**LAPORAN**

**CASEBASE PEMBELAJARAN MESIN ALGORITMA UNSUPERVISED LEARNING DENGAN MENGGUNAKAN MODEL KMEANS CLUSTERING**

**2021/2022**



**Disusun oleh :**

Muhammad Rafli Ramadhan (1301200204)

**(IF 44 04)**

**S1 INFORMATIKA**

**FAKULTAS INFORMATIKA**

**BAB I**

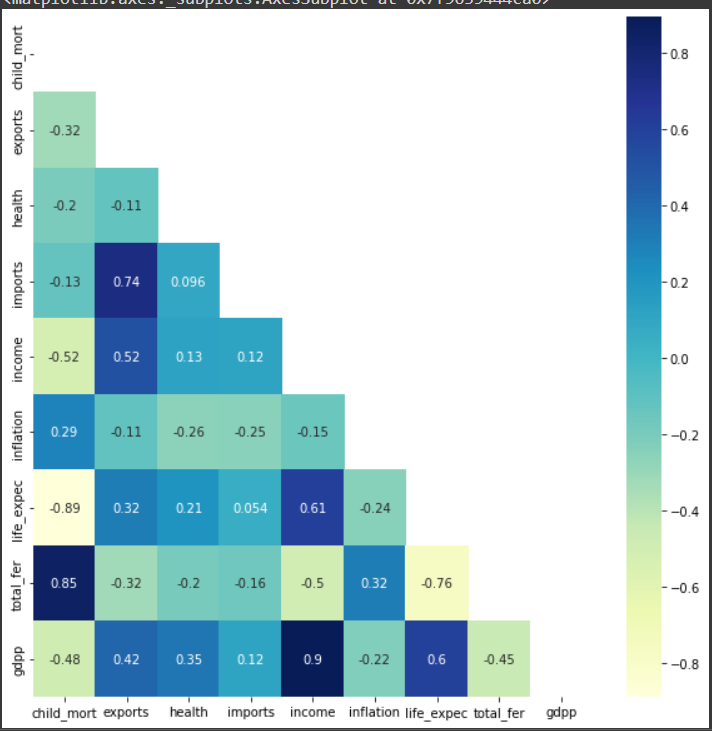
**IKHTISAR KUMPULAN DATA YANG DIPILIH**

# 1.1 Dataset

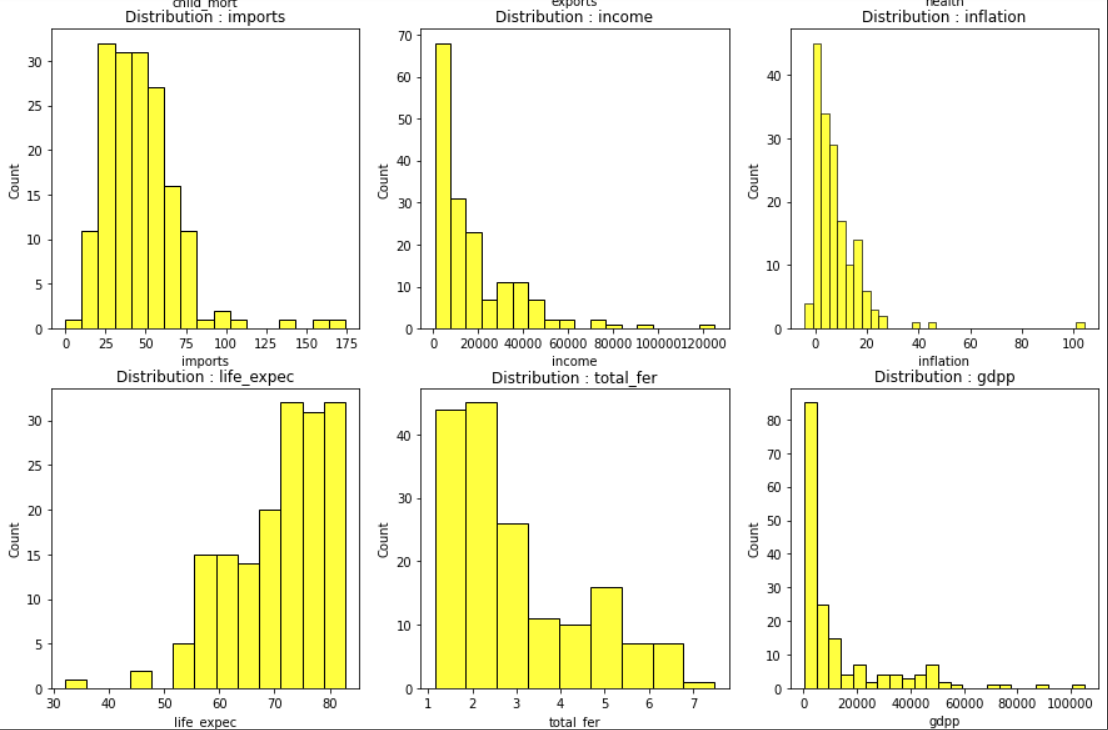
Data yang diambil adalah data Country Data, yaitu data pada negara. Datanya meiliki banyak kolom 10 dan record datanya adalah 167. Dataset ini sudah lumayan bagus, deskripsi pada dataset sudah tersedia dan data kosongnya pun tidak ada.

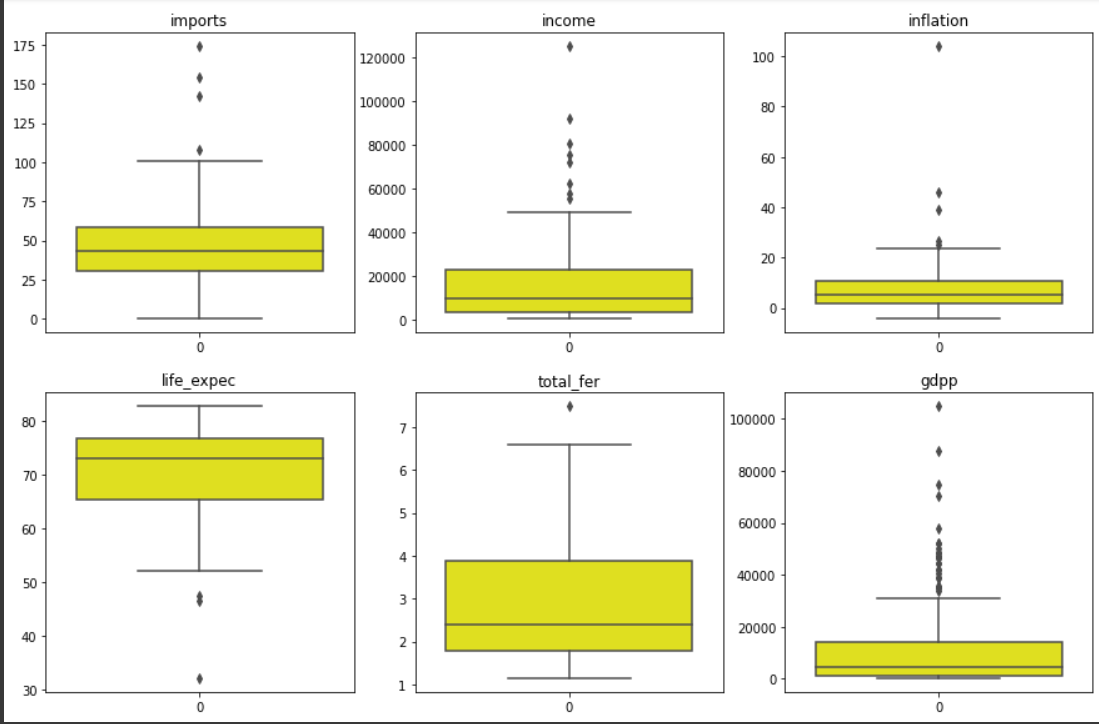
# 1.2 Pra-Processing Data

Pada unsupervised learning, pra-processing data yang cocok adalah pendekatan Korelasi, Histogram dan box plot.

  
diagram korelasi

Data korelasi paling terbesar pada countrydata adalah income dan gdpp, terbesar kedua adalah total\_fer dan child mort. Korelasi tersebut bisa dijadikan acuan untuk kasus unsupervised learning.

  
Diagram Histogram

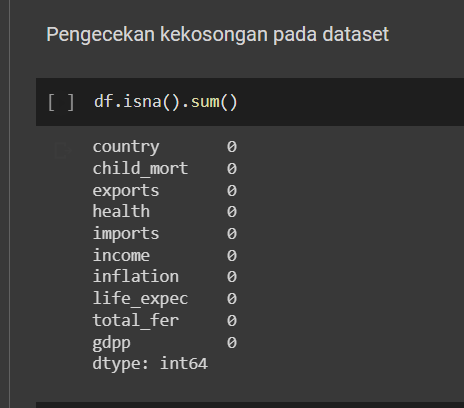
  
Diagram Boxplot

**BAB II**

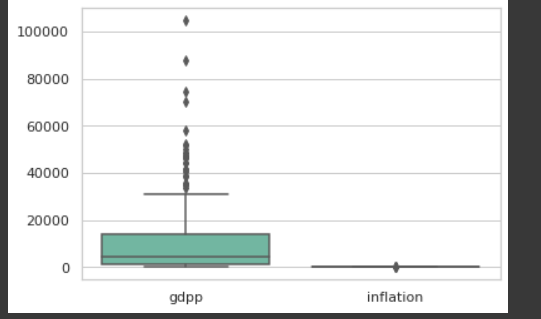
**RINGKASAN PRA-PEMPROSESAN DATA YANG DIUSULKAN**

**2.1 Data Cleaning**

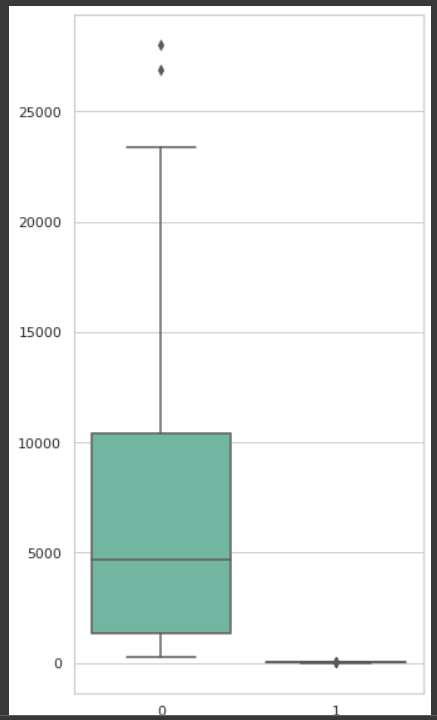
Data yang ingin dibahas pada kasus ini adalah pendapatan perkapita (gdpp) dan inflasi, maka data cleaning akan dilakukan pada atribut yang ingin diolah



Data sudah clean, tetapi kasus ini, memerlukan data cleaning pada atribut gdpp dan inflasi. Karena jika dibandingkan dengan boxplot, datanya sangat timpang sekali

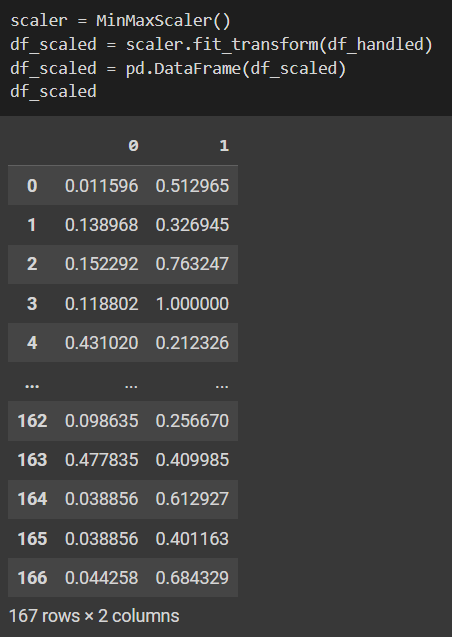


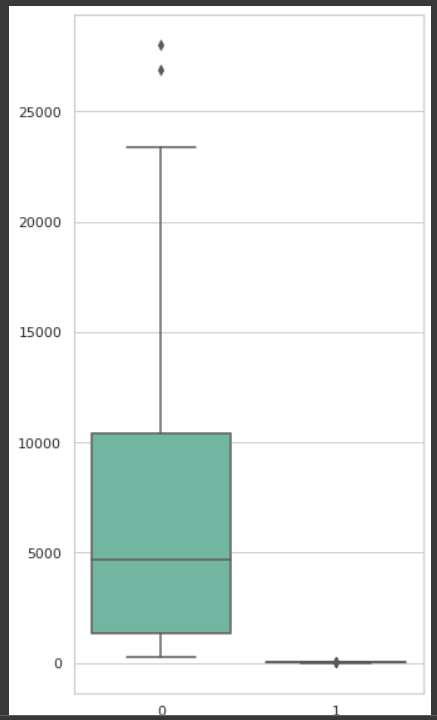
Maka, dilakukanlah reduksi pada outliners, jadilah seperti ini. Walaupun outliners udah bersih, tetapi data masih terlihat sangat timpang, maka normalisasi data disini adalah hal penting yang harus dilakukan



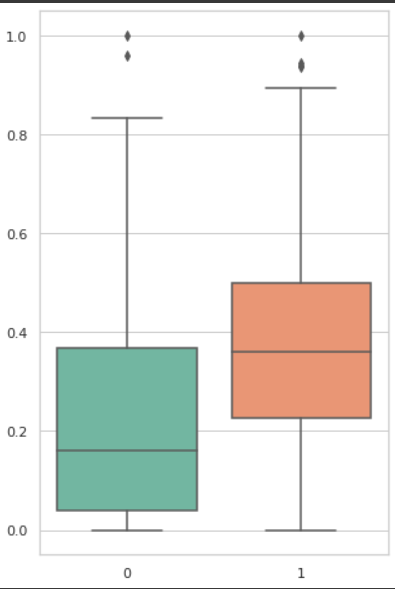
**2.2 Normalisasi Data**

Ketimpangan data adalah suatu masalah, maka normalisasi diperlukan untuk mengatur max dan minimum agar tidak terlalu berjauhan pada data. Pada kasus ini akan menggunakan minmax scaller





Sebelum Menggunakan normalisasi



Sesudah menggunakan normalisasi

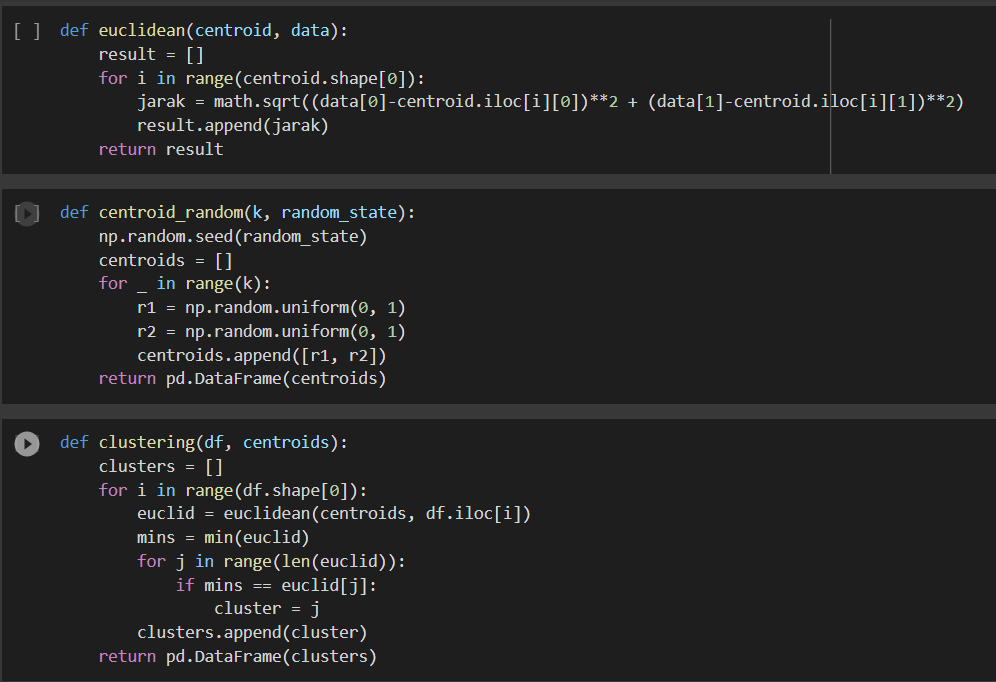
Perbedaanya sangat segnifikan, data terlihat cocok dan bagus pada kasus kali ini

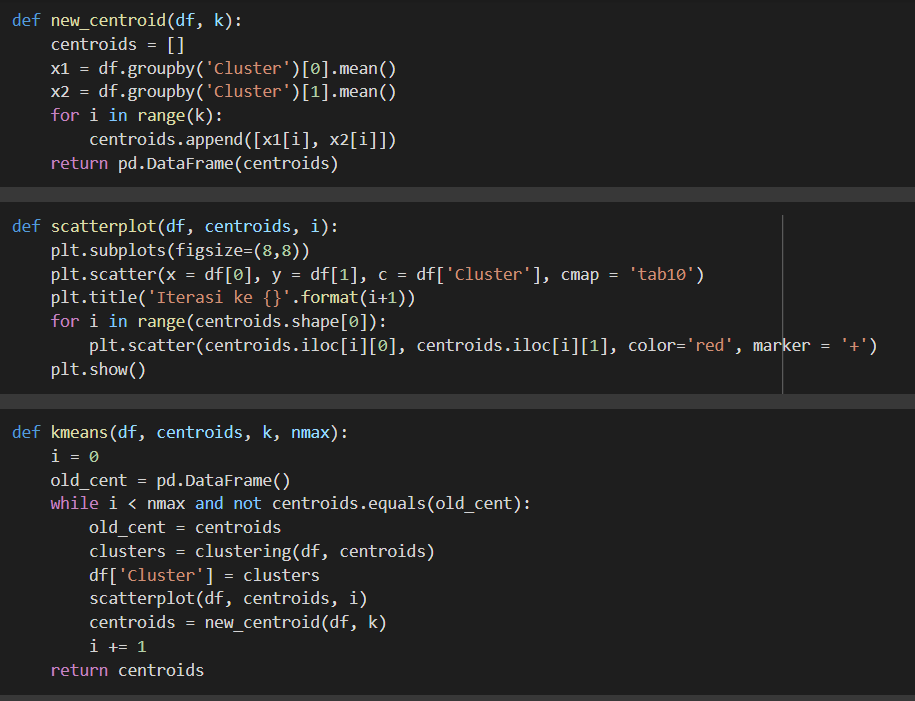
**BAB III**

**MENERAPKAN ALGORITMA YANG DI PILIH**

**3.1 Algoritma Kmeans**

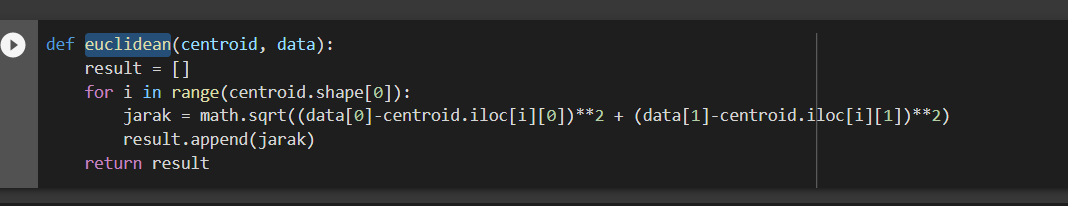
Algoritma Kmeans adalah algoritma yang lumayan popular untuk pembuatan clustering karena sifatnya trainingnya yang sangat bagus digabungkan dengan rumus 1





**3.2 Algoritma Eucledian**

Ini adalah algoritma untuk menghitung jarak pada titik dengan titik lainnya dengan rumus eucledian, rumus ini akan digunakan pada penghitungan jarak dengan kmeans

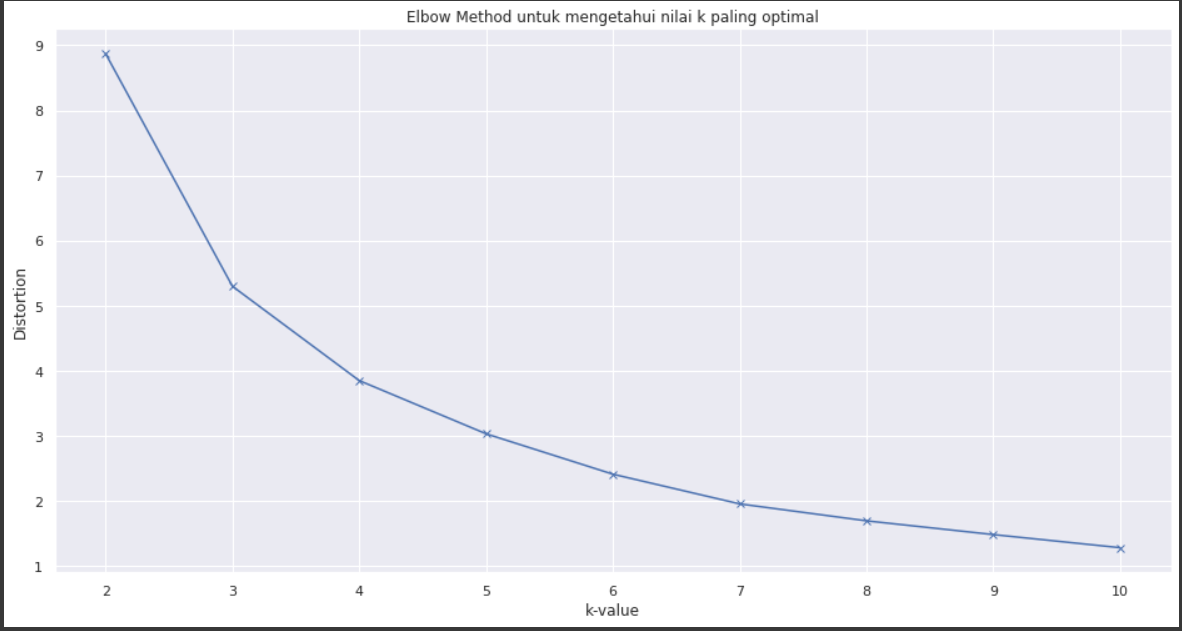


**BAB IV**

**EVALUASI HASIL**

**4.1 Elbow Method**

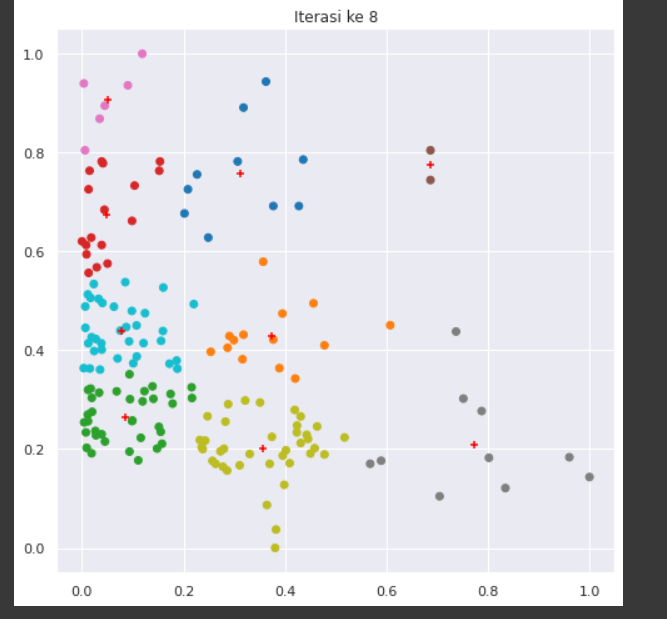
Ini adalah method yang digunakan untuk mendapatkan k terbaik pada data yang telah diberikan. Elbow digunakan untuk mendapatkan k pada kmeans



**4.2 Clustering**

****

Ini adalah hasil dari centroid

****Dan ini adalah hasil pada iterasi ke 8 pada kmeans

**BAB V**

**LINK & PRESENTASI VIDEO**

**5.1 Link Presentasi**

Video https://drive.google.com/file/d/1i1D1chywLl49uodBkRbApedJp-BwS3oN/view?usp=sharing 5.2

Link Colab https://colab.research.google.com/drive/1TyO4qFW8zSNmMsWuKdfNy1vELY\_CqTTl?usp =sharing 5.3

Link Github https://github.com/raflidev/Machine-Learning-Case-Base-1