

Arya Prima Al AUFAR

Final Project Sanbercode

## **Formulasi Masalah**

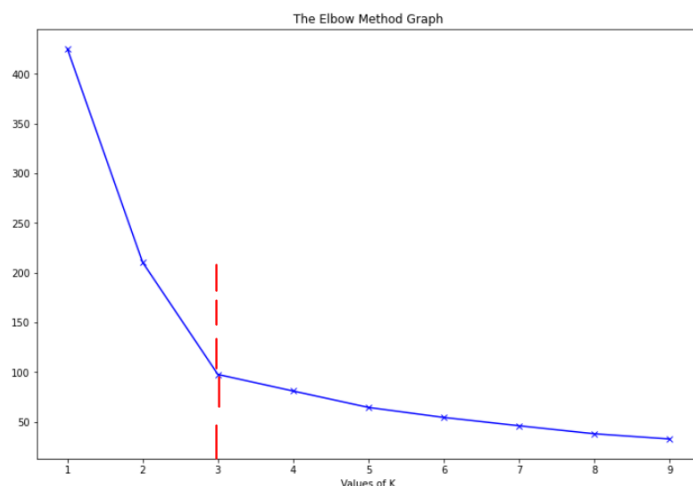
Dalam kasus ini, CEO LSM perlu memutuskan bagaimana menggunakan uang secara strategis dan efektif dari hasil kumpulan dari HELP International, dan tugas saya adalah untuk mengkategorikan negara menggunakan faktor sosial ekonomi dan Kesehatan yang menentukan perkembangan negara secara keseluruhan, dan juga menyarankan negara mana saja yang perlu menjadi fokus si CEO.

## **Eksplorasi dan Pre-Processing**

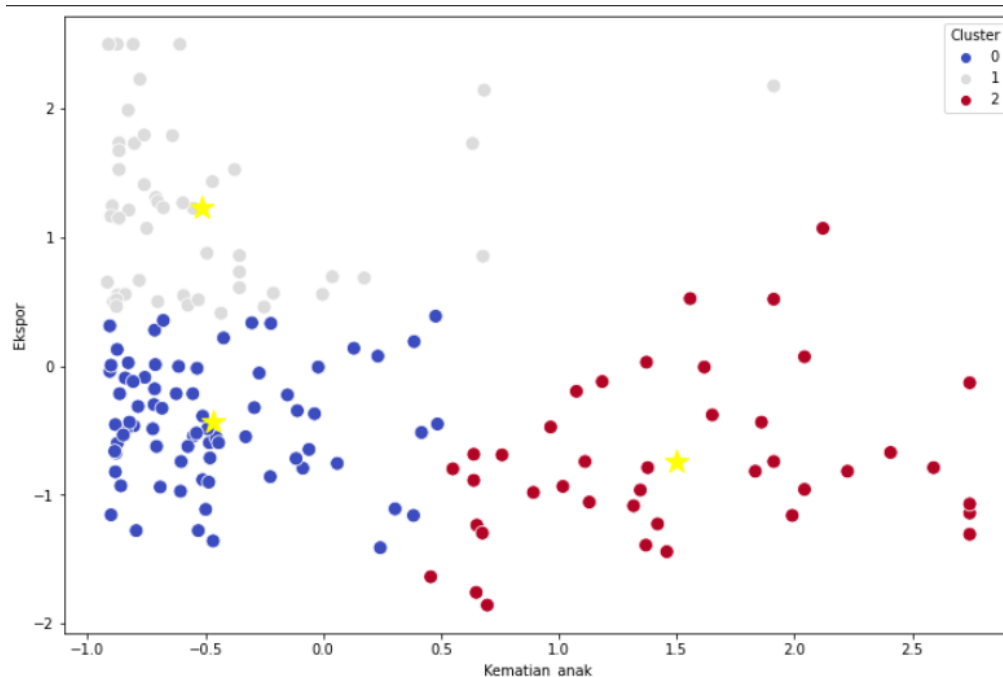
- melakukan `df_train.info()` , digunakan untuk mengecek elemen apa saja yang ada di data
- Lalu, melakukan drop terhadap kolom Negara, karena kolom negara adalah yang ingin kita jadikan target nantinya.
- Setelah itu, membuat Heatmap untuk melihat korelasi antar data
- Setelah itu, melakukan Feature Engineering, untuk mendapatkan kolom feature yang ingin kita uji dalam Clustering ini, dan saya mendapatkan kolom Kematian\_anak dan Ekspor
- Setelah terambil kolom feature, ternyata di kolom Kematian\_anak dan Ekspor terdapat Outliers, sehingga saya harus mengatasinya dengan metode Interquartile sehingga Outliernya hilang.
- Setelah itu melakukan Scalling, scalling yang saya gunakan menggunakan StandardScaler

## **Pemodelan dan Evaluasi**

Setelah data nya melewati tahap eksplorasi dan preprocessing, saya mencari Elbow nya dengan menggunakan library KMeans, dan hasil Elbow nya seperti ini:



Terlihat dari Elbow tersebut, bahwa pada saat  $K = 3$  terjadi lekukan yang cukup signifikan, yang berarti clustering dengan  $K = 3$  merupakan yang optimal berdasarkan Sum Squared Distance (SSE), maka kita akan melakukan modelling dengan  $K = 2$ , maka hasilnya akan seperti berikut:



Analisis : Terlihat bahwa ada 3 Cluster yang sudah di plotkan, yaitu Cluster 0, 1 , dan 2. Dalam kasus ini mengkategorikan negara menggunakan faktor sosial ekonomi dan Kesehatan yang menentukan perkembangan negara secara keseluruhan, berdasarkan data diatas,

- Negara yang berada di Cluster 0, merupakan negara yang memiliki rata-rata Kematan\_anak kecil dan rata-rata Ekspor kecil, namun tidak lebih kecil daripada rata-rata Ekspor Cluster 2.
- Negara yang berada di Cluster 1, merupakan negara yang memiliki rata-rata Kematan\_anak yang kecil dan rata-rata Ekspor terbesar.
- Negara yang berada di Cluster 2, merupakan negara yang memiliki rata-rata kematan\_anak terbesar dan rata-rata ekspor terkecil.

Jadi berdasarkan data dan penjelasan diatas, jika diurut kan berdasarkan sosial Ekonominya maka saya akan membuat peringkat Cluster Negara berdasarkan Sosial Ekonominya mulai dari yang terbaik hingga kurang baik, yaitu :

1. Cluster 1 (Kematan\_anak Terkecil, Ekspor Terbesar)
2. Cluster 0 (Kematan\_anak Normal, Ekspor Normal)
3. Cluster 2 (Kematan\_anak Terbesar, Ekspor Terkecil)

Saran dari saya untuk negara mana saja yang perlu menjadi Fokus CEO adalah, negara yang berada di Cluster 2, dan inilah list negara yang berada di Cluster 2

Hasil Akhir

```
[64] databarin = pd.DataFrame([df_train['Negara'], dftrain_scal['Cluster']])  
      databarin = databarin.transpose()
```

```
[65] databarin[databarin['Cluster'] == 2]['Negara']
```

```
0      Afghanistan  
3      Angola  
17     Benin  
25     Burkina Faso  
26     Burundi  
28     Cameroon  
31  Central African Republic  
32     Chad  
36     Comoros  
37  Congo, Dem. Rep.  
40  Cote d'Ivoire  
50     Eritrea  
56     Gambia  
59     Ghana  
63     Guinea  
64  Guinea-Bissau  
66     Haiti  
69     India  
80     Kenya  
81     Kiribati  
84     Lao  
87     Lesotho  
88     Liberia  
93     Madagascar  
94     Malawi  
97     Mali  
99     Mauritania  
106    Mozambique  
107    Myanmar  
112    Niger  
113    Nigeria  
116    Pakistan  
126    Rwanda  
129    Senegal  
132    Sierra Leone  
142    Sudan  
147    Tanzania  
149    Timor-Leste  
150    Togo  
155    Uganda  
166    Zambia  
Name: Negara, dtype: object
```

## Kesimpulan

Jadi kesimpulannya adalah berdasarkan Feature kolom yang saya pilih yaitu Kematian\_anak (Kesehatan) dan Ekspor(Sosial Ekonomi), dan setelah melalui proses Clustering saya mendapatkan data bahwa dari 3 Cluster yang saya buat, Cluster 1 merupakan Cluster yang terbaik, lalu diikuti Cluster 2, dan untuk yang terakhir adalah Cluster 3, sehingga Negara yang lebih menjadi fokus CEO untuk mendapatkan fasilitas dan bantuan dasar bagi masyarakatnya pada saat terjadi bencana dan bencana alam adalah Negara yang berada di Cluster 2.