Stunting adalah kondisi ketika anak lebih pendek dibandingkan anak-anak lain seusianya, atau dengan kata lain, tinggi badan anak berada di bawah standar. Standar yang dipakai sebagai acuan adalah kurva pertumbuhan yang dibuat oleh Badan Kesehatan Dunia (WHO).

### **Penyebab Stunting** **pada Anak**

Stunting terjadi karena kurangnya asupan gizi pada anak dalam [1000 hari pertama](https://www.alodokter.com/cara-merawat-bayi-di-1000-hari-pertama" \t "_blank) kehidupan, yaitu semenjak anak masih [di dalam kandungan](https://www.alodokter.com/cara-mencegah-stunting-pada-anak-sejak-masa-kehamilan" \t "_blank) hingga anak berusia 2 tahun. Salah satu penyebabnya adalah [kurangnya asupan protein](https://www.alodokter.com/mengenal-protein-dan-dampak-kekurangan-protein-bagi-tubuh" \t "_blank).

Stunting pada anak bisa disebabkan oleh masalah pada saat kehamilan, melahirkan, menyusui, atau setelahnya, seperti [pemberian MPASI](https://www.alodokter.com/informasi-penting-tentang-pemberian-mpasi-pada-bayi" \t "_blank) yang tidak mencukupi asupan nutrisi.

Selain nutrisi yang buruk, stunting juga bisa disebabkan oleh kebersihan lingkungan yang buruk, sehingga anak sering terkena infeksi. Pola asuh yang kurang baik juga ikut berkontribusi atas terjadinya stunting. Buruknya pola asuh orang tua sering kali disebabkan oleh kondisi ibu yang masih terlalu muda, atau jarak antar kehamilan terlalu dekat.

1. NumPy

Library yang pertama ini merupakan singkatan dari Numerical Python. Fungsi dari modul ini untuk membantu para data scientist menangani permasalahan angka-angka. kemampuan dalam membentuk objek N-dimensional array yang mirip dengan MATLAB. NumPy memudahkan penggunanya dalam operasi Aljabar Linear seperti Vektor dan Matriks. terutama operasi matematika

### 2.  Pandas

Library berdasarkan sistem dataframe, modul ini dapat memuat sebuah file ke dalam tabel virtual menyerupai spreadsheet. Pandas juga berfungsi mengolah suatu data seperti teknik join, distinct, group by, agregasi, dan teknik lainnya seperti pada SQL. Bedanya, ini dilakukan pada tabel. Kelebihan dari library ini juga dapat membaca file dari berbagai format seperti .txt, .csv, dan .tsv. Penanganan, manipulasi, dan analisis data

3. Matplotlib

Library Matplotlib ini akan membantu dalam menampilkan hasil analisis berupa visual grafik berwarna dengan lebih rapi dan menarik. Ada dua jenis plot dalam menampilkan data yaitu secara 2D dan 3D. Sehingga data bisa ditampilkan sesuai dengan kebutuhan. Matplotlib ini merupakan library yang paling sering digunakan oleh data science dalam menyajikan datanya ke dalam visual yang lebih menarik.

4. Seaborn

adalah library untuk membuat grafik dan statistik dengan menggunakan Python. Library ini di bangun berdasarkan library matplotlib

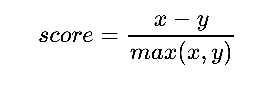
5. sklearn

Scikit Learn difokuskan pada Machine Learning

Cara yang baik untuk mengukur algoritme pengelompokan adalah dengan melihat seberapa baik kluster dipisahkan. **Apakah cluster terpisah dengan baik? Apakah titik data dalam cluster cukup ketat?**

Kami membutuhkan metrik yang dapat mengukur perilaku ini. Kami akan menggunakan metrik yang disebut dengan **skor koefisien siluet** . Skor ini ditentukan untuk setiap titik data;

Fungsi **sklearn.metrics.silhouette\_score** menghitung  
koefisien siluet rata-rata dari semua sampel. Untuk setiap sampel, dua jarak dihitung: rata-rata jarak antar klaster (x), dan rata-rata jarak klaster terdekat (y). Koefisien siluet untuk sampel diberikan oleh persamaan berikut:



Pada dasarnya, y adalah jarak antara sampel dan cluster terdekat yang tidak menyertakan sampel tersebut.

**x** adalah jarak rata-rata antara titik data saat ini dan semua titik data lainnya dalam cluster yang sama

**y** adalah jarak rata-rata antara titik data saat ini dan semua titik data  
di cluster terdekat berikutnya.

## **Masih ada lagi…**

Nilai terbaik adalah 1, dan nilai terburuk adalah -1. 0 mewakili cluster yang tumpang tindih, sedangkan nilai kurang dari 0 berarti sampel tertentu telah dilampirkan ke cluster yang salah.

Perintah berikut untuk menyatakan bahwa jika cell tersebut adalah nilai hilang makan bernilai

df.isnull()

## Menghilang sample dimana ada fitur yang nilainya hilang (missing values)

Salah satu strategi sederhananya adalah menghilangkan sample ( baris dalam tabel) atau fitur ( kolom dari tabel) dimana terdapat nilai hilang (missing values) didasarkan beberapa kriteria

# menghilangkan sample atau baris

df.dropna()

# Menghilang fitur atau kolom

df.dropna(axis=1)

# hanya menghilang baris dimana semua kolom/fitur adalah NaN

df.dropna(how='all')

# Menhgilangkan baris yang memiliki paling sedikit mangandung 4 fitur yang tidak NaN

df.dropna(thresh=4)

# hanya menghilangkan baris dimana NaN muncul pada kolom tertentu (disini kolom : 'C')

df.dropna(subset=['C'])

Menghilangkan data mungkin tidak diharapkan karena data akan menjadi lebih sedikit, oleh karena itu perlu penyelesaian data hilang dengan cara mengimputkan data. Teknik ini disebut dengan teknik inputer. Salah satunya dengan menggantikan dengan nilai statistik

from sklearn.preprocessing import Imputer

# pilihan dari library imputer terdiri dari niai mean, median, modus ( nilai yang paling dering/most\_frequent)

imr = Imputer(missing\_values='NaN', strategy='mean', axis=0)

imr = imr.fit(df)

imputed\_data = imr.transform(df.values)

imputed\_data

displot digunakan untuk membuat histogram.

Boxplot adalah salah satu cara dalam statistik deskriptif untuk menggambarkan secara grafik dari data numeris melalui lima ukuran sebagai berikut:

* Nilai observasi terkecil,
* Kuartil terendah atau kuartil pertama (Q1), yang memotong 25% dari data terendah
* Median (Q2) atau nilai pertengahan,
* Kuartil tertinggi atau kuartil ketiga (Q3), yang memotong 25% dari data terbesar
* Niilai observasi terbesar.

Untuk menggunakan **scaterplot()** pertama kita harus cari korelasi antar atribut data dengan fungsi **corr()**

dari hasil korelasi kita dapat membuat **scaterplot()** untuk melihat data lebih dalam misalnya korelasi bisa dilihat dari perbedaan atau kedalaman warna

untuk dataset kita gunakan dataset tips

## **K-Means Clustering**

K-Means Clustering adalah salah satu algoritma dalam menentukan klasifikasi terhadap objek berdasarkan attribut / fitur dari objek tersebut kedalam **K** kluster/partisi. K adalah angka positif yang menyatakan jumlah grup/kluster/partisi terhadap objek. Pemartisian data dilakukan dengan mencari nilai jarak minimum antara data dan nilai ***centroid*** yang telah di set baik secara random atau pun dengan***Initial Set of Centroids,***kita juga dapat menentukan nilai centroid berdasarkan **K object** yang berurutan.

[***Centroid***](https://en.wikipedia.org/wiki/Centroid)adalah nilai rata-rata aritmetik dari sebuah bentuk objek dari seluruh titik dalam objek tersebut. Penerapan K-Means Clustering ini dapat dilakukan dengan prosedur step by step berikut :

 Siapkan data training berbentuk vector.

 Set nilai K cluster.

 Set nilai awal centroids.

 Hitung jarak antara data dan centroid menggunakan rumus [***Euclidean Distance***](https://en.wikipedia.org/wiki/Euclidean_distance).

 Partisi data berdasarkan nilai minimum.

 Kemudian lakukan iterasi selama partisi data masih bergerak (tidak ada lagi objek yang bergerak ke partisi lain), bila masih maka ke poin 3.

 Bila grup data sekarang sama dengan grup data sebelumnya, maka hentikan iterasi.

 Data telah dipartisi sesuai nilai centroid akhir.