

NATIONAL STATISTICS CHALLENGE 2023

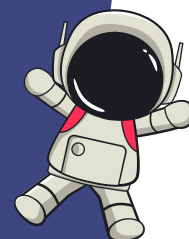


ANALISIS DATA



NOMOR PESERTA

NSC23A008



NAMA ANGGOTA

Kamal Muftie Yafi
Nasywa Safira Ardanty

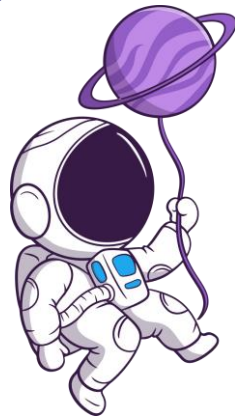


ASAL INSTANSI

Universitas Indonesia



NATIONAL STATISTICS CHALLENGE



ANALISIS UMUM

Data terdiri dari faktor-faktor yang mempengaruhi angka *stunting*. Adapun variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

Xa1 : 0 bulan

Xb1 : 0-5 bulan

Xc1 : 6-11 bulan

Xd1 : 12-23 bulan,

Xe1 : 24-35 bulan,

Xf1 : 36-47 bulan,

Xg1 = 48-59 bulan

X2 : Presentase Bayi yang Menerima ASI Eksklusif (%)

X3 : Rata-Rata Kalori yang Dikonsumsi per Kapita (Kkal)

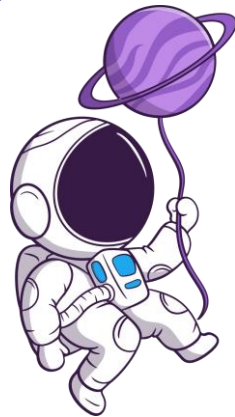
X4 : Rata-Rata Protein yang Dikonsumsi per Kapita (Gram)

X5 : Presentase Rumah Tangga Memanfaatkan Air Bekas untuk Keperluan Lain (%)

X6 : Presentase Gizi Kurang Balita (%)

X7 : Rata-Rata Umur Ibu Ketika Kehamilan Pertama (Tahun)

NATIONAL STATISTICS CHALLENGE



ANALISIS UMUM

Tujuan dari studi kasus ini adalah:

- Mengelompokkan provinsi menjadi dua kelompok, yaitu kelompok dengan faktor stunting yang tinggi dan rendah
- Mengetahui faktor yang menjadi pembeda antar kelompok

NATIONAL STATISTICS CHALLENGE



SOFTWARE



JAWABAN DAN PENYELESAIAN

1. Metode yang paling tepat untuk mengelompokkan provinsi berdasarkan persentase bayi *stunting* adalah analisis klaster (clustering analysis). Provinsi-provinsi pada data akan dikelompokkan menjadi dua kelompok, yaitu kelompok provinsi dengan angka *stunting* tinggi dan rendah.
2. Asumsi yang diperlukan untuk analisis klaster adalah:
 - a) Asumsi Independensi Klaster
 - b) Asumsi Skala Data
 - c) Asumsi Jumlah Klaster
 - d) Asumsi Multikolinearitas

```
> ols_vif_tol(model)
  Variables  Tolerance    VIF
1      Xa1 0.60068390  1.664769
2      Xb1 0.30276538  3.302888
3      Xc1 0.18070787  5.533793
4      Xd1 0.11340851  8.817680
5      Xe1 0.10266435  9.740479
6      Xf1 0.06717318 14.886894
7      Xg1 0.05338847 18.730636
```

Multikolinearitas terjadi ketika dalam sebuah analisis data, dua atau lebih variabel tampaknya sangat berkorelasi atau saling terkait satu sama lain. Variabel yang memiliki nilai vif lebih dari 10 dapat dikatakan multikolinearitas.

JAWABAN DAN PENYELESAIAN

3. Langkah analisis data:

a. Preprocessing

Pengubahan nama-nama kolom yang ada menjadi Xa1, Xb1, Xc1, Xd1, Xe1, Xf1, Xg1, X2, X3, X4, X5, X6, X7. Kolom provinsi tetap menjadi kolom provinsi.

b. Analisis Data

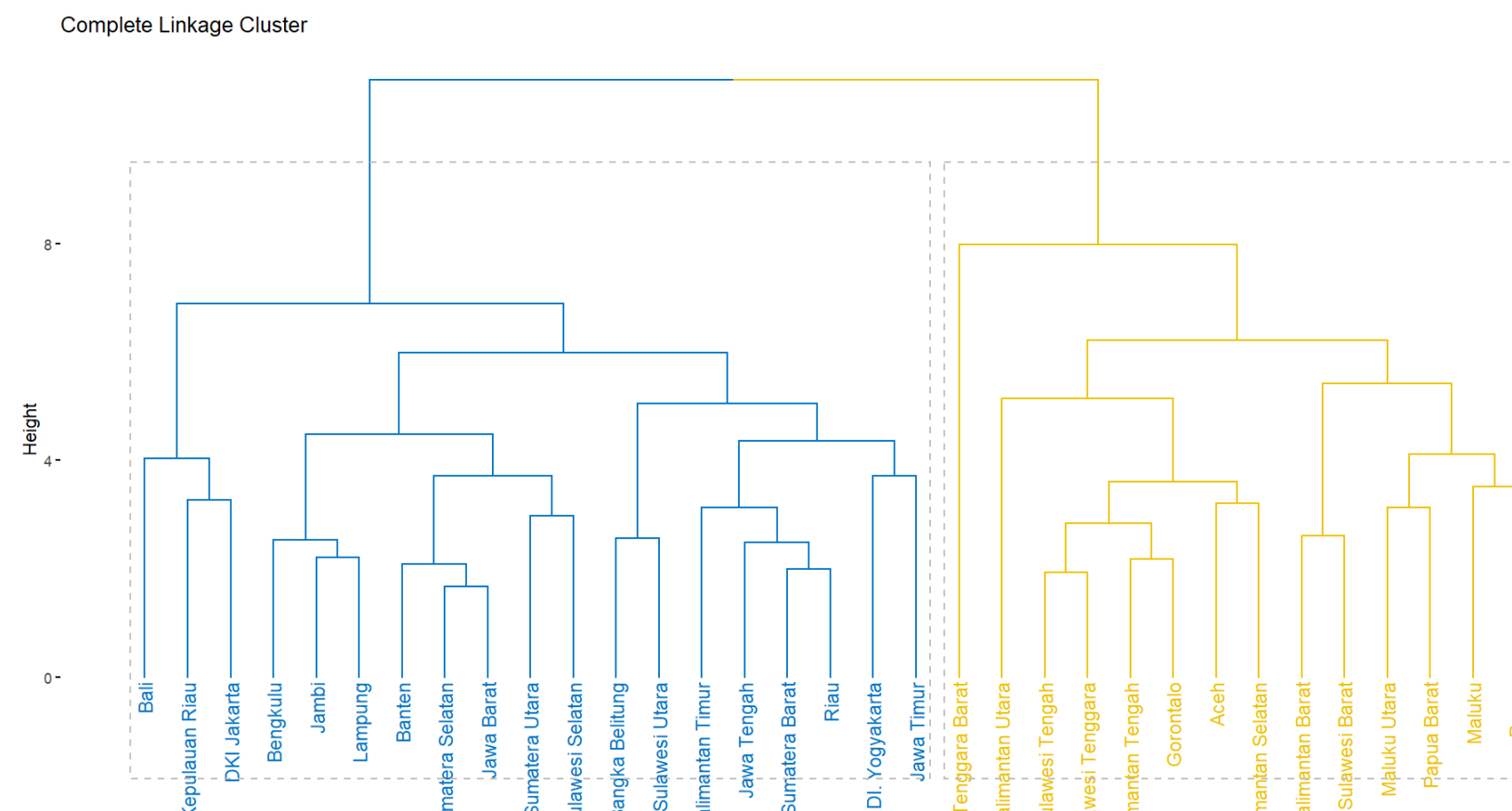
- Penentuan metode *hierarchical clustering* dengan cara pengelompokan *complete*
- Penentuan jumlah anggota kluster, yaitu 2 yang menyatakan faktor stunting tinggi dan rendah
- Melakukan standardisasi data
- Menghitung *distance* antarvariabel

JAWABAN DAN PENYELESAIAN

3. Langkah analisis data:

b. Analisis Data

- Diperoleh clustering sebagai berikut:



Klaster yang berwarna biru (dengan label '2') adalah klaster dengan provinsi yang memiliki persentase angka stunting rendah. Sementara, klaster yang berwarna kuning (dengan label '1') adalah klaster dengan provinsi yang memiliki persentase angka stunting tinggi.

JAWABAN DAN PENYELESAIAN

3. Langkah analisis data:

c. Uji kecocokan model

Setelah diuji, didapatkan jumlah klaster terbaik adalah 2 dan 4. Maka, pemilihan 2 klaster sudah cukup baik.

Optimal Scores:

	Score	Method	Clusters
Connectivity	9.9183	agnes	2
Dunn	0.4748	agnes	4
Silhouette	0.2357	agnes	2

```
> optimalScores(stability)
```

	Score	Method	Clusters
APN	0.08784198	agnes	2
AD	3.56044706	agnes	4
ADM	0.53293222	agnes	2
FOM	0.83570745	agnes	4

JAWABAN DAN PENYELESAIAN

4. Untuk mengetahui faktor yang menjadi pembeda antar kelompok, metode yang paling tepat digunakan adalah analisis diskriminan. Analisis diskriminan adalah teknik statistik yang memungkinkan peneliti untuk memahami kontribusi variabel-variabel independen dalam membedakan atau mengklasifikasikan observasi ke dalam kelompok-kelompok yang berbeda. Dalam konteks data, peneliti dapat mengidentifikasi faktor-faktor yang mempengaruhi kelompok angka stunting tinggi dan rendah.

JAWABAN DAN PENYELESAIAN

5. Asumsi yang diperlukan dalam analisis diskriminan:

Asumsi Normalitas

Analisis diskriminan mengasumsikan bahwa variabel-variabel independen (prediktor) dalam setiap kelompok atau kategori mengikuti distribusi normal. Akan diuji dengan Shapiro test. Hasilnya, seluruh variabel independent berdistribusi normal, kecuali persentase rumah tangga memanfaatkan air bekas untuk keperluan lain dan rata-rata umur ibu ketika kehamilan pertama.

JAWABAN DAN PENYELESAIAN

5. Pengujian asumsi normalitas

H_0 : variabel berdistribusi normal

H_1 : variabel tidak berdistribusi normal

Dipilih $\alpha = 0.05$. H_0 akan ditolak apabila $p - value < \alpha$

```
> shapiro.test(nsc$X2)
```

Shapiro-Wilk normality test

```
data: nsc$X2
W = 0.9742, p-value = 0.6041
```

```
> shapiro.test(nsc$X3)
```

Shapiro-Wilk normality test

```
data: nsc$X3
W = 0.94056, p-value = 0.07051
```

```
> shapiro.test(nsc$X4)
```

Shapiro-Wilk normality test

```
data: nsc$X4
W = 0.97483, p-value = 0.6237
```

```
> shapiro.test(nsc$X5)
```

Shapiro-Wilk normality test

```
data: nsc$X5
W = 0.89475, p-value = 0.003904
```

```
> shapiro.test(nsc$X6)
```

Shapiro-Wilk normality test

```
data: nsc$X6
W = 0.94346, p-value = 0.08568
```

```
> shapiro.test(nsc$X7)
```

Shapiro-Wilk normality test

```
data: nsc$X7
W = 0.89913, p-value = 0.005057
```

Variabel yang tidak berdistribusi normal: X5 dan X7

JAWABAN DAN PENYELESAIAN

6. Analisis dan Interpretasi:

a) Pembagian data

Data dibagi menjadi data 'train' sebanyak 70% dari total data dan data 'test' sebanyak 30% dari total data. Data 'train' digunakan untuk membangun model prediksi, sementara data 'test' digunakan untuk menguji model. Tujuan pembagian data ini adalah untuk mengukur kinerja model secara obyektif dan menghindari masalah overfitting

b) *Modelling*

Model yang dihasilkan adalah sebagai berikut

$$Y = 0.036X_2 + 0.002X_3 - 0.012X_4 - 0.171X_5 - 0.551X_6 + 1.365X_7$$

Dari model, terlihat bahwa persentase bayi yang menerima ASI eksklusif (X_2), Presentase Rumah Tangga Memanfaatkan Air Bekas untuk Keperluan Lain (X_5), Presentase Gizi Kurang Balita (X_6), dan Rata-Rata Umur Ibu Ketika Kehamilan Pertama (X_7) berkontribusi besar dalam pengelompokkan provinsi.

JAWABAN DAN PENYELESAIAN

7. Kesimpulan:


Provinsi-provinsi di Indonesia dapat dikelompokkan menjadi 2 dengan menggunakan complete hierarchical clustering berdasarkan persentase stunting pada bayi. Dua kelompok ini adalah kelompok dengan angka stunting tinggi dan rendah. Untuk kelompok yang memiliki angka stunting yang tinggi dapat diberi bantuan.

JAWABAN DAN PENYELESAIAN


8. Rekomendasi:

Provinsi Jawa Timur masuk ke dalam klaster provinsi dengan angka stunting rendah. Saran atau rekomendasi dari kami untuk Provinsi Jawa Timur adalah mengoptimalkan, mempertahankan, bahkan meningkatkan kualitas faktor-faktor yang mempengaruhi stunting. Terlihat pada model bahwa persentase bayi yang menerima ASI eksklusif berkontribusi tinggi pada pengelompokan provinsi. Maka, perlu ditingkatkan penyuluhan pada ibu dan calon ibu untuk mengoptimalkan pemberian ASI eksklusif. Selain itu, perlu juga penyuluhan untuk seluruh warga agar dapat menciptakan lingkungan yang aman bagi ibu hamil dan baru melahirkan karena lingkungan dapat mempengaruhi kualitas dan kuantitas ASI pada ibu. Selain itu, Rata-Rata Umur Ibu Ketika Kehamilan Pertama (Tahun) juga berkontribusi besar. Maka perlu diadakan aturan yang ketat dalam pernikahan diri. Tujuannya agar umur ibu ketika kehamilan pertama tidak terlalu muda (cukup umur).

KESIMPULAN



Dengan analisis klaster, provinsi di Indonesia dapat dibagi menjadi 2 berdasarkan persentase *stunting* pada bayi dari tiap rentang usia. Dapat diketahui pula kontribusi faktor-faktor yang mendasari angka *stunting*. Dari model, diketahui bahwa kontribusi faktor terbesar adalah Rata-Rata Umur Ibu Ketika Kehamilan Pertama. Jadi, diperlukan penyuluhan tentang bahaya kehamilan dini dan risikonya terhadap bayi *stunting* kepada warga negara Indonesia.



THANK YOU

