

# CLUSTERING THE COUNTRIES BY USING K-MEANS FOR HELP INTERNATIONAL

## A. LATAR BELAKANG

HELP International adalah LSM kemanusiaan internasional yang berkomitmen untuk memerangi kemiskinan dan menyediakan fasilitas dan bantuan dasar bagi masyarakat di negara-negara terbelakang saat terjadi bencana alam.

HELP International telah berhasil mengumpulkan sekitar \$10 juta dan perlu memutuskan bagaimana menggunakan dana ini secara strategis dan efektif dengan memilih negara yang paling membutuhkan bantuan berdasarkan faktor sosial ekonomi dan kesehatan yang menentukan perkembangan negara secara keseluruhan.

## B. TUJUAN

Untuk mengkategorikan negara menggunakan faktor sosial ekonomi dan kesehatan yang menentukan pembangunan negara secara keseluruhan.

## C. RINGKASAN DATA

Data yang digunakan merupakan data keadaan sosial ekonomi dan kesehatan dari seluruh negara yang menjadi kandidat penerima bantuan dari HELP. Data awal terdiri dari 10 kolom dan 167 baris. Kolom-kolom tersebut terdiri dari Negara, Kematian\_anak, Ekspor, Kesehatan, Impor, Pendapatan, Inflasi, Harapan\_hidup, Jumlah\_fertiliti, dan GDPperkapita.

	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200
...	...	...	...	...	...	...	...	...	...	...
162	Vanuatu	29.2	46.6	5.25	52.7	2950	2.62	63.0	3.50	2970
163	Venezuela	17.1	28.5	4.91	17.6	16500	45.90	75.4	2.47	13500
164	Vietnam	23.3	72.0	6.84	80.2	4490	12.10	73.1	1.95	1310
165	Yemen	56.3	30.0	5.18	34.4	4480	23.60	67.5	4.67	1310
166	Zambia	83.1	37.0	5.89	30.9	3280	14.00	52.0	5.40	1460

Data Kematian\_anak berada di rentang 2.6 – 208 dengan rata-rata 38.27 dan standar deviasi 40.33. Data Ekspor berada di rentang 0.11 – 200 dengan rata-rata 41.11 dan

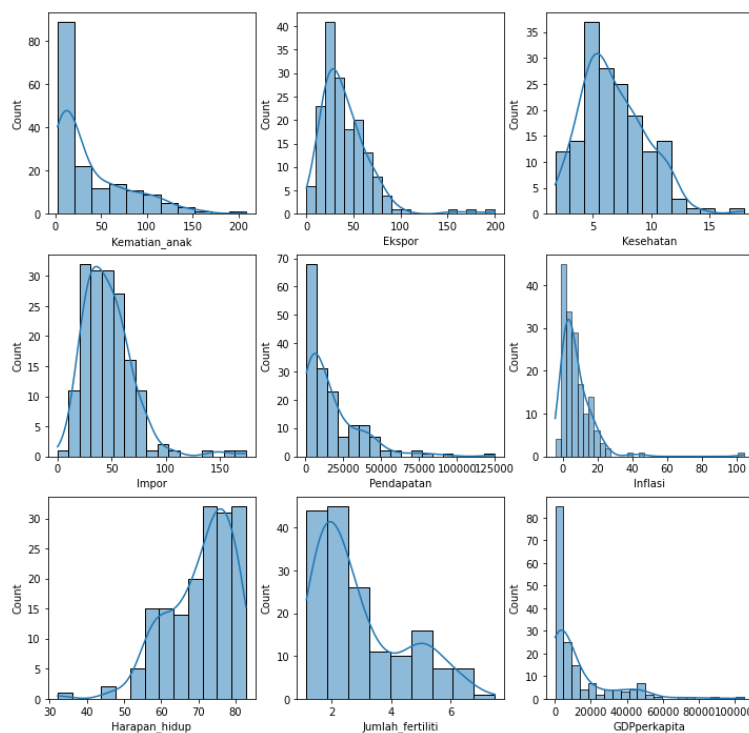
standar deviasi 27.41. Data Kesehatan berada di rentang 1.81 – 17.9 dengan rata-rata 6.82 dan standar deviasi 2.75. Data Impor berada di rentang 0.07 – 174 dengan rata-rata 46.89 dan standar deviasi 24.21. Data Pendapatan berada di rentang 609 – 125000 dengan rata-rata 17.144 dan standar deviasi 19.278. Data Inflasi berada di rentang -4.21 – 104 dengan rata-rata 7.78 dan standar deviasi 10.57. Data Harapan\_hidup berada di rentang 32.1 – 82.8 dengan rata-rata 70.56 dan standar deviasi 8.89. Data Jumlah\_fertiliti berada di rentang 1.15 – 7.49 dengan rata-rata 2.95 dan standar deviasi 1.51. Data GDPperkapita berada di rentang 231 – 105000 dengan rata-rata 12964.16 dan standar deviasi 18238.7.

	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
count	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000
mean	38.270060	41.108976	6.815689	46.890215	17144.688623	7.781832	70.555689	2.947964	12964.155689
std	40.328931	27.412010	2.746837	24.209589	19278.067698	10.570704	8.893172	1.513848	18328.704809
min	2.600000	0.109000	1.810000	0.065900	609.000000	-4.210000	32.100000	1.150000	231.000000
25%	8.250000	23.800000	4.920000	30.200000	3355.000000	1.810000	65.300000	1.795000	1330.000000
50%	19.300000	35.000000	6.320000	43.300000	9960.000000	5.390000	73.100000	2.410000	4660.000000
75%	62.100000	51.350000	8.600000	58.750000	22800.000000	10.750000	76.800000	3.880000	14050.000000
max	208.000000	200.000000	17.900000	174.000000	125000.000000	104.000000	82.800000	7.490000	105000.000000

Data terisi lengkap tanpa adanya missing values atau nilai yang kosong.

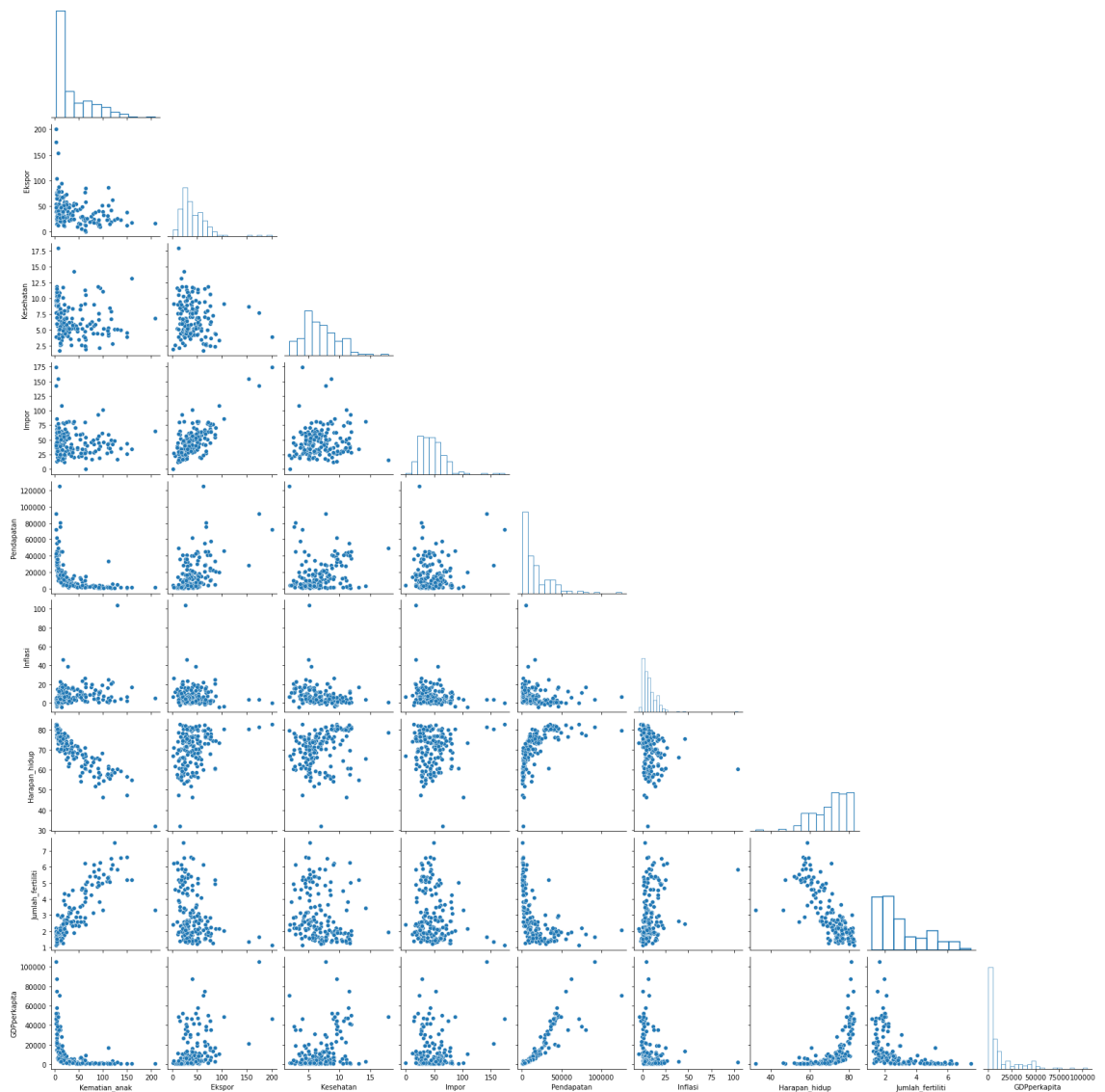
## D. ANALISA DATA

### 1. Analisa Univariate



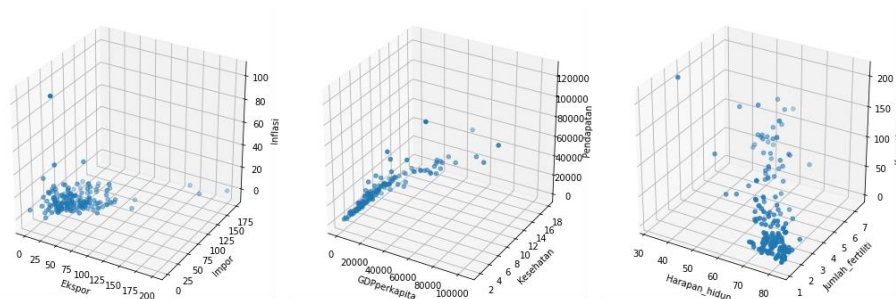
Dari grafik dapat dilihat bahwa sebagian besar data, seperti Kematian\_anak, Ekspor, Kesehatan, Pendapatan, Inflasi, Jumlah\_fertiliti, dan GDPperkapita memiliki kemiringan ke kanan atau positive skewness yang menandakan sebagian besar distribusi berada di nilai rendah. Sementara untuk Harapan\_hidup tampak miring ke kiri atau negative skewness sehingga sebagian besar data terdistribusi di nilai tinggi. Selain itu hanya data impor yang terlihat simetris atau hampoir seperti zero skewness, yang mana penyebaran datanya rata. Sementara untuk keruncingannya hamper semua tampak berbentuk leptokurtik.

## 2. Analisa Bivariate



Dilihat dari bivariate-nya tampak sebagian besar data sebenarnya memiliki hubungan yang rendah dan non-linear dengan outlier yang cukup banyak. Hal ini sebenarnya akan membuat data agak cukup sulit diklasterisasi. Terlihat hanya Kematian\_anak dan Jumlah\_fertiliti serta Pendapatan dan GDPperkapita yang memiliki hubungan linear positif artinya jika satu nilai variable naik maka variable lainnya pun akan ikut naik. Ada juga Kematian\_anak dan Harapan\_hidup yang memiliki hubungan linear negative, yang artinya jika satu nilai variable naik maka variable akan turun. Selain itu data yang lainnya seperti sudah disebutkan memiliki hubungan yang rendah dan non-linear.

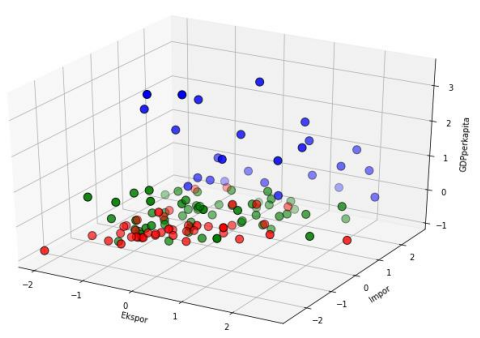
### 3. Analisa Multivariate



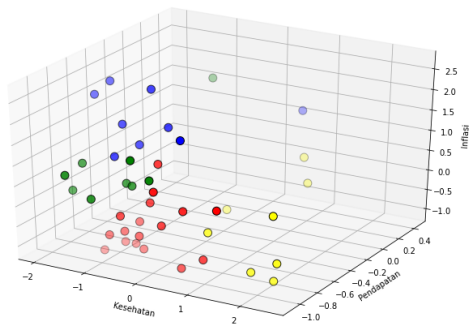
Dilihat berdasarkan hubungan multivariatnya, antara Ekspor, Impor, dan Inflasi agaknya memiliki hubungan lemah yang non linier, sedangkan GDPperkapita, Kesehatan, dan Pendapatan memiliki hubungan linear positif, dan Harapan\_hidup, Jumlah\_fertiliti, dan Kematian\_anak memiliki hubungan linear negatif.

## E. KLASSTERING K-MEANS

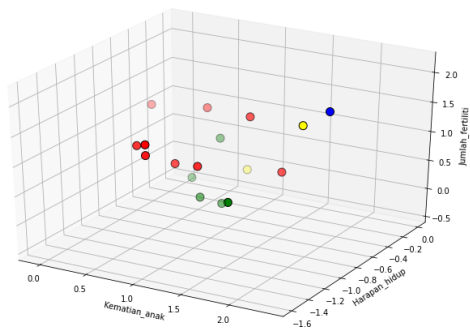
Pada proses klastering kali ini awalnya diambil tiga data sebagai pembentuk klaster awal yaitu Ekspor, Impor dan GDPperkapita.



Hasil dari klaster tersebut kemudian di klasterisasi lagi berdasarkan Kesehatan, Pendapatan dan Inflasi.



Hasil dari klaster tersebut diklasterkan lagi untuk yang terakhir kalinya



Hingga hasil yang didapat adalah sebagai berikut:

HELP	
Negara	
Yemen	0
Congo, Rep.	0
Eritrea	0
Gabon	0
Ghana	0
Sudan	0
Pakistan	0
Iraq	0
Madagascar	0
Angola	1
Guinea	1
Myanmar	2
Nepal	2
India	2
Tajikistan	2
Lao	2
Mauritania	3

Urutan negara di atas sudah diurutkan dari yang paling direkomendasikan, terutama untuk yang bernilai 0 dan 1.

#### **F. KESIMPULAN**

Berdasarkan hasil klastering maka negara yang direkomendasikan menerima bantuan adalah Yemen, Congo. Rep, Eritrea, Gabon, Ghana, Sudan, Pakistan, dan Iraq. Sementara untuk alternative berikutnya Angola dan Guinea.