

# Clustering Provinsi di Indonesia Berdasarkan Indikator Kesehatan Ibu Tahun 2022 Menggunakan Metode Klaster Hirarki dan Non Hirarki

Chafshoh Nafilah<sup>1\*</sup>, Meysa Endies Irenia Putri<sup>2</sup>, dan Salma Zaura Baraza<sup>3</sup>

<sup>1</sup>Chafshoh Nafilah: Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

<sup>2</sup>Meysa Endies Irenia Putri: Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

<sup>3</sup>Salma Zaura Baraza: Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

\*Penulis Korespondensi: 5003211011@student.its.ac.id

**ABSTRAK** – Kesehatan ibu merupakan faktor krusial yang telah menjadi perhatian pemerintah mengingat jumlah kematian ibu Indonesia mencapai 4.005 pada tahun 2022. Salah satu target SDGs tahun 2030 adalah mereduksi Angka Kematian Ibu (AKI) karena peran penting ibu dalam mendidik penerus bangsa dengan kualitas terbaik. Penelitian ini bertujuan untuk menentukan klaster dari 34 provinsi di Indonesia berdasarkan indikator kesehatan ibu yang bersumber dari publikasi BPS berjudul Profil Kesehatan Ibu dan Anak 2022 guna memprioritaskan program pemerintah mendatang untuk kesehatan ibu Indonesia pada wilayah yang kritis. Pendekatan yang digunakan dalam penelitian ini yaitu metode klaster hirarki (*Linkage*) dan non hirarki (*K-Means*). Metode *K-Means* dengan empat klaster merupakan metode terbaik dengan nilai *pseudo-f statistics* tertinggi yaitu 10,977. Klaster pertama mengindikasikan provinsi dengan tingkat kesehatan ibu sangat rendah dan klaster keempat menunjukkan provinsi dengan tingkat kesehatan ibu tertinggi. Provinsi yang termasuk dalam klaster pertama meliputi Sulawesi Utara, Sulawesi Tengah, Maluku, Maluku Utara, Papua Barat, dan Papua yang merupakan wilayah Indonesia bagian timur, sehingga keenam provinsi ini perlu menjadi prioritas pemerintah dalam implementasi program kesehatan Ibu di masa mendatang.

**Kata kunci**– Kesehatan Ibu Indonesia, *Linkage*, *K-Means*, Program Kesehatan Ibu

**ABSTRACT** – Maternal health is a crucial factor that has become the government's concern considering the number of maternal deaths in Indonesia reached 4,005 in 2022. One of the targets of the SDGs in 2030 is to reduce the maternal mortality rate (MMR) because of the important role of mothers in educating the nation's next generation with the best quality. This study aims to determine the cluster of 34 provinces in Indonesia based on maternal health indicators sourced from BPS publication entitled Maternal and Child Health Profile 2022 to prioritize future government programs for Indonesian maternal health in critical areas. The approaches used in this study are hierarchical (*Linkage*) and non-hierarchical (*K-Means*) cluster methods. The *K-Means* method with four clusters was the best method with the highest *pseudo-f statistics* value of 10.977. The first cluster indicates provinces with very low maternal health levels and the fourth cluster indicates provinces with the highest maternal health levels. The provinces included in the first cluster include North Sulawesi, Central Sulawesi, Maluku, North Maluku, West Papua, and Papua which are the eastern part of Indonesia, so these six provinces need to be the government's priority in implementing maternal health programs in the future.

**Keywords**– Maternal Health Indonesia, *Linkage*, *K-Means*, Maternal Health Program

## I. PENDAHULUAN

Mengingat jumlah kematian ibu yang akan mencapai 4.005 pada tahun 2022, maka kesehatan ibu menjadi faktor penting yang menarik perhatian pemerintah Indonesia [17]. Tingginya angka kematian ibu (AKI) ini mencerminkan tantangan besar dalam sistem kesehatan Indonesia, khususnya dalam memberikan layanan kesehatan yang tepat dan berkualitas kepada ibu hamil dan melahirkan. Salah satu tujuan pembangunan berkelanjutan (SDGs) pada tahun 2030 adalah menurunkan Angka Kematian Ibu (AKI), karena ibu berperan penting dalam memberikan pendidikan berkualitas terbaik bagi generasi masa depan negara [20].

Menurut Profil Kesehatan Indonesia 2021 yang diterbitkan Kementerian Kesehatan, penyebab utama kematian ibu di Indonesia adalah perdarahan, hipertensi kehamilan, infeksi, dan komplikasi aborsi [19]. Situasi ini semakin diperparah dengan kesenjangan akses terhadap layanan kesehatan antara wilayah perkotaan dan perdesaan, serta wilayah barat dan timur Indonesia. Misalnya saja di Papua dan Maluku, tantangan geografis dan infrastruktur sering kali menyulitkan akses terhadap fasilitas kesehatan yang memadai, sehingga mengakibatkan penurunan kualitas layanan kesehatan ibu di wilayah tersebut [18].

Tingkat pendidikan ibu dan kesadaran kesehatan juga berperan penting terhadap kesehatan ibu. Ibu dengan tingkat pendidikan yang lebih tinggi cenderung lebih sadar akan pentingnya perawatan prenatal dan *post-natal* serta lebih mampu mengenali tanda-tanda bahaya pada masa kehamilan [5]. Namun data menunjukkan bahwa tingkat pendidikan perempuan masih relatif rendah di beberapa provinsi di Indonesia, sehingga berdampak negatif terhadap kesehatan ibu dan anak [20]. Selain itu, status sosial ekonomi juga merupakan faktor penting bagi kesehatan ibu. Keluarga dengan status ekonomi rendah sering kali memiliki akses terbatas terhadap layanan kesehatan berkualitas dan nutrisi yang tepat selama kehamilan [4]. Menurut laporan Bank Dunia, sekitar 25,1% penduduk Indonesia masih hidup di bawah garis kemiskinan nasional, yang berdampak langsung pada kesehatan ibu dan anak [22]. Mengatasi tantangan-tantangan ini

memerlukan analisis terperinci yang dapat mengidentifikasi bidang-bidang yang paling memerlukan intervensi. Teknik *clustering* dapat digunakan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator kesehatan ibu sehingga program pemerintah dapat lebih tepat sasaran dan efektif [7]. Pendekatan ini memungkinkan kita untuk mengidentifikasi bagian kelompok provinsi dari Indonesia dengan karakteristik serupa, yang dapat menjadi fokus upaya peningkatan kesehatan ibu. Penelitian ini menggunakan metode *clustering* hirarki (*Linkage*) dan non hirarki (*K-Means*) untuk menganalisis data kesehatan ibu yang diperoleh dari publikasi Badan Pusat Statistik berjudul Profil Kesehatan Ibu dan Anak 2022. Berdasarkan data indikator-indikator yang diperoleh dilakukan penentuan metode terbaik untuk mengelompokkan 34 provinsi di Indonesia. Pendekatan dengan metode terbaik diharapkan dapat menemukan pola yang jelas dalam sebaran permasalahan kesehatan ibu di berbagai provinsi sehingga intervensi pemerintah yang akan dilakukan akan lebih tepat sasaran dan efektif.

## II. KAJIAN PUSTAKA

### A. Statistika Deskriptif

Statistika deskriptif adalah metode pengumpulan dan menampilkan data untuk memberikan informasi yang bermanfaat. Tujuan mendeskripsikan secara statistika adalah untuk mengetahui ciri-ciri suatu keadaan serta menciptakan gambaran yang sistematis dan akurat mengenai fakta data yang diteliti. Contoh menampilkan data dalam statistik deskriptif antara lain tabel, bagan, dan grafik [10]. Penelitian ini menggunakan metode deskriptif berupa rata-rata, standar deviasi, nilai maksimum, dan nilai minimum untuk mendeskripsikan masing-masing variabel yang merupakan indikator-indikator kesehatan ibu di Indonesia. Statistika deskriptif disajikan dalam bentuk tabel, *bar chart*, *scatter plot*, *heatmap*, histogram, dan *boxplot*.

### B. Pengujian Multikolinearitas

Uji multikolinearitas bertujuan untuk mengetahui apakah terdapat korelasi tinggi atau sempurna antara variabel independen dalam suatu model. Pengujian ini dilakukan dengan mengevaluasi nilai toleransi dan *variance inflation factor* (VIF). Keputusan mengenai adanya multikolinearitas didasarkan pada kriteria yaitu nilai VIF yang kurang dari 10 menunjukkan tidak terjadi multikolinearitas, nilai VIF yang lebih dari 10 menunjukkan adanya multikolinearitas, dan nilai koefisien korelasi variabel independen yang lebih dari 0,8 menunjukkan terjadinya multikolinearitas [1].

### C. Analisis Clustering

*Clustering* adalah proses pengelompokan sekumpulan observasi dengan karakteristik yang memiliki tingkat kemiripan yang cukup tinggi dan disertai dengan properti yang sama dalam satu kelompok. *Clustering* merupakan salah satu metode utama yang ada pada *data mining*. *Cluster* merupakan kelompok yang tercipta dengan kemiripan satu sama lain antar anggota yang tidak dimiliki oleh anggota kelompok lainnya. Kompleksitas data dari segi *volume* observasi memiliki tingkat kesulitannya sendiri dapat terbantu dengan menggunakan metode ini. Salah satu algoritma konvensional namun efektif dalam proses pengclusteran adalah algoritma *K-Means* [1].

### D. Jarak Euclidean

Jarak *Euclidean* yang digunakan untuk mengukur jarak antar objek dengan data tipe numerik. Salah satu algoritma yang menggunakan jarak ini adalah algoritma *K-Means* [12]. Persamaan dari jarak *Euclidean* sebagai berikut.

$$d_{ij} = \left[ \sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}}$$

Dengan  $d_{ij}$  merupakan jarak *Euclidean* antara objek ke- $i$  dengan objek ke- $j$ . Nilai objek ke- $i$  pada peubah ke- $k$  digambarkan oleh  $x_{ik}$  dan nilai objek ke- $j$  pada peubah ke- $k$  digambarkan oleh  $x_{jk}$  dengan banyaknya peubah yang diamati disimbolkan dengan  $p$ .

### E. Metode Hirarki Linkage

Teknik pengelompokan hirarki atau *hierarchical clustering* merupakan salah satu teknik penting dalam analisis kluster yang mengelompokkan data ke dalam hirarki kluster bertingkat. Keuntungan utama pengelompokan hirarki adalah memungkinkan dalam memvisualisasikan dengan jelas struktur hirarki *cluster*. Metode ini tidak memerlukan penentuan jumlah *cluster* sebelumnya, sehingga interpretasi hasil lebih fleksibel. Namun, metode ini mempunyai kelemahan yaitu membutuhkan waktu komputasi yang lama terutama untuk data dalam jumlah besar [8]. Metode hirarki *clustering* yang dipakai dalam analisa ini meliputi *single linkage*, *complete linkage*, *average linkage*, dan *ward linkage* [8].

### F. Metode K-Means

Algoritma *K-Means* merupakan metode kluster non hirarki yang dapat menentukan kluster setiap observasi berdasarkan nilai rata-rata terdekat. Langkah-langkah algoritma *K-Means* adalah membagi observasi kedalam  $k$  kluster awal, melakukan klustering kedalam kluster yang memiliki nilai rata-rata terdekat, lalu melakukan perhitungan ulang nilai rata-rata untuk kluster dapat menerima observasi baru ataupun kehilangan observasi sebelumnya hingga tidak

terdapat perpindahan signifikan lagi terhadap observasi yang ada [11].

### G. Pseudo-F -Statistics

Nilai *Pseudo-F Statistics* merupakan salah satu kriteria untuk menentukan jumlah kelompok optimum dalam metode *clustering* [23]. Persamaan untuk mendapatkan nilai *Pseudo-F Statistics* ini digambarkan dalam persamaan berikut.

$$Pseudo - F Statistics = \frac{\left( \frac{R^2}{k-1} \right)}{\left( \frac{1-R^2}{n-k} \right)}$$

Dengan nilai

$$R^2 = \frac{SST - SSW}{SST}$$

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_j)^2$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2$$

SST (*Sum Square Total*) merupakan jumlah dari kuadrat jarak sampel terhadap rata-rata keseluruhan sedangkan SSW (*Sum Square Within*) merupakan jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya. Dengan n menunjukkan banyaknya sampel, c adalah banyak variabel, dan p banyak kelompok.

### H. Metode Elbow

*Elbow Method* merupakan salah satu metode yang memperlihatkan sebuah penurunan drastis pada titik-titik tertentu yang digambarkan dalam sebuah grafik yang terdapat lekukan tajam. Nilai k optimum diambil saat terjadi 'elbow', yakni penurunan tajam pada nilai *cost (inertia)* [12]. Dalam menentukan jumlah *cluster*, perhitungan dilakukan dengan menggunakan nilai SSE pada setiap k dengan persamaan sebagai berikut.

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} |X_i - C_k|_2^2$$

Dengan k merupakan jumlah *cluster* yang diinisiasi di awal hingga dinaikkan,  $X_i$  adalah nilai atribut data ke-i, dan  $C_k$  menggambarkan jumlah *cluster* i pada klaster ke-k.

### I. Pengujian Kruskal Wallis

Kruskal Wallis merupakan salah satu metode statistika nonparametrik, yaitu metode yang digunakan untuk analisis data ketika tidak memenuhi asumsi distribusi normal. pengujian hipotesis dengan metode Kruskal-Wallis merupakan bentuk pengembangan dari metode ANOVA satu arah, yang digunakan ketika beberapa asumsi yang diperlukan untuk analisis parametrik tidak terpenuhi [13]. Beberapa asumsi lain yang memenuhi untuk diterapkan pengujian ini meliputi data merupakan sampel acak, jumlah sampel tidak sama besar, skala data ordinal, masing-masing tidak saling memengaruhi, dan merupakan variabel acak kontinyu dengan hipotesis sebagai berikut.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_n$$

$$H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_n$$

Persamaan statistik uji untuk pengujian Kruskal Walli sebagai berikut.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

Nilai H adalah perhitungan Kruskal Wallis,  $R_i$  adalah jumlah rank dari kelompok ke-i,  $n_i$  menggambarkan banyaknya variabel dalam sampel pada kelompok ke-i, k merupakan banyaknya kelompok, dan N adalah jumlah seluruh observasi.

## III. METODOLOGI

### A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Profil Kesehatan Ibu dan Anak 2022 yang merupakan katalog yang dipublikasikan oleh Badan Pusat Statistik Indonesia. Publikasi ini memuat informasi mengenai gambaran kesehatan penduduk Indonesia, membangun bangsa melalui kesehatan ibu, kesehatan anak yang merupakan investasi masa depan bangsa, hingga pada pola pengeluaran kesehatan penduduk Indonesia. Penelitian ini mengerucut pada pembentukan klaster dari 34 provinsi di Indonesia menggunakan indikator-indikator kesehatan ibu di Indonesia pada tahun 2022 yang dijelaskan pada Tabel 1.

### A. Variabel Penelitian

Penelitian ini menggunakan tujuh variabel yang merupakan provinsi di Indonesia dan indikator-indikator kesehatan ibu Indonesia dengan masing-masing skala sebagai berikut.

**Tabel 1** Variabel Penelitian

Notasi	Variabel	Skala
Provinsi	Provinsi di Indonesia	Nominal
$X_1$	Rata – rata jumlah rokok yang dihisap ibu per minggu	Rasio
$X_2$	Persentase ibu merokok (dari semua ibu perokok) dengan jumlah $\geq 15$ batang per hari	Rasio
$X_3$	Persentase ibu yang melahirkan anak meninggal ( $X_4$ )	Rasio
$X_4$	Persentase ibu yang melahirkan anak dengan berat badan rendah	Rasio
$X_5$	Persentase ibu yang rawat inap dengan jaminan kesehatan	Rasio
$X_6$	Persentase ibu yang memiliki keluhan sakit dan rawat jalan	Rasio

### B. Struktur Data

Struktur data dalam penelitian ini dijelaskan dalam Tabel 2 yang disajikan sebagai berikut.

**Tabel 2** Struktur Data Variabel Penelitian

No	Provinsi	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$
1	Aceh	70	15,27	30,51	15,48	96,70	59,01
2	Sumatera Utara	56	10,80	43,39	11,93	68,99	23,10
3	Sumatera Barat	97	45,80	37,30	10,17	72,73	41,93
4	Riau	76	17,77	40,86	11,12	70,83	45,55
...	...	...	...	...	...	...	...
33	Papua Barat	34	6,39	49,83	14,51	84,12	66,68
34	Papua	38	5,73	50,66	13,13	78,48	67,61

### C. Langkah Analisis

Penelitian ini melakukan beberapa tahapan analisis yaitu sebagai berikut.

1. Mengidentifikasi dan merumuskan masalah
2. Menentukan variabel indikator-indikator kesehatan ibu Indonesia tahun 2022
3. *Pre-processing* dan eksplorasi data
4. Melakukan pengujian multikolinearitas
5. Melakukan analisis kluster hirarki menggunakan metode *Linkage*
6. Melakukan analisis kluster non hirarki menggunakan metode *K-Means*
7. Menentukan model terbaik menggunakan nilai *Pseudo-F Statistics*
8. Melakukan pengujian Kruskal Wallis antar kluster yang didapatkan
9. Visualisasi hasil kluster
10. Menarik kesimpulan

## IV. HASIL ANALISIS DAN PEMBAHASAN

### A. Statistika Deskriptif

Tahun 2022 Indonesia terdiri dari 34 provinsi dimana setiap provinsi memiliki karakteristik kesehatan ibu yang berbeda-beda, sehingga perlu dilakukan analisis mengenai karakteristik provinsi berdasarkan pada indikator kesehatan ibu sesuai skala datanya. Berikut disajikan tabel mengenai karakteristik kesehatan ibu di Indonesia.

**Tabel 3** Statistika Deskriptif

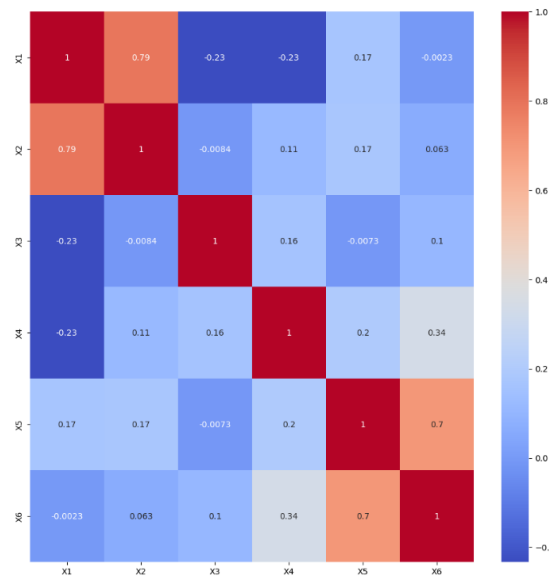
Variabel	Rata - Rata	Standar deviasi	Minimum	Maksimum
$X_1$	76,50	24,57	34	133
$X_2$	26,32	17,71	3,59	84,47
$X_3$	37,11	6,32	26,41	51,13
$X_4$	13,06	2,30	10,17	19,19
$X_5$	78,77	9,94	60,99	96,70
$X_6$	46,78	13,36	21,28	67,61

Tabel diatas menunjukkan bahwa indikator kesehatan ibu memiliki rata – rata yang berbeda. Rata – rata yang terendah yaitu persentase ibu yang melahirkan anak dengan berat badan rendah ( $X_4$ ) yakni sebesar 13,06%. Hal tersebut menunjukkan bahwa pemahaman ibu mengenai kesehatan dan perawatan selama kehamilan sudah cukup baik namun perlu dilakukan sosialisasi lebih lanjut supaya persentase bisa lebih menurun lagi. Sedangkan rata – rata tertinggi yaitu persentase ibu yang rawat inap dengan jaminan kesehatan ( $X_5$ ), hal ini menunjukkan bahwa cakupan masyarakat khususnya ibu hamil dalam kepesertaan jaminan kesehatan sudah cukup baik. Pada standar deviasi dapat terlihat bahwa nilai yang dihasilkan cukup beragam dan beberapa variabel menunjukkan nilai yang cukup besar. Dimana hal tersebut mengindikasikan bahwa terdapat ketidakmerataan pada kondisi kesehatan ibu di tiap provinsi Indonesia dan dapat dikatakan pula bahwa kualitas kesehatan ibu di Indonesia sangat beragam.

### B. Visualisasi Data

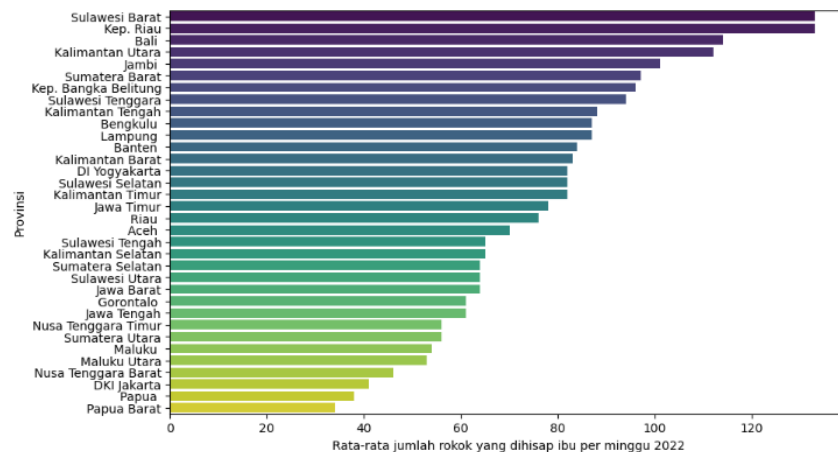
Visualisasi data merupakan proses menyajikan data dalam bentuk grafik, diagram, atau peta untuk memudahkan

pemahaman dan analisis. Tujuan utama dari visualisasi data adalah untuk mengubah data kompleks menjadi informasi yang mudah dipahami dan diinterpretasikan. Berikut disajikan beberapa hasil visualisasi data.



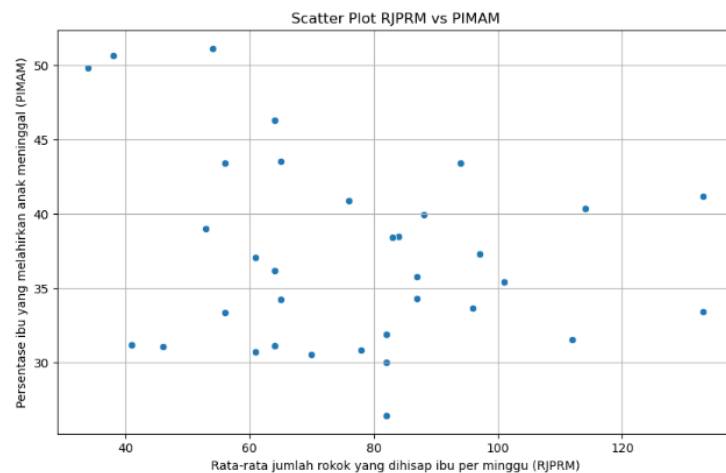
**Gambar 1** Heatmap Korelasi

Heatmap korelasi yang ditampilkan menunjukkan hubungan antara enam variabel yang diberi label  $X_1$  hingga  $X_6$ . Skala warna pada sisi kanan heatmap menunjukkan nilai korelasi, di mana warna merah tua mewakili korelasi positif yang kuat (mendekati 1), dan warna biru tua mewakili korelasi negatif yang kuat (mendekati -1). Warna-warna yang lebih terang menunjukkan korelasi yang lebih lemah atau tidak ada korelasi. Berdasarkan hasil terlihat bahwa  $X_1$  dan  $X_2$ ,  $X_5$  dan  $X_6$  memiliki korelasi positif yang sangat kuat menunjukkan ketika salah satu variabel naik maka variabel lain juga akan naik. Sebaliknya,  $X_1$  dan  $X_3$ , serta  $X_1$  dan  $X_4$  memiliki korelasi negatif yang relatif signifikan, masing-masing dengan nilai -0.23, mengindikasikan bahwa ketika nilai  $X_1$  naik, nilai  $X_3$  dan  $X_4$  cenderung turun. Korelasi antara variabel lainnya, seperti  $X_2$  dan  $X_3$ , serta  $X_2$  dan  $X_4$ , menunjukkan hubungan yang sangat lemah dengan nilai-nilai mendekati 0, mengindikasikan tidak ada hubungan linear yang signifikan antara variabel-variabel tersebut.

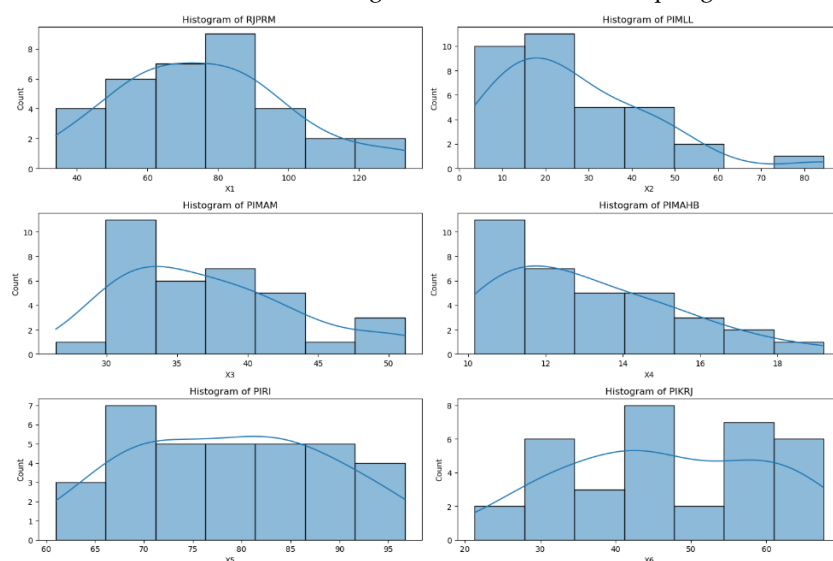


**Gambar 2** Barchart Rata-Rata Jumlah Rokok yang Dihisap Ibu Berdasarkan Provinsi

Visualisasi tersebut menampilkan rata – rata jumlah rokok yang dihisap oleh ibu per minggu di berbagai provinsi Indonesia pada tahun 2022. Terlihat dari grafik bahwa Sulawesi Barat dan Kepulauan Riau menempati posisi teratas dengan rata – rata jumlah rokok yang dihisap ibu per minggu paling tinggi. Hal tersebut menunjukkan bahwa ibu – ibu di provinsi tersebut memiliki resiko untuk memiliki masalah kesehatan yang lebih tinggi daripada provinsi lain. Sedangkan posisi terbawah ditempati oleh Papua Barat. Perbedaan tersebut bisa saja disebabkan karena perbedaan budaya dan kebiasaan merokok di tiap provinsi. Oleh karena itu, perlu dilakukan kebijakan pemerintah untuk mengurangi kebiasaan tersebut mengingat tingginya bahaya rokok terhadap kesehatan ibu.


 Gambar 3 Scatter Plot  $X_1$  dan  $X_3$ 

Scatter plot di atas menunjukkan hubungan antara rata – rata jumlah rokok yang dihisap ibu per minggu dan presentase yang melahirkan anak meninggal di berbagai provinsi. Dari visualisasi ini, terlihat bahwa tidak ada pola yang jelas atau kuat yang menunjukkan korelasi langsung antara jumlah rokok yang dihisap ibu dan persentase ibu yang melahirkan anak meninggal. Titik-titik data tersebar di seluruh area plot, menunjukkan variasi yang cukup besar dalam kedua variabel. Meskipun ada beberapa titik data yang menunjukkan bahwa provinsi dengan tingkat merokok yang lebih tinggi memiliki persentase kematian anak yang lebih tinggi, pola ini tidak konsisten di seluruh provinsi. Variabilitas ini menunjukkan bahwa faktor lain selain merokok mungkin berkontribusi terhadap angka kematian anak saat lahir.

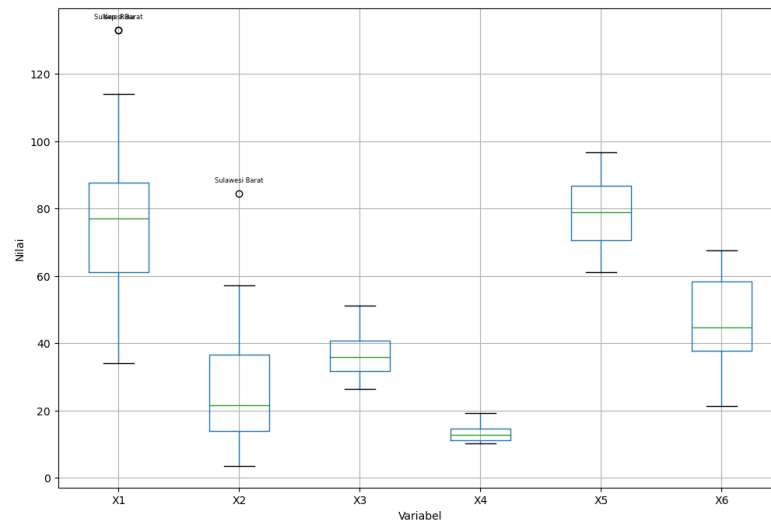


Gambar 4 Histogram

Histogram diatas menampilkan distribusi tiap variabel yang berpengaruh terhadap kesehatan ibu di tiap provinsi dengan rinciapi tiap variabel sebagai berikut.

- Pada variabel  $X_1$  menunjukkan bahwa distribusi data di tiap provinsi memiliki rata – rata konsumsi rokok ibu per minggu antara 60 hingga 80 batang, dengan beberapa provinsi menunjukkan konsumsi yang tinggi hingga 120 batang per minggu.
- Variabel  $X_2$  menunjukkan bahwa persentase ibu yang merokok (dari semua ibu perokok) dengan jumlah  $\geq 15$  batang per hari di tiap provinsi berada pada rentang 10% – 30%, hanya sedikit provinsi yang memiliki persentase diatas 50%.
- Persentase ibu yang melahirkan anak meninggal berdasarkan hasil histogram tiap provinsi memiliki persentase antara 30% hingga 45% dengan distribusi data menunjukkan kecenderungan yang lebih tinggi pada kisaran 35%
- Distribusi data variabel  $X_4$  menunjukkan bahwa sebagian besar provinsi memiliki persentase antara 10% hingga 14%.
- Data persentase ibu yang rawat inap dengan jaminan kesehatan menunjukkan distribusi yang relatif merata antara 60% hingga 95%. Tidak ada rentang tertentu yang mendominasi, menunjukkan variasi yang cukup besar antar provinsi.
- Distribusi data pada variabel  $X_6$  menunjukkan konsentrasi pada rentang 30% hingga 70% dengan variasi pada data yang cukup besar.





Gambar 5 Boxplot Variabel Penelitian

Boxplot setiap variabel penelitian menunjukkan distribusi data yang sangat berbeda dari masing-masing variabel karena memiliki skala satuan yang berbeda, oleh karena itu sebelum dilakukan analisa lebih lanjut dilakukan standarisasi data menggunakan algoritma *standar scaler*. Variabel  $X_1$  menunjukkan bahwa Provinsi Sulawesi Barat dan Kepulauan Riau memiliki nilai yang cukup tinggi dibandingkan provinsi lainnya menunjukkan kemungkinan tingkat kesehatan ibu di provinsi ini cukup rendah berdasarkan rata-rata jumlah rokok yang dihisap ibu setiap minggunya. Variabel  $X_2$  menunjukkan bahwa Provinsi Sulawesi barat juga memiliki persentase yang paling tinggi untuk ibu merokok dengan jumlah batang yang dihisap lebih dari 15 dibandingkan dengan 33 provinsi lainnya di Indonesia.

### C. Uji Multikolinearitas

Pemeriksaan multikolinearitas data bertujuan untuk mengidentifikasi ada atau tidaknya kolerasi antar variabel prediktornya, Salah satu kriteria untuk mendeteksi adanya multikolinearitas adalah nilai VIF dari masing-masing variabel prediktor, Berikut ini adalah nilai VIF dari semua variabel prediktor.

Tabel 4 Nilai VIF Data

Variabel	VIF
$X_1$	4,300290
$X_2$	3,871620
$X_3$	1,172487
$X_4$	1,605153
$X_5$	2,087840
$X_6$	2,187375

Berdasarkan nilai VIF dari masing-masing variabel di atas, dapat dilihat semua variabel prediktor memiliki nilai VIF kurang dari 10, sehingga dapat disimpulkan bahwa tidak terdapat multikolinearitas.

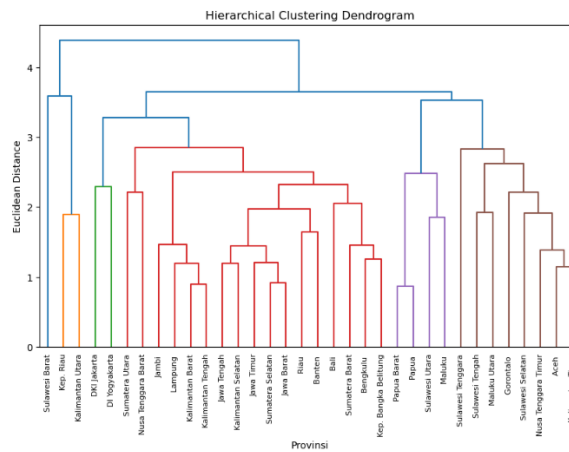
### D. Cluster Hirarki Linkage Spatial

Cluster hirarki linkage memiliki 4 jenis metode yaitu *single*, *ward*, *complete*, dan *average linkage*. Oleh karena itu, dicari metode dan jumlah cluster yang optimum. Pemilihan banyak kluster yang optimum menggunakan *Pseudo-F Statistics* dengan banyak kelompok 2, 3, 4, dan 5. Berdasarkan metode tersebut didapatkan hasil sebagai berikut.

Tabel 5 Nilai *Pseudo-F Statistics* Hirarki

Jumlah Kluster	<i>Single Linkage</i>	<i>Ward Linkage</i>	<i>Complete Linkage</i>	<i>Average Linkage</i>
2	3,9194	8,5608	3,9194	6,2126
3	2,6834	9,6333	6,4806	10,0387
4	2,1158	9,8051	7,7765	7,6213
5	2,6769	9,9334	7,9625	8,5960

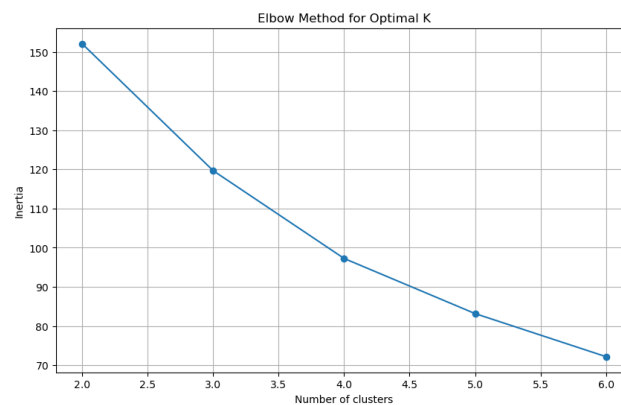
Berdasarkan Tabel x, nilai *Pseudo-F Statistics* pada metode hirarki tertinggi dari semua kombinasi metode dan jumlah cluster adalah 10,0387, yang ditemukan pada kelompok dengan 3 cluster menggunakan metode *average linkage*. Nilai Pseudo F-statistics yang tinggi ini menunjukkan bahwa 3 cluster dapat menghasilkan keragaman yang sangat homogen dalam setiap kelompok, sementara antar kelompok sangat heterogen. Hal ini menandakan bahwa jumlah cluster optimum yang terbentuk adalah 3 cluster. Hasil dari visualisasi cluster dengan menggunakan metode *average linkage* dengan jumlah 3 cluster adalah sebagai berikut.



Gambar 6 Cluster Hirarki Metode Average Linkage

### E. Cluster K-Means Spatial

Pada penelitian pengelompokan menggunakan metode *cluster* non-hirarki *K-Means*. Untuk penentuan jumlah optimum *cluster* dapat dilakukan dengan banyak cara, namun dalam penelitian ini akan menggunakan metode elbow dan berdasarkan Within-Cluster Sum of Squares. Dengan menggunakan visualisasi plot elbow adalah sebagai berikut.



Gambar 7 Elbow Method K-Means

Dari hasil *elbow method* di atas dapat dilihat bahwa grafik membentuk elbow pada *cluster* ke 4 sehingga dapat diduga bahwa jumlah *cluster* optimum untuk *cluster K-Means* adalah sebanyak 4 *cluster*. Setelah mendapatkan jumlah optimum *cluster* berdasarkan elbow method, sama pada metode hirarki, penentuan jumlah optimum *cluster* dapat dilakukan dengan menggunakan metode Pseudo-F Statistics. Sehingga, didapatkan nilai *Pseudo-F Statistics* pada metode *K-Means* sebagai berikut.

Tabel 6 Nilai *Pseudo-F Statistics* K-Means

Jumlah Klaster	<i>K-Means</i>
2	10,9281
3	10,9225
4	10,9767
5	10,5310

Berdasarkan Tabel 6, nilai *Pseudo-F Statistics* pada metode *K-Means* tertinggi dari semua jumlah *cluster* adalah 10,9767, yang terlihat pada kelompok dengan 4 *cluster*. Nilai Pseudo F-statistics yang tinggi ini menunjukkan bahwa 4 *cluster* dapat menghasilkan keragaman yang sangat homogen dalam setiap kelompok, sementara antar kelompok sangat heterogen.

### F. Metode Cluster Terbaik

Dari kedua metode *cluster* yang telah digunakan di atas baik *cluster K-Means* dan Hirarki Average Linkage. Pemilihan metode *cluster* terbaik didapatkan berdasarkan nilai *Pseudo-F Statistics* didapatkan hasil sebagai berikut.

Tabel 7 Nilai *Pseudo-F Statistics* Metode Clustering

Metode	Jumlah Cluster	<i>Pseudo-F Statistics</i>
Hierarchical Clustering Average Linkage	3	10,0387
K-Means	4	10,9767

Dari tabel hasil perbandingan nilai *Pseudo-F Statistics* kedua metode *cluster* tersebut, dapat dilihat bahwa metode



*cluster* terbaik yaitu metode *K-Means* dengan jumlah *cluster* sebanyak 4. Sehingga, *cluster* terbaik yang digunakan yaitu hasil dari *clustering* menggunakan metode *K-Means*. Berikut merupakan anggota *cluster* yang terbentuk.

Tabel 8 Anggota Setiap Cluster

Cluster	Anggota
0	Sulawesi Utara, Sulawesi Tengah, Maluku, Maluku Utara, Papua Barat, dan Papua.
1	Aceh, DKI Jakarta, DI Yogyakarta, Nusa Tenggara Timur, Kalimantan Timur, Sulawesi Selatan, dan Gorontalo.
2	Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, dan Kalimantan Selatan.
3	Kep. Riau, Bali, Kalimantan Utara, Sulawesi Tenggara, dan Sulawesi Barat.

### G. Uji Perbedaan Rata-rata

Setelah didapatkan model cluster yang terbaik, dilakukan *profiling* tiap *cluster* dengan melihat perbedaan dari rata-rata variabel yang digunakan. Sebelum itu, akan dilakukan pengecekan apakah data telah berdistribusi normal. Berikut hasil uji normalitas pada data.

Tabel 9 Uji Normalitas

Cluster	Statistics	P-Value	Keputusan	Kesimpulan
0	0,8571	0,0000	Tolak $H_0$	Tidak Normal
1	0,8571	0,0000	Tolak $H_0$	Tidak Normal
2	0,9772	0,0000	Tolak $H_0$	Tidak Normal
3	0,9987	0,0000	Tolak $H_0$	Tidak Normal

Karena data tidak berdistribusi normal, pengujian perbedaan rata-rata akan dilakukan dengan menggunakan metode Kruskal Wallis sebagai berikut.

### Hipotesis

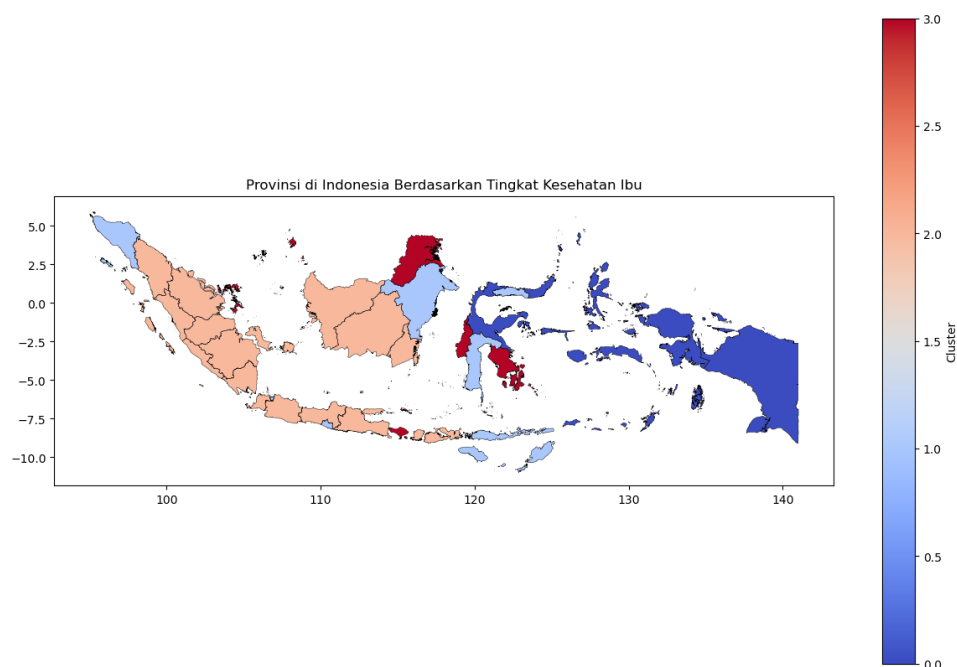
$H_0$  : Tidak terdapat perbedaan signifikan antara median atau distribusi kelompok-kelompok yang diuji.

$H_1$  : Terdapat perbedaan signifikan antara median atau distribusi setidaknya dua kelompok yang diuji.

Dengan nilai  $\alpha$  sebesar 0,05 dan nilai statistik uji H sebesar 0,6312 serta *p-value* 0,8892 memutuskan untuk gagal menolak  $H_0$  sehingga dengan taraf signifikansi 95% disimpulkan bahwa tidak terdapat perbedaan signifikan antara median atau distribusi antara kelompok-kelompok yang diuji.

### H. Hasil Cluster

Berdasarkan analisis yang telah dilakukan di atas, didapatkan empat *cluster* provinsi di Indonesia berdasarkan tingkat kesehatan ibu. Berikut ini akan disajikan visualisasi hasil *clustering*.



Gambar 8 Peta Hasil Cluster

Terlihat pada peta Indonesia di atas bahwa Indonesia yang dibagi berdasarkan tingkat kesehatan ibu menurut provinsi. Warna pada peta merepresentasikan kluster dari tingkat kesehatan ibu, dengan skala warna yang berkisar dari biru (*cluster* rendah) hingga merah (*cluster* tinggi). Dari peta tersebut, terlihat bahwa provinsi di wilayah timur Indonesia, seperti Sulawesi Utara, Sulawesi Tengah, Maluku, Maluku Utara, Papua Barat, dan Papua, berada pada *cluster* dengan tingkat kesehatan ibu yang paling rendah (ditandai dengan warna biru tua). Di sisi lain, beberapa provinsi di wilayah tengah dan barat Indonesia, seperti Kep. Riau, Bali, Kalimantan Utara, Sulawesi Tenggara, dan Sulawesi Barat, menunjukkan *cluster* dengan tingkat kesehatan ibu sangat tinggi (ditandai dengan warna merah). Hal ini mencerminkan perbedaan dalam akses dan kualitas layanan kesehatan ibu jika dilihat dari wilayah berdasarkan provinsi di Indonesia, dengan daerah-daerah di bagian timur yang relatif lebih tertinggal dibandingkan daerah-daerah di bagian barat dan tengah utamanya dalam segi kesehatan ibu Indonesia sejalan dengan sulitnya akses dan tingkat pendidikan yang rendah di enam provinsi di atas. Sehingga program kesehatan pemerintah untuk membangun kesehatan ibu Indonesia ke depannya dapat diprioritaskan pada keenam provinsi di Indonesia bagian timur dengan tingkat kesehatan ibu yang rendah guna menjaga sumber daya manusia Indonesia saat ini yang akan menjadi langkah awal untuk menciptakan generasi selanjutnya yang berkualitas demi Indonesia maju, aman, dan tentram.

## V. KESIMPULAN DAN SARAN

### A. Kesimpulan

Berdasarkan hasil analisa data indikator-indikator kesehatan ibu Indonesia tahun 2022, beberapa hal yang menjadi kesimpulan dalam penelitian ini sebagai berikut.

1. *Clustering* provinsi di Indonesia berdasarkan indikator-indikator kesehatan ibu menunjukkan bahwa metode *K-Means* adalah yang terbaik dalam penelitian ini. Dengan nilai *Pseudo-F Statistics* sebesar 10,977 didapatkan bahwa provinsi-provinsi di Indonesia dikelompokkan dalam empat kluster meliputi kluster 0 adalah provinsi dengan tingkat kesehatan ibu paling rendah, kluster 1 adalah provinsi dengan tingkat kesehatan ibu rendah, kluster 2 adalah provinsi dengan tingkat kesehatan ibu sedang, dan kluster 3 adalah provinsi dengan tingkat kesehatan ibu tinggi.
2. Provinsi-provinsi yang berada pada kluster 0 dengan tingkat kesehatan ibu yang sangat rendah meliputi Sulawesi Utara, Sulawesi Tengah, Maluku, Maluku Utara, Papua Barat, dan Papua yang secara geografis keenam provinsi ini berada di wilayah Indonesia Timur. Dengan demikian, program kesehatan pemerintah untuk membangun ibu yang sehat di Indonesia ke depannya dapat diprioritaskan pada keenam provinsi yang kritis ini.

### B. Saran

Berdasarkan hasil analisis, disarankan untuk penelitian selanjutnya supaya dapat memperluas cakupan data dengan memasukkan lebih banyak variabel yang relevan dan menggunakan metode lain dengan evaluasi yang lebih baik dan efektif. Saran untuk pemerintah adalah meningkatkan program kesehatan, upaya sosialisasi, hingga edukasi kesehatan ibu, terutama di provinsi dengan pada kluster dengan tingkat kesehatan paling rendah. Pemerintah dan pihak terkait perlu memperkuat program jaminan kesehatan dan memastikan akses yang merata ke fasilitas kesehatan bagi semua ibu di Indonesia.

## DAFTAR PUSTAKA

- [1] I, Ghozali, "Analisis Multivariat dengan Program IBM SPSS 23," Semarang: Badan Penerbit Universitas Diponegoro, 2016,
- [2] J, Supranto, "Statistik Deskriptif," Jakarta: Airlangga, 1988
- [3] A, Septianingsih, "Pemetaan Kabupaten Kota di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan Agglomeratif Hierarchical Clustering," vol, 3, no, 2, 2022,
- [4] C, G, Victora, dkk, "Maternal and child health in Brazil: progress and challenges," *The Lancet*, 377(9780): 1863-1876, 2012,
- [5] C, R, Titaley, dkk, "Determinants of neonatal mortality in Indonesia," *BMC Public Health*, 8(1): 232, 2010,
- [6] E, P, Wigner, "Theory of traveling-wave optical laser," *Phys, Rev*, vol, 134, pp, A635-A646, 1965,
- [7] G, L, Darmstadt, dkk, "Evidence-based, cost-effective interventions: how many newborn babies can we save? *The Lancet*," 365(9463): 977-988, 2005,
- [8] H, Annur, "Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering (Studi Kasus Toko Luxor Variasi Gorontalo)," 2019,
- [9] Nooraeni, R, "Metode Cluster Menggunakan Kombinasi Algoritma Cluster K-Prototype Dan Algoritma Genetika Untuk Data Bertipe Campuran," *Jurnal Aplikasi Statistika & Komputasi Statistik*, 7(2): 81-98, 2016,
- [10] Novidianto, R, dan Kartika, F, "Algoritma ClusterMix K-Prototypes Untuk Menangkap Karakteristik Pasien Berdasarkan Variabel Penciri Mortalitas Pasien Dengan Gagal Jantung," *INFERENSI*, 4(1): 37-46, 2021,
- [11] R, A, Johnson dan D, W, Wichern, "Applied multivariate statistical analysis," vol, 5, no, 8, Prentice hall Upper Saddle River, NJ, 2002,
- [12] R, Wijayanti, dan D, R, S, Saputro, "Clustering Data Campuran Numerik dan Kategorik Menggunakan Algoritma K-Prototype," *Peosiding Seminar Nasional Matematika*, 6: 702-706, 2023,
- [13] A, Quraisy, dkk, "Analisis Kruskal-Wallis Terhadap Kemampuan Numerik Siswa," *Journal of Statistics and Its Application on Teaching and Research*, 3(3): 156-161, 2021,
- [14] Lukiasuti, dkk, "Statistika Non Parametris Aplikasinya dalam Bidang Ekonomi dan Bisnis," Yogyakarta, 2012,
- [15] V, N, Maurya, dkk, "A Survey Report On Nonparametric Hypothesis Testing Including Kruskal-Wallis, ANOVA, and Kolmogorov-Smirnov Goodness Fit Test," *International Journal of Information*, 2013,

- [16] Subhan, A, dkk, 2022, Clustering Item Fast Moving Dan Slow Moving Pada Produk Unilever Menggunakan Algoritma K-Prototype (Studi Kasus: Yogya Purwakarta), *Jurnal Mahasiswa Teknik Informatika*, 6(2): 629-634,
- [17] Badan Pusat Statistik, "Profil Kesehatan Ibu dan Anak 2022," 2022,
- [18] Badan Pembangunan Nasional, "Laporan Pembangunan Berkelanjutan Indonesia," 2019,
- [19] Kementerian Kesehatan, "Profil Kesehatan Indonesia 2021," 2021,
- [20] UNICEF, "Laporan Tahunan UNICEF Indonesia 2020," 2020,
- [21] World Health Organization, "Sustainable Development Goals (SDGs)," 2015,
- [22] Bank Dunia, "Indonesia Economic Quarterly: Towards Inclusive Growth," 2018,
- [23] I, N, L, Fitriana, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Keluarga Sehat Menggunakan Metode Kluster Hirarki dan Non Hirarki," *Jurnal Paradigma: Jurnal Multiidipliner Mahasiswa Pascasarjana Indonesia*, 2(1): 27-36, 2021.



© 2022 by the authors, This work is licensed under a Creative Commons Attribution-ShareAlike 4,0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>),