

BAB III

METODOLOGI PENELITIAN

3.1. Jenis Penelitian dan Pengumpulan Data

Penelitian yang dilakukan menggunakan metode penelitian kuantitatif. Penelitian kuantitatif merupakan suatu pendekatan untuk menguji teori objekrif dengan mengkaji hubungan di antara variabel dengan variabel-variabel tersebut dapat diukur sehingga dapat ditulis secara numerik dan dianalisis menggunakan prosedur statistik (Creswell, 2012). Penelitian kuantitatif juga adalah sebuah proses penggalian informasi berupa pengetahuan yang menggunakan data dengan jenis angka sebagai alat untuk menganalisis keterengan terkait hal yang ingin kita dapatkan (Kasiram, 2010).

Data yang dipakai untuk penelitian ini adalah data sekunder. Data sekunder merupakan data yang diperoleh dari dokumen-dokumen melalui instansi terkait yang berkompeten (Riduwan, 2013). Data sekunder yang digunakan pada penelitian ini adalah data jumlah tenaga kesehatan di Indonesia pada tahun 2023. Data ini merupakan data sekunder yang diambil dari situs Badan Pusat Statistik Republik Indonesia.

3.2. Variabel Penelitian

Penelitian ini mengkaji variabel yang berkaitan dengan jumlah tenaga kesehatan pada setiap provinsi di Indonesia. Menurut Kementerian Kesehatan Republik Indonesia, terdapat 13 jenis tenaga kesehatan. Jenis-jenis tenaga kesehatan tersebut merupakan berbagai kategori profesi, seperti perawat, bidan, tenaga kefarmasian (apoteker), tenaga kesehatan masyarakat, tenaga gizi, tenaga medis (dokter spesialis, dokter umum, dan dokter gigi), tenaga kesehatan lingkungan, serta tenaga teknik biomedika. Selain itu, penelitian ini juga mencakup tenaga kefarmasian lainnya serta tenaga kesehatan di bidang laboratorium medik dan keterampilan fisik. Data dikumpulkan berdasarkan jumlah tenaga kesehatan di masing-masing provinsi,

sehingga dapat menggambarkan distribusi tenaga medis dan tenaga kesehatan lainnya di seluruh Indonesia.

Selain jumlah tenaga kesehatan berdasarkan jenis profesinya, penelitian ini juga memasukkan provinsi sebagai unit analisis utama. Setiap provinsi memiliki karakteristik tenaga kesehatan yang berbeda, yang mencerminkan variasi dalam ketersediaan sumber daya kesehatan di berbagai wilayah. Variabel-variabel tersebut memungkinkan analisis lebih lanjut terkait perbandingan jumlah tenaga kesehatan antarprovinsi serta ketersediaan profesi tenaga kesehatan di Indonesia. Variabel-variabel yang digunakan beserta definisi operasionalnya disajikan dalam tabel berikut.

Tabel 3.1 Variabel dan Definisi Operasional

Kode	Variabel	Definisi	Keterangan	Jenis
P	Provinsi	Wilayah administratif tempat tenaga kesehatan bekerja	Nama	Kategorik
X1	Perawat	Jumlah tenaga kesehatan yang bertugas dalam bidang keperawatan di suatu provinsi	Orang	Kuantitatif Diskrit
X2	Bidan	Jumlah tenaga kesehatan yang bertugas dalam bidang kebidanan di suatu provinsi	Orang	Kuantitatif Diskrit
X3	Tenaga Kefarmasian	Jumlah tenaga kesehatan yang bekerja dalam bidang kefarmasian, termasuk apoteker dan tenaga farmasi lainnya	Orang	Kuantitatif Diskrit

Kode	Variabel	Definisi	Keterangan	Jenis
X4	Tenaga Kesehatan Masyarakat	Jumlah tenaga kesehatan yang bertugas dalam bidang kesehatan masyarakat, termasuk edukasi dan promosi kesehatan	Orang	Kuantitatif Diskrit
X5	Tenaga Kesehatan Lingkungan	Jumlah tenaga kesehatan yang bertugas dalam bidang kesehatan lingkungan, seperti sanitasi dan pengendalian penyakit	Orang	Kuantitatif Diskrit
X6	Tenaga Gizi	Jumlah tenaga kesehatan yang bertugas dalam bidang gizi dan nutrisi masyarakat	Orang	Kuantitatif Diskrit
X7	Tenaga Medis	Jumlah tenaga kesehatan yang terdiri dari dokter umum, dokter spesialis, dan dokter gigi	Orang	Kuantitatif Diskrit
X8	Tenaga Ahli Teknologi Laboratorium Medik	Jumlah tenaga kesehatan yang bertugas dalam melakukan pemeriksaan spesimen biologis manusia untuk tujuan diagnosis dan pemantauan kesehatan	Orang	Kuantitatif Diskrit
X9	Tenaga Kesehatan Psikologi Klinis	Jumlah tenaga kesehatan yang bertugas dalam bidang psikologi klinis, termasuk layanan terapi psikologis	Orang	Kuantitatif Diskrit
X10	Tenaga Keterampilan Fisik	Jumlah tenaga kesehatan yang bertugas dalam bidang fisioterapi dan rehabilitasi fisik	Orang	Kuantitatif Diskrit

Kode	Variabel	Definisi	Keterangan	Jenis
X11	Tenaga Keteknisan Medis	Jumlah tenaga kesehatan yang bertugas dalam bidang teknis medis, seperti radiologi dan laboratorium klinis	Orang	Kuantitatif Diskrit
X12	Tenaga Teknik Biomedika	Jumlah tenaga kesehatan yang bertugas dalam bidang pengelolaan dan pemeliharaan alat medis	Orang	Kuantitatif Diskrit
X13	Tenaga Kesehatan Tradisional	Jumlah tenaga kesehatan yang berpraktik dalam pengobatan tradisional, seperti jamu dan akupunktur	Orang	Kuantitatif Diskrit

3.3. Prapemrosesan Data

Prapemrosesan data mencakup serangkaian teknik yang diterapkan pada data mentah agar sesuai untuk analisis atau pemodelan (Muslikh dkk., 2023). Proses ini mencakup pembersihan, transformasi, dan penataan data untuk meningkatkan kualitas, relevansi, dan efisiensi untuk tugas-tugas analisis selanjutnya (Pacovský, 2017). Tujuan utamanya adalah untuk memastikan data bersih dan konsisten, yang sangat penting untuk analisis yang akurat (Madrakhimov dkk., 2021), mengurangi waktu dan sumber daya komputasi yang dibutuhkan untuk *data mining* (García dkk., 2016), dan meningkatkan keandalan dan efisiensi pelatihan model dan hasil (Maharana dkk., 2022).

3.4. Standardisasi Variabel

Standardisasi penting ketika variabel independen kontinu diukur pada skala yang berbeda, karena hal ini dapat menyebabkan variabel-variabel tersebut

memberikan kontribusi yang tidak merata pada analisis (Kaufman & Rousseeuw, 2009). Misalnya, dalam segmentasi pelanggan, di mana tujuannya adalah mengelompokkan berdasarkan karakteristik yang sama, standardisasi melibatkan penyesuaian variabel untuk memastikan mereka memiliki rentang atau varian yang sebanding.

Untuk menghitung standardisasi dapat menggunakan metode dengan menghitung nilai z-score terlebih dahulu. Dalam hal ini, metode dilakukan dengan mengubah skala variabel asli agar memiliki rata-rata nol dan deviasi standar satu (Hastie dkk., 2009). Rumus untuk menghitung nilai z-score dapat ditulis:

$$z = \frac{x - \mu}{\sigma} \quad (3.1)$$

dengan:

z : nilai standardisasi variabel

x : nilai data

μ : nilai rata-rata

σ : nilai simpangan baku

3.5. Menentukan Ukuran Jarak

Pengukuran jarak adalah metode yang digunakan untuk menentukan kedekatan antara dua objek, memberikan nilai kuantitatif untuk menggambarkan kemiripan atau ketidakmiripan objek tersebut (P.-N. Tan dkk., 2006). Di antara berbagai metrik jarak, jarak Euclidean banyak digunakan baik dalam teknik klasterisasi hirarki maupun nonhirarki. Hal ini umumnya diterapkan dalam metode seperti *k-means* untuk pengelompokan nonhirarki, serta dalam banyak algoritma pengelompokan hirarki untuk mengukur ketidaksamaan antara titik data atau klaster (Hartigan & Wong, 1979). Jarak Euclidean sangat cocok untuk data dengan skala interval atau rasio,

menjadikannya salah satu ukuran yang paling sering digunakan dalam analisis kuantitatif. Secara matematis, jarak Euclidean dihitung sebagai akar kuadrat dari jumlah perbedaan kuadrat antara variabel-variabel yang sesuai (Johnson & Wichern, 2007).

3.6. Algoritma *Centroid Linkage*

Centroid Linkage adalah metode klusterisasi hirarki yang menghitung rata-rata semua objek dalam sebuah cluster untuk menentukan jarak antarcluster (Liantoni & Cahyani, 2017). Jarak antara dua kluster ditentukan dengan mengukur jarak antara *centroid* masing-masing. *Centroid* kluster mewakili nilai rata-rata pengamatan di seluruh variabel dalam kluster tertentu (Veeraiah & Vasumathi, 2014). Dengan menggunakan metode ini, *centroid* dihitung ulang setiap kali sebuah kluster baru dibuat, dan terus berlanjut hingga kluster menjadi stabil (Rahmawati, 2013).

Centroid mewakili rata-rata dari semua elemen dalam sebuah kluster. Setiap kali objek baru ditambahkan ke dalam kluster, *centroid* dihitung ulang, sehingga menyebabkannya diperbarui setiap kali ada penambahan (Johnson & Wichern, 2007). Keunggulan dari metode ini adalah tingkat sensitifitas yang rendah terhadap outlier dibandingkan dengan beberapa metode yang lain (Silvi, 2018). Selain itu, *centroid linkage* dapat diimplementasikan secara efisien, seperti yang ditunjukkan oleh algoritma yang mencapai kecepatan yang signifikan (hingga 36 kali lebih cepat) dibandingkan dengan metode tradisional dengan mengurangi jumlah perbandingan jarak yang diperlukan (Batani dkk., 2024). Metode ini juga efektif dalam menghasilkan pengelompokan yang optimal, sebagaimana dibuktikan dengan penerapannya dalam pengelompokan provinsi di Indonesia berdasarkan indikator sosial. Algoritma ini berhasil mengidentifikasi jumlah kluster yang optimal dengan koefisien silhouette yang masuk akal, yang mengindikasikan validitas kluster yang baik (Candra dkk., 2024).

Langkah-langkah dalam *centroid linkage*:

1. Perlakukan setiap pengamatan atau individu sebagai klasternya sendiri. Buatlah matriks jarak $D = \{d_{ik}\}$, yang merepresentasikan jarak antar klaster.
2. Tentukan jarak terkecil antara dua klaster, yang dilambangkan sebagai d_{uv} (jarak antara klaster u dan v).
3. Gabungkan klaster u dan v , dan perbarui matriks jarak yang sesuai.
4. Lakukan langkah 2 dan 3 secara berulang hingga $N - 1$ penggabungan selesai, catat jarak pada setiap penggabungan.
5. Pilih nilai *cut off* untuk menentukan jumlah klaster. Gunakan dendrogram untuk memvisualisasikan proses dan memutuskan di mana titik potongnya.

3.7. Metode *Elbow*

Metode *Elbow* adalah teknik yang digunakan untuk mengoptimalkan algoritma klusterisasi dan menentukan nilai parameter yang optimal. Dalam pengelompokan *k-means*, metode ini membantu mengidentifikasi jumlah klaster yang optimal dengan menganalisis jumlah kuadrat kesalahan dalam klaster (SSE) (Winarta & Kurniawan, 2021). Metode ini telah diterapkan pada berbagai domain, termasuk analisis data pengguna narkoba dan prediksi stroke. Untuk prediksi stroke, Metode *Elbow* digunakan untuk mengoptimalkan algoritma *K-Nearest Neighbors*, sehingga menghasilkan akurasi dan presisi yang lebih baik (Sutomo dkk., 2023).

Metode ini memberikan gagasan dengan cara memilih nilai klaster dan kemudian menambah nilai klaster tersebut untuk dijadikan model data dalam penentuan klaster terbaik, selain itu persentase perhitungan yang dihasilkan menjadi pembandingan antara jumlah klaster yang ditambah (Shi dkk., 2020). Hasil persentase yang berbeda dari setiap nilai klaster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Nilai klaster paling baik ketika nilai klaster pertama dan nilai klaster kedua membentuk sudut pada grafik atau nilainya paling kecil (Herdiana

dkk., 2025). Untuk melakukan perbandingan, kita perlu menghitung SSE dari setiap nilai kluster.

1. Hitung nilai kluster dengan berbagai jumlah kluster yang bervariasi misalkan nilai $k = 1$ hingga $k = 5$.
2. Kemudian, hitung nilai SSE dengan persamaan berikut:

$$SSE = \sum_{i=1}^k \sum_{x_i \in C_i} D(\vec{x}_i, \vec{C}_i)^2 \quad (3.2)$$

dengan:

k : jumlah kluster

d : nilai keanggotaan titik data ke- i ke pusat kluster \vec{C}_i

\vec{C}_i : pusat kluster ke- i

$D(\vec{x}_i, \vec{C}_i)^2$: jarak dari titik \vec{x}_i ke kelompok \vec{C}_i yang diikuti

3. Plot nilai SSE berdasarkan jumlah nilai k kluster.
4. Pilih nilai k optimal dengan cara mengambil nilai k yang mengalami penurunan paling besar dan selanjutnya hasil dari nilai k akan turun secara perlahan-lahan sampai hasil dari nilai k tersebut stabil maka didapat nilai k yang optimal.

3.8. Algoritma *K-Means*

Metode *k-means* adalah teknik pengelompokan nonhirarki yang banyak digunakan karena kesederhanaannya dalam implementasi (Arthur & Vassilvitskii, 2007). Pendekatan ini memungkinkan peneliti untuk menentukan jumlah kluster yang ingin mereka buat (Steinley, 2006). Entitas pada awalnya ditugaskan ke- k kluster, sering kali secara acak (Macqueen, 1967) Selanjutnya, rata-rata untuk setiap kluster dihitung, dan jarak setiap entitas dari pusat kluster (dihitung sebagai rata-rata kluster) diukur (Hartigan & Wong, 1979).

Langkah-langkah melakukan klusterisasi dengan algoritma *k-means* adalah sebagai berikut:

1. Tentukan jumlah kluster k berdasarkan hasil analisis yang diterapkan pada metode hirarki.
2. Pilih pusat awal kluster. Secara umum, pemilihan pusat dilakukan secara acak. Namun, karena pendekatan ini menggabungkan klusterisasi hirarki dan *k-means*, pusat kluster pada algoritma *k-means* ditentukan sebagai rata-rata data dalam satu kluster hasil klusterisasi hirarki (Khan & Ahmad, 2004).
3. Kelompokkan seluruh data atau objek ke dalam kluster terdekat. Kedekatan antar objek dihitung berdasarkan jarak antara objek tersebut, biasanya menggunakan metode jarak Euclidean. Begitu pula, kedekatan suatu data dengan kluster ditentukan dari jarak antara data tersebut dan pusat kluster.
4. Perbarui pusat kluster berdasarkan keanggotaan kluster baru. Pusat kluster dihitung sebagai rata-rata dari semua data atau objek dalam kluster tersebut.
5. Ulangi proses pengelompokan untuk setiap objek menggunakan pusat kluster baru. Jika pusat kluster tidak mengalami perubahan, proses klusterisasi dianggap selesai. Jika tidak, kembali ke langkah sebelumnya hingga pusat kluster menjadi konstan.

3.9. Validasi Kluster

Validasi kluster melibatkan berbagai teknik yang digunakan untuk mengevaluasi kualitas dan kinerja algoritma kluster. Metode-metode ini sangat penting untuk menganalisis kluster yang dihasilkan, mengidentifikasi jumlah kluster yang optimal, dan mendapatkan wawasan tentang konsistensi dan keunikan titik-titik data di dalam kluster. Pendekatan validasi umumnya dibagi menjadi tiga kategori: pendekatan internal, eksternal, dan indeks relatif. Pendekatan internal menilai hasil pengelompokan berdasarkan struktur inheren dataset, dengan mempertimbangkan aspek-aspek seperti kohesi di dalam kluster dan separasi antarkluster. Sebaliknya, pendekatan eksternal bergantung pada label kebenaran dasar yang telah ditentukan

untuk mengukur seberapa baik klusterisasi selaras dengan pengelompokan yang sebenarnya (Hassan dkk., 2024)

3.9.1. *Silhouette Coefficient*

Silhouette coefficient adalah metode yang pertama kali diperkenalkan oleh (Kaufmann & Rousseeuw, 1987) untuk interpretasi dan validasi kluster pada objek-objek. Metode ini digunakan untuk menentukan jumlah kluster optimal dengan menghitung rata-rata nilai *silhouette*, yang bertujuan untuk menilai kualitas kluster yang terbentuk (Hidayat & Hakim, 2021). Perhitungan *silhouette coefficient* dilakukan dengan menghitung nilai indeks *silhouette* untuk setiap data ke- i . Nilai *silhouette coefficient* diperoleh dengan menentukan nilai maksimum dari indeks *Silhouette global* untuk jumlah kluster yang bervariasi dari 2 hingga $n-1$, sesuai dengan persamaan dari Kaufmann & Rousseeuw (1987) sebagai berikut.

$$SC = \max_k SI(k) \quad (3.3)$$

dengan,

SC : *Silhouette Coefficient*

SI : *Silhouette Indeks Global*

k : Jumlah kluster

Nilai *Silhouette Index* (SI) untuk sebuah data ke- i terdiri dari dua komponen utama, yaitu a_i dan b_i . Komponen a_i merepresentasikan rata-rata jarak antara data ke- i dan seluruh data lain dalam kluster yang sama. Sementara itu, b_i dihitung sebagai rata-rata jarak antara data ke- i dan semua data dalam kluster lain yang berbeda, kemudian dipilih nilai terkecil dari rata-rata tersebut.

$$a_i^j = \frac{1}{m_j - 1} \sum_{\substack{r=1 \\ r \neq i}}^{m_j} d(\vec{x}_{ij}, \vec{x}_{rj}) \quad (3.4)$$

dengan,

j : Klaster

i : Indeks data ($i = 1, 2, \dots, m_j$)

a_{ij} : Rata-rata jarak data ke- i terhadap semua data dalam klaster j

m_j : Jumlah data dalam klaster j

$d(\vec{x}_{ij}, \vec{x}_{rj})$: Jarak data ke- i dengan data ke- r dalam klaster j

Kemudian menghitung nilai b_i dengan persamaan berikut.

$$b_{ij} = \min_{\substack{n=1, \dots, k \\ n \neq j}} \left[\frac{1}{m_n - 1} \sum_{\substack{r=1 \\ r \neq i}}^{m_n} d(\vec{x}_{ij}, \vec{x}_{rn}) \right] \quad (3.5)$$

dengan,

j : Klaster

i : Indeks data ($i = 1, 2, \dots, m_j$)

b_{ij} : Rata-rata jarak data ke- i terhadap semua data yang tidak dalam satu klaster yaitu klaster j dengan data ke- i

m_n : Jumlah data dalam klaster n

$d(\vec{x}_{ij}, \vec{x}_{rn})$: Jarak data ke- i dalam klaster j dengan data ke- r dalam klaster n

Nilai *Silhouette Index* data ke- i dalam kluster j yaitu SI_i^j , diperoleh dengan menggunakan persamaan berikut.

$$SI_{ij} = \frac{b_{ij} - a_{ij}}{\max(a_{ij}, b_{ij})} \quad (3.6)$$

dengan,

SI_{ij} : *Silhouette Index* data ke- i dalam satu kluster j

a_{ij} : Rata-rata jarak data ke- i terhadap semua data dalam kluster j

b_{ij} : Rata-rata jarak data ke- i terhadap semua data yang tidak dalam satu kluster yaitu kluster j dengan data ke- i

Adapun nilai rata-rata *silhouette index* kluster j , SI_j dihitung menggunakan persamaan berikut:

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_{ij} \quad (3.7)$$

dengan,

SI_j : Rata-rata *Silhouette Index* dari kluster j

SI_{ij} : *Silhouette Index* data ke- i dalam satu kluster j

m_j : Jumlah data dalam kluster j

i : Index data ($i = 1, 2, \dots, m_j$)

Lebih jauh nilai rata-rata *Silhouette Index* dari dataset, SI, diberikan oleh persamaan berikut:

$$SI = \frac{1}{k} \sum_{j=1}^k SI_j \quad (3.8)$$

dengan,

SI : Rata-rata *Silhouette Index* dari data

SI_j : Rata-rata *Silhouette Index* dari kluster j

k : Jumlah kluster

Kriteria pengukuran pengelompokan berdasarkan *Silhouette Coefficient* (SC) menurut Kaufman dan Rousseeuw (2009) pada tabel berikut.

Tabel 3.2 Kriteria Pengukuran Pengelompokan

Nilai <i>Silhouette Coefficient</i>	Kriteria
$0,70 \leq SC \leq 1,00$	Struktur Kuat
$0,50 \leq SC < 0,70$	Struktur Baik
$0,25 \leq SC < 0,50$	Struktur Lemah
$SC < 0,25$	Struktur Buruk

3.10. Interpretasi dan Profilisasi Kluster

Setelah melakukan klusterisasi, kluster yang terbentuk perlu diprofilisasi agar dapat memberikan pemahaman lebih lanjut. Menurut Kaufman dan Rousseeuw (2009), profilisasi kluster memberikan rangkuman dari karakteristik utama dari setiap kluster, biasanya melalui nilai rata-rata atau dari variabel-variabel yang diteliti. Profilisasi kluster tidak hanya mencakup perbandingan statistik, tetapi juga interpretasi dalam konteks domain aplikasi agar hasilnya dapat digunakan secara praktis (Bacher dkk., 2004).

Profilisasi kluster mencakup analisis terhadap *centroid*, yaitu rata-rata nilai objek dalam kluster untuk setiap variabel. Nilai *centroid* ini memungkinkan deskripsi

setiap klaster dengan memberikan label tertentu. Label suatu klaster juga dapat didasarkan pada manfaat yang akan dicari (Putriana dkk., 2016).

3.11. Analisis Profil

Analisis profil berkaitan dengan situasi di mana serangkaian p perlakuan (seperti tes, pertanyaan, dan sejenisnya) diberikan kepada dua atau lebih kelompok subjek. Semua respons harus dinyatakan dalam satuan yang sama. Selain itu, diasumsikan bahwa respons dari kelompok yang berbeda bersifat independen satu sama lain. Dalam analisis profil, pertanyaan mengenai kesetaraan vektor rata-rata ini diklasifikasikan ke dalam beberapa kemungkinan tertentu (Johnson & Wichern, 2007).

1. Uji Keparalelan Profil

Uji Keparalelan Profil adalah uji utama yang menjadi fokus dalam analisis profil berdasarkan kelompok (klaster) karena tahap pengujian ini mengidentifikasi dan mengevaluasi keidentikan setiap segmen dalam suatu profil. Segmen ini merujuk pada perbedaan nilai variabel yang sama di berbagai titik waktu atau perbedaan antar beberapa variabel pada satu titik waktu tertentu. Dengan demikian, segmen tersebut merupakan kemiringan garis antara rata-rata dua variabel yang berdekatan. Jika hipotesis nol uji kesejajaran ditolak, terdapat interaksi signifikan antara keanggotaan kelompok dengan variabel, atau antara keanggotaan kelompok dengan titik waktu (misalnya, dalam tes yang dilakukan berulang kali). Artinya, perubahan kenaikan atau penurunan antar pengukuran variabel berturut-turut berbeda setidaknya pada satu kelompok (Bulut & Desjardins, 2020).

Uji Keparalelan Profil untuk dua populasi yang berdistribusi normal dapat dituliskan sebagai berikut:

Hipotesis Uji:

$H_{01}: C_{\mu_1} = C_{\mu_2}$, artinya profil populasi kelompok pertama dan kedua paralel;

$H_{11}: C_{\mu_1} \neq C_{\mu_2}$, artinya profil populasi kelompok pertama dan kedua tidak paralel.

C adalah matriks kontras yang didapat dari persamaan berikut.

$$C_{(p-l) \times p} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \quad (3.9)$$

Untuk sampel independen dengan ukuran n_1 dan n_2 dari dua populasi, hipotesis nol dapat diuji dengan membangun observasi yang ditransformasi sebagai berikut:

$$C_{x_{1j}}, j = 1, 2, \dots, n_1 \quad (3.10)$$

dan

$$C_{x_{2j}}, j = 1, 2, \dots, n_2 \quad (3.11)$$

Masing-masing memiliki vektor rata-rata sampel $C_{\bar{x}_1}$ dan $C_{\bar{x}_2}$, serta matriks kovarians gabungan $CS_{pooled}C'$. Karena dua himpunan pengamatan yang ditransformasi masing-masing memiliki distribusi $N_{p-1}(C\mu_1, C\Sigma C')$ dan $N_{p-1}(C\mu_2, C\Sigma C')$, maka pengujian untuk keparalelan profil adalah sebagai berikut.

Tolak $H_{01}: C_{\mu_1} = C_{\mu_2}$ pada taraf α jika

$$T^2 = (\bar{x}_1 - \bar{x}_2)' C' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) CS_{pooled} C' \right]^{-1} C (\bar{x}_1 - \bar{x}_2) > c^2 \quad (3.12)$$

dengan:

$$c^2 = \frac{(n_1 + n_2 - 2)(p - 1)}{n_1 + n_2 - p} F_{p-1, n_1+n_2-p}(\alpha) \quad (3.13)$$

2. Uji Keberimpitan Profil

Apabila profil-profil tersebut paralel, profil yang pertama berada tepat di atas profil yang kedua ($\mu_{1i} > \mu_{2i} \forall i$) atau sebaliknya. Pada kondisi ini, profil-profil dapat berimpit jika total ketinggian $\mu_{11} + \mu_{12} + \cdots + \mu_{1p} = 1'\mu_1$ dan $\mu_{21} + \mu_{22} + \cdots + \mu_{2p} = 1'\mu_2$ adalah sama (Johnson & Wichern, 2007).

Uji keberimpitan profil yang paralel untuk dua populasi yang berdistribusi normal dapat dituliskan sebagai berikut:

Hipotesis Uji:

H_{02} : $1'\mu_1 = 1'\mu_2$, artinya profil populasi kelompok pertama dan kedua berimpit;

H_{12} : $1'\mu_1 \neq 1'\mu_2$, artinya profil populasi kelompok pertama dan kedua tidak berimpit

Tolak H_{02} : $1'\mu_1 = 1'\mu_2$ (profil berimpit) pada taraf signifikansi α jika

$$T^2 = 1'(\bar{x}_1 - \bar{x}_2) \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) 1'S_{pooled}1 \right]^{-1} 1'(\bar{x}_1 - \bar{x}_2) \quad (3.14)$$

$$> t_{n_1+n_2-2}^2(\alpha) = F_{1,n_1+n_2-2}(\alpha)$$

$$T^2 = \left(\frac{1'(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) 1'S_{pooled}1}} \right)^2 > t_{n_1+n_2-2}^2(\alpha) = F_{1,n_1+n_2-2}(\alpha) \quad (3.15)$$

Nilai S_{pooled} didapatkan dari perhitungan menggunakan formula berikut:

$$S_{pooled} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2} \quad (3.16)$$

3. Uji Kesamaan Tingkat Profil

Untuk profil-profil yang berimpit, langkah berikutnya adalah menguji kesamaan rata-rata setiap variabel pada masing-masing profil. Jika rata-rata variabel memiliki kesamaan tingkat, maka profil adalah sama (Johnson & Wichern, 2007).

$$\bar{x} = \frac{1}{n_1 + n_2} \left(\sum_{j=1}^{n_1} x_{1j} + \sum_{j=1}^{n_2} x_{2j} \right) = \frac{n_1}{n_1 + n_2} \bar{x}_1 + \frac{n_2}{n_1 + n_2} \bar{x}_2 \quad (3.17)$$

Jika profil adalah sama, maka $\mu_1 = \mu_2 = \dots = \mu_p$.

Hipotesis uji:

$H_{03}: C_{\mu} = 0$, artinya profil populasi kelompok pertama dan kedua memiliki rata-rata yang sama;

$H_{13}: C_{\mu} \neq 0$, artinya profil populasi kelompok pertama dan kedua memiliki rata-rata yang berbeda.

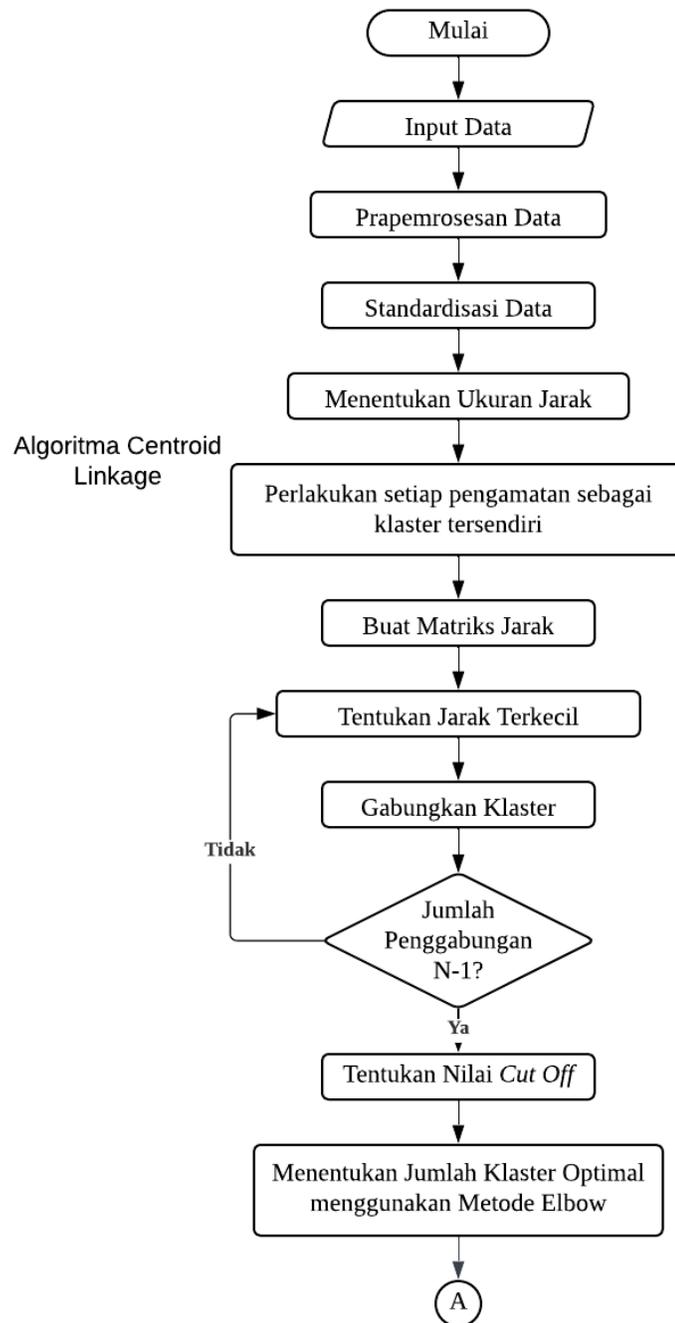
Tolak $H_{03}: C_{\mu} = 0$ (profil sama) pada taraf α jika

$$(n_1 + n_2)\bar{x}'C'[CSC']^{-1}C\bar{x} > c^2 \quad (3.18)$$

dengan S adalah sampel matriks kovarians dari semua $n_1 + n_2$ pengamatan, dan

$$c^2 = \frac{(n_1 + n_2 - 1)(p - 1)}{n_1 + n_2 - p + 1} F_{p-1, n_1+n_2-p+1}(\alpha) \quad (3.19)$$

3.12. Diagram Alir

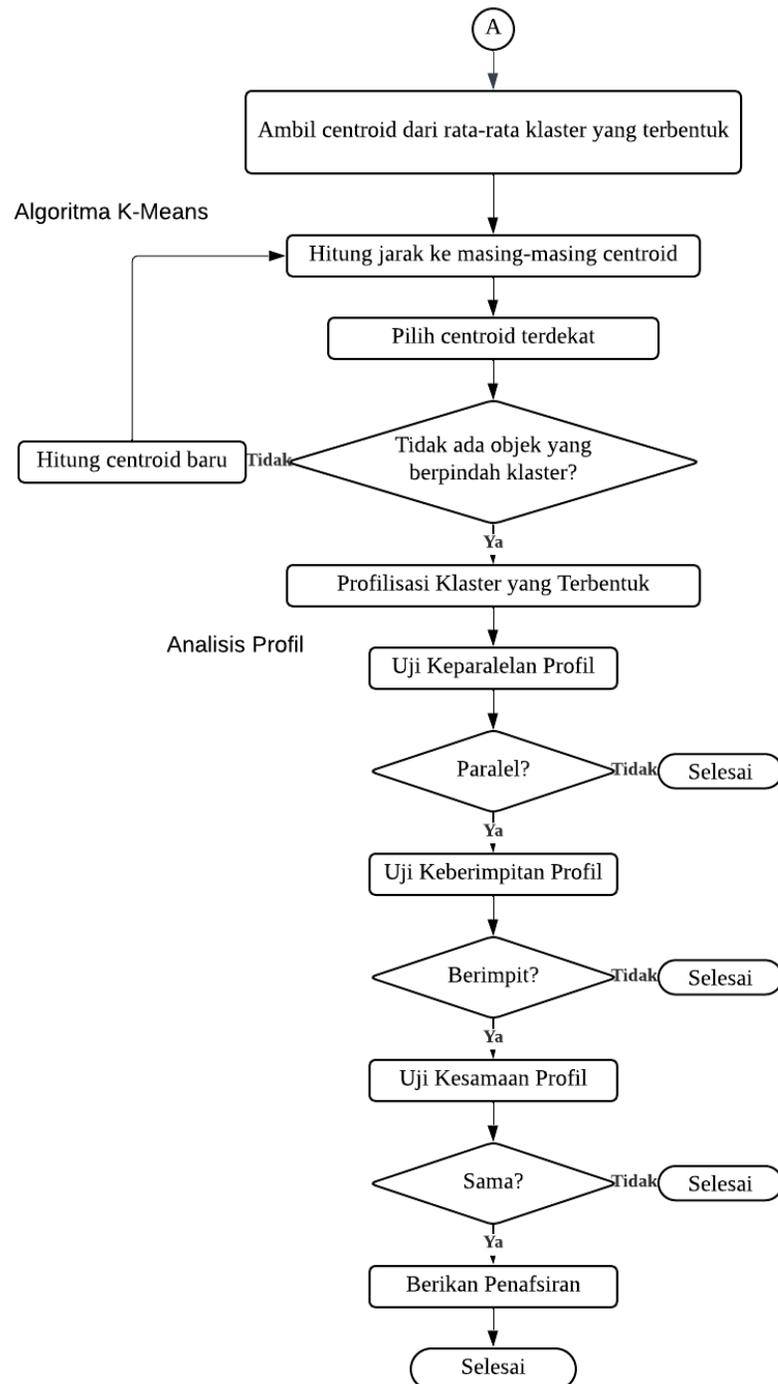


Gambar 3.1 Diagram Alir Penelitian

Barqy Muhammad Ilhan, 2025

KLASTERISASASI PROVINSI DI INDONESIA MENGGUNAKAN METODE K-MEANS DENGAN CENTROID LINKAGE BERDASARKAN JUMLAH TENAGA KESEHATAN TAHUN 2023

Univeristas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu



Gambar 3.2 Diagram Alir Penelitian (Lanjutan)

Barqy Muhammad Ilhan, 2025

KLASTERISASASI PROVINSI DI INDONESIA MENGGUNAKAN METODE K-MEANS DENGAN CENTROID LINKAGE BERDASARKAN JUMLAH TENAGA KESEHATAN TAHUN 2023

Univeristas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu