

Nama : Ahmad Nabilul As'ad

NIM : A11.2022.14488

1. Pemahaman Proses Bisnis (Business Understanding)

- **Masalah:** Mengidentifikasi faktor-faktor yang mempengaruhi status gizi balita.
- **Tujuan:** Membangun model prediktif untuk memprediksi status gizi balita berdasarkan usia, jenis kelamin, dan tinggi badan.

2. Pemahaman Data (Data Understanding)

Dataset ini siap untuk analisis lebih lanjut karena tidak ada missing values dan distribusi data yang tampaknya wajar.

3. Data Preparation

Dataset : data_balita.csv

Setelah memahami data, kita dapat melakukan pembersihan data dan menyiapkan data untuk pemodelan.

Data balita ini tidak memiliki nilai yang hilang, dan distribusi dari beberapa fitur utama seperti berikut:

- **Umur (bulan):** Rentang dari 0 hingga 60 bulan (5 tahun).
- **Tinggi Badan (cm):** Rentang tinggi badan dari 40.01 cm hingga 128 cm.
- **Jenis Kelamin:** Terdapat dua kategori, yaitu "laki-laki" dan "perempuan."
- **Status Gizi:** Kategori status gizi terdiri dari:
 - *stunted* (pendek)
 - *tinggi*
 - *normal*
 - *severely stunted* (sangat pendek).

4. Pemodelan (Modelling)

```

Hasil
(Umur (bulan)          0
Jenis Kelamin          0
Tinggi Badan (cm)     0
Status Gizi            0
dtype: int64,
      Umur (bulan)  Tinggi Badan (cm)
count  120999.000000    120999.000000
mean     30.173803       88.655434
std      17.575119       17.300997
min        0.000000       40.010437
25%      15.000000       77.000000
50%      30.000000       89.800000
75%      45.000000      101.200000
max       60.000000      128.000000,
array(['laki-laki', 'perempuan'], dtype=object),
array(['stunted', 'tinggi', 'normal', 'severely stunted'],

```

Saya akan memulai dengan encoding kolom **Jenis Kelamin** dan **Status Gizi** serta menyiapkan data untuk pelatihan model.

Data telah siap untuk pemodelan, dengan kolom **Jenis Kelamin** dan **Status Gizi** yang sudah di-encode. Berikut rincian encoding:

- **Jenis Kelamin:** 0 untuk laki-laki, 1 untuk perempuan.
- **Status Gizi:**
 - 0: *normal*
 - 1: *severely stunted*
 - 2: *stunted*
 - 3: *tinggi*

Langkah berikutnya adalah melatih model untuk memprediksi status gizi. Saya akan menggunakan algoritma seperti **Decision Tree** untuk membangun model dan mengevaluasi kinerjanya.

Hasil

```
(      Umur (bulan)  Jenis Kelamin  Tinggi Badan (cm)
110274           55           1         92.800000
21987            11           0         82.700000
69900            35           0         91.000000
1385              0           1         40.029406
42463            21           1         80.400000,
110274      1
21987       3
69900       0
1385        1
42463       0
Name: Status Gizi, dtype: int64)
```