanjutan (TPB)*. Jakarta: Kementerian Perencanaan dan Pembangunan Nasional/Bappenas, 2017. Accessed: Dec. 04, 2021. [Online]. Available: https://sdgs.bappenas.go.id/wp-content/uploads/2017/09/Buku_Terjemahan_Baku_Tujuan_dan_Target_Global_TPB.pdf
- [5] S. Sumarmi, "Model Sosio Ekologi Perilaku Kesehatan Dan Pendekatan Continuum Of Care Untuk Menurunkan Angka Kematian Ibu," *The Indonesian Journal of Public Health*, vol. 12, no. 1, pp. 129–141, Dec. 2017, doi: 10.20473/IJPH.V12I1.2017.129-141.
- [6] UNICEF, "Maternal and newborn health and COVID-19 - UNICEF DATA," May 2020. <https://data.unicef.org/topic/maternal-health/covid-19/> (accessed Dec. 04, 2021).
- [7] R. Hida Nurriszka *et al.*, "Akses Ibu Hamil terhadap Pelayanan Kesehatan di Masa Pandemi COVID-19," *Jurnal Kebijakan Kesehatan Indonesia : JKKI*, vol. 10, no. 2, pp. 94–99, Jun. 2021, doi: 10.22146/JKKI.62752.
- [8] Cici Suhaeni, Anang Kurnia, and Ristiyanti, "Perbandingan Hasil Pengelompokan menggunakan Analisis Cluster Berhierarchy, K-Means Cluster, dan Cluster Ensemble (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil)," *Jurnal Media Infotama*, vol. 14, no. 1, pp. 31–38, Feb. 2018, doi: <https://doi.org/10.37676/jmi.v14i1.469>.
- [9] K. B. Aditya, D. Puspitaningrum, and Y. Setiawan, "Sistem Informasi Geografis Pemetaan Faktor-faktor Yang Mempengaruhi Angka Kematian Ibu (Aki) Dan Angka Kematian Bayi (Akb) Dengan Metode K-means Clustering (Studi Kasus: Provinsi Bengkulu) - Neliti," *Jurnal Teknik Informatika*, pp. 59–65, 2017, Accessed: Dec. 05, 2021. [Online]. Available:

- <https://www.neliti.com/publications/133712/sistem-informasi-geografis-pemetaan-faktor-faktor-yang-mempengaruhi-angka-kemati>
- [10] A. Nufus, “Pengelompokan Kabupaten/Kota Di Jawa Timur Berdasarkan Program Pelayanan Kesehatan Ibu,” *Medical Technology and Public Health Journal*, vol. 3, no. 1, pp. 9–16, Mar. 2019, doi: 10.33086/MTPHJ.V3I1.934.
 - [11] N. Thamrin, A. W. Wijayanto, S. Politeknik, and I. Stis, “Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island,” *Indonesian Journal of Statistics and Its Applications*, vol. 5, no. 1, pp. 141–160, Mar. 2021, doi: 10.29244/IJSA.V5I1P141-160.
 - [12] A. R. Damayanti and A. W. Wijayanto, “Comparison of Hierarchical and Non-Hierarchical Methods in Clustering Cities in Java Island using the Human Development Index Indicators year 2018,” *Eigen Mathematics Journal*, vol. 4, no. 1, pp. 8–17, Jun. 2021, doi: 10.29303/EMJ.V4I1.89.
 - [13] A. M. Sikana and A. W. Wijayanto, “Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering,” *Jurnal Ilmu Komputer*, vol. 14, no. 2, pp. 66–78, Sep. 2021, doi: 10.24843/JIK.2021.V14.I02.P01.
 - [14] N. Afira and A. W. Wijayanto, “Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019,” *Komputika : Jurnal Sistem Komputer*, vol. 10, no. 2, pp. 101–109, Sep. 2021, doi: 10.34010/KOMPUTIKA.V10I2.4317.
 - [15] E. Luthfi and A. W. Wijayanto, “Analisis perbandingan metode hirearchical, k-means, dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia,” *Jurnal Inovasi*, vol. 17, no. 4, pp. 761–773, Dec. 2021, Accessed: Dec. 26, 2021. [Online]. Available: <https://journal.feb.unmul.ac.id/index.php/INOVASI/article/view/10106>
 - [16] Kementerian Kesehatan RI, “Profil Kesehatan Indonesia Tahun 2020,” Jakarta, 2021.
 - [17] D. N. Gujarati and D. C. Porter, *Basic Econometric*, 5th ed. Douglas Reiner, 2009.
 - [18] A. T. Fitriyah, “Penerapan Metode Agglomerative Hierarchical Clustering Untuk Klasifikasi Program Studi Berdasarkan Kualitas Pelayanan Mahasiswa,” *Jurnal Ekonomi Syariah*, vol. 8, no. 2, pp. 194–202, Dec. 2018.
 - [19] J. F. Hair Jr, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 7th ed. New Jersey: Pearson Prentice Hall, 2010.
 - [20] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. 2007.
 - [21] S. Saraçlı, N. Do, ~ Gan, ~ Ismet, and D. ~ Gan, “Comparison of hierarchical cluster analysis methods by cophenetic correlation,” *Journal of Inequalities and Applications*, p. 203, 2013, Accessed: Dec. 20, 2021. [Online]. Available: <http://www.journalofinequalitiesandapplications.com/content/2013/1/203>
 - [22] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data : An Introduction to Cluster Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1990. doi: 10.1002/9780470316801.