

Clustering Negara dengan Menggunakan K-Means untuk HELP International

Untuk mengkategorikan negara menggunakan faktor sosial ekonomi dan kesehatan yang menentukan pembangunan negara secara keseluruhan. Permasalahan yang ada di dalam dataset ini perlu memutuskan bagaimana menggunakan uang ini secara strategis dan efektif. Jadi, harus mengambil keputusan untuk memilih negara yang paling membutuhkan bantuan. Dari permasalahan diatas untuk proses clustrering terdiri dari beberapa tahap:

1. Mengimport data drive lalu membaca data HELP dengan cara sebagai berikut

```
from google.colab import drive
drive.mount('/content/drive')
```

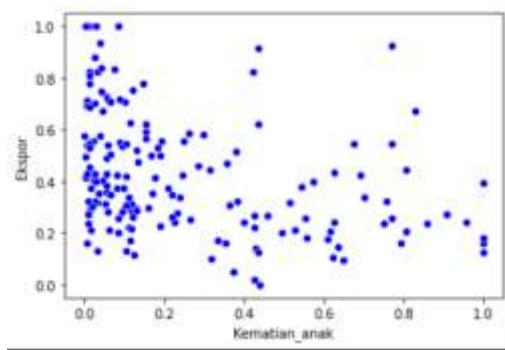
```
df_help = pd.read_csv('/content/drive/MyDrive/SanberCode/Dataset/Data_Negara_HELP.csv')
```

2. Setelah mengimport data hal yang saya lakukan menghapus/mendrop data dari table negara karena menurut saya tidak akan terpakai juga.
3. Setelah itu melihat korelasi data menggunakan headmap
4. Langkah selanjutnya melakukan seleksi fitur menggunakan codingan sebagai berikut

```
columns = np.full((df_help_corr.shape[0],), True, dtype=bool)
for i in range(df_help_corr.shape[0]):
    for j in range(i+1, df_help_corr.shape[0]):
        if df_help_corr.iloc[i,j] >= -0.2:
            if columns[j]:
                columns[j] = False
selected_columns = X.columns[columns]
df_help_features = X[selected_columns]
df_help_features.info()
```

Dan didapatkanlah kematian anak dan ekspor dari seleksi fitur diatas untuk di clustering

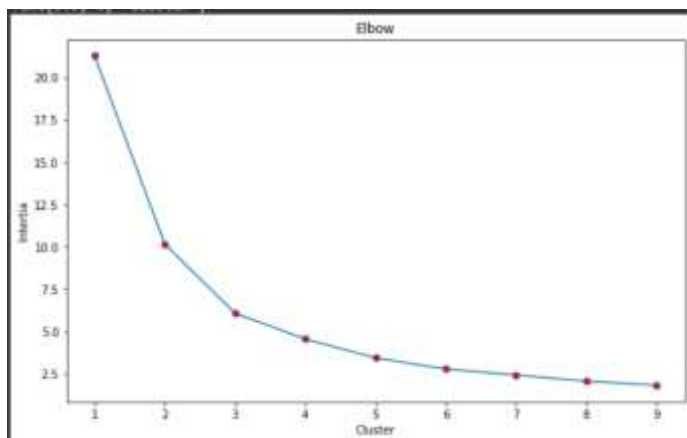
5. Setelah mendapatkan data kematian anak dan ekspor langkah selanjutnya mengecek outlier dari data kematian anak dan ekspor , jika ada outlier hilangkan terlebih dahulu outliernya agar bisa di proses ke tahap selanjutnya.
6. Setelah outlier dari kematian anak dan ekspor hilang , langkah selanjutnya melakukan proses scaling menggunakan scaling minmax
7. Setelah data di scaling langkah selanjutnya adalah memplot sebaran data dari kematian anak dan ekspor sebelum di clustering. Berikut sebaran data kematian anak dan ekspor sebelum di clustering



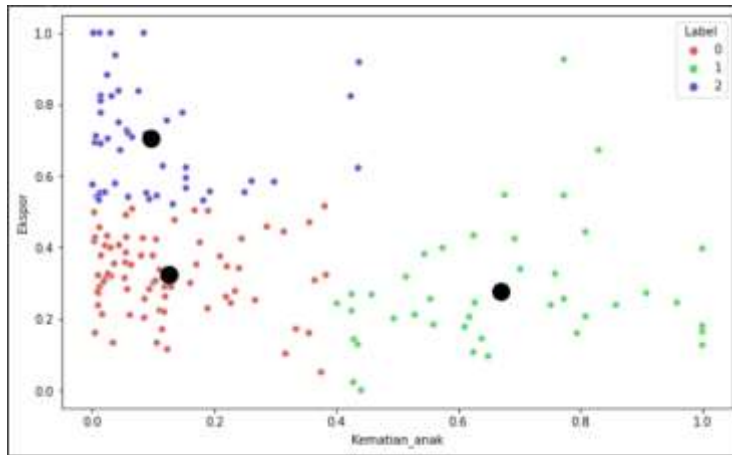
8. Setelah proses eksplorasi data dan praproses data selesai langkah selanjutnya adalah proses clustering menggunakan kmeans dengan cara mengimport library yang dibutuhkan dan library yang dibutuhkan sebagai berikut

```
from sklearn.cluster import KMeans
from sklearn import metrics
from scipy.spatial.distance import cdist
import numpy as np
import matplotlib.pyplot as plt
```

9. Setelah library di import langkah selanjutnya yaitu membuat grafik elbow untuk menentukan n_cluster yang paling bagus . Ciri-ciri cluster nilai n_cluster yang bagus dilihat dari patahan grafik yang paling menonjol. Pada grafik dibawah ini terlihat pada saat n_cluster =3 memiliki patahan yang cukup bagus saat menuju kedua.



10. Dari grafik elbow didapatkanlah $n_cluster$ sama dengan 3 dan ini merupakan sebaran data setelah dilakukannya proses clustering



Kesimpulan : Dapat disimpulkan bahwa cluster negara dengan label 1 perlu memerlukan bantuan karena jumlah kematiannya cukup banyak dan ekspornya sedikit. Dan negara negara yang perlu memerlukan bantuan sebagai berikut:

```
array(['Afghanistan', 'Angola', 'Benin', 'Burkina Faso', 'Burundi',  
      'Cameroon', 'Central African Republic', 'Chad', 'Comoros',  
      'Congo, Dem. Rep.', 'Cote d'Ivoire', 'Equatorial Guinea',  
      'Eritrea', 'Gambia', 'Ghana', 'Guinea', 'Guinea-Bissau', 'Haiti',  
      'India', 'Kenya', 'Kiribati', 'Lao', 'Lesotho', 'Liberia',  
      'Madagascar', 'Malawi', 'Mali', 'Mauritania', 'Mozambique',  
      'Myanmar', 'Niger', 'Nigeria', 'Pakistan', 'Rwanda', 'Senegal',  
      'Sierra Leone', 'Sudan', 'Tanzania', 'Timor-Leste', 'Togo',  
      'Uganda', 'Zambia'], dtype=object)
```