

# ANALISIS TINGKAT KEMAKMURAN NEGARA MENGGUNAKAN ALGORITMA *K-MEANS* DAN FAKTOR-FAKTOR YANG MEMPENGARUHINYA

Miftahur Rizki<sup>#1</sup>, Mohammad Ikmal Akbar<sup>\*2</sup>, Mugni Hidayah<sup>#3</sup>

<sup>#1</sup>*SI Sains Data, Telkom University*

*Jl. Telekomunikasi No. 1, Terusan Buahbatu - Bojongsoang, Sukapura, Dayeuh Kolot, Bandung, Jawa Barat, Indonesia*

<sup>1</sup>[miftahurrizki@student.telkomuniversity.ac.id](mailto:miftahurrizki@student.telkomuniversity.ac.id)

<sup>2</sup>[lokololo@student.telkomuniversity.ac.id](mailto:lokololo@student.telkomuniversity.ac.id)

<sup>3</sup>[mugnihidayah@student.telkomuniversity.ac.id](mailto:mugnihidayah@student.telkomuniversity.ac.id)

**Abstrak**— Penelitian ini bertujuan untuk menganalisis pengaruh Gross Domestic Product (GDP) terhadap pertumbuhan ekonomi di negara-negara dunia. Latar belakang penelitian adalah bahwa GDP adalah indikator utama yang digunakan untuk mengukur pertumbuhan ekonomi, dan semakin tinggi nilai GDP, semakin baik kondisi perekonomian di negara tersebut. Permasalahan yang akan diselesaikan adalah mengetahui bagaimana GDP mempengaruhi pertumbuhan ekonomi di negara menggunakan metode *K-means*. Tujuan penelitian adalah untuk menganalisis tinggi dan rendahnya suatu GDP di negara-negara dunia, serta fitur-fitur yang mempengaruhi tinggi rendahnya nilai GDP di negara. Tahapan penelitian termasuk data cleaning, Exploratory Data Analysis, pemilihan fitur, data export ke dalam bentuk file csv, visualisasi data menggunakan Looker Studio, dan analisis mendalam.

**Kata Kunci**— GDP per capita, Pertumbuhan ekonomi, K-means, Exploratory Data Analysis, Looker.

## I. LATAR BELAKANG

Gross Domestic Product (GDP) Per Kapita adalah salah satu indikator utama yang digunakan untuk mengukur pendapatan rata-rata yang dihasilkan oleh setiap individu dalam suatu negara. Ini dihitung dengan rumus GDP negara/jumlah penduduk. GDP per kapita dapat memberikan gambaran tentang tingkat kesejahteraan rata-rata penduduk suatu negara.

Semakin tinggi nilai GDP per kapita suatu negara, semakin makmur masyarakat di negara tersebut, begitu juga sebaliknya. Banyak faktor yang mempengaruhi naik dan turunnya nilai GDP per kapita suatu negara.

Proyek ini akan menganalisis tingkat kemakmuran suatu negara berdasarkan nilai GDP per kapita nya menggunakan metode *K-Means*. Selain itu, proyek ini akan menganalisis faktor-faktor yang mempengaruhi nilai GDP Per Capita suatu negara.

Algoritma *K-Means* merupakan pilihan yang tepat untuk menganalisis GDP Per Capita seluruh negara di dunia karena sederhana, efisien, dan mudah diinterpretasikan. Dengan kemampuannya mengatasi data numerik, skala besar, dan bentuk kluster yang beragam, KMeans memberikan kontrol parameter yang memudahkan eksplorasi jumlah *cluster* yang optimal. Selain itu, sebagai algoritma *unsupervised learning*, *K-Means* dapat bekerja tanpa memerlukan label sebelumnya, menjadikannya cocok untuk studi kasus analisis GDP Per Capita yang mungkin tidak memiliki informasi label kluster sebelumnya.

Proyek ini menggunakan metode *K-Means* untuk mengelompokkan negara berdasarkan pengaruh GDP Per Capita terhadap pertumbuhan ekonomi suatu negara. *K-Means* akan digunakan untuk membagi negara-negara berdasarkan tingkat kemakmuran masyarakat di suatu negara, yang diukur dengan GDP Per Capita mereka. Dengan mengelompokkan negara, dapat dilihat bagaimana GDP Per Capita mempengaruhi tingkat kesejahteraan masyarakat negara tersebut, serta bagaimana negara yang memiliki GDP Per Capita tinggi berbeda dengan negara yang memiliki nilai rendah. Analisis ini diharapkan dapat membantu dalam merumuskan kebijakan ekonomi yang lebih efektif, mengidentifikasi area dimana perbaikan diperlukan, serta memberikan kontribusi yang berharga bagi penelitian ekonomi dan pembahasan kebijakan publik yang lebih luas.

## II. PERMASALAHAN

Rumusan masalah yang dihadapi adalah sebagai berikut:

1. Apa saja faktor-faktor yang mempengaruhi pertumbuhan GDP per kapita di suatu negara?
2. Bagaimana kita dapat mengelompokkan negara-negara berdasarkan tingkat GDP per kapita menggunakan algoritma *K-Means*?

## III. TUJUAN

Dengan melakukan analisis faktor-faktor yang mempengaruhi nilai GDP dan melakukan pengelompokan negara-negara berdasarkan nilai GDP mereka menggunakan algoritma *K-Means*, kami berharap dapat memperoleh pemahaman yang lebih baik tentang pola pertumbuhan ekonomi di berbagai wilayah serta membandingkan kinerja ekonomi antara kelompok negara yang berbeda. Hasil analisis ini diharapkan dapat membantu dalam merumuskan kebijakan ekonomi yang lebih efektif, mengidentifikasi area dimana perbaikan diperlukan, serta memberikan kontribusi yang berharga bagi penelitian ekonomi dan pembahasan kebijakan publik yang lebih luas.

## IV. DATASET

Dataset yang digunakan dalam proyek ini adalah dataset "The World Factbook" yang disediakan oleh Badan Intelijen Pusat Amerika Serikat (CIA) yang terdiri dari 30 kolom dan 149 baris. Dataset ini mencakup berbagai variabel ekonomi, sosial, dan politik dari negara-negara di seluruh dunia.

## V. METODE

### A. *K-Means*

*K-Means* merupakan salah satu jenis algoritma *unsupervised learning*. Algoritma *K-Means* berfungsi untuk mengelompokkan suatu data berdasarkan nilai *K*. Caranya adalah dengan melihat jumlah grup yang diwakili variabel *K*.

Untuk bisa mengelompokkan data, algoritma ini bekerja secara berulang (iteratif) dari setiap titik data ke grup *K* yang telah dibuat.

Umumnya, data akan dikelompokkan menurut kesamaan fitur. Semakin mirip fitur yang dimiliki suatu data, maka makin besar peluangnya untuk berada dalam satu kelompok.

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

Di mana:

$\mu_k$  = titik centroid dari cluster ke-*K*

$N_k$  = banyaknya data pada cluster ke-*K*

$x_q$  = data ke-*q* pada cluster ke-*K*

## VI. TAHAPAN

- Pertama, dataset akan dibersihkan, atau biasa disebut *data cleaning*. Proses ini dilakukan untuk memastikan bahwa data yang akan digunakan untuk analisis tidak mengandung *missing value* sehingga keluaran analisis dapat lebih baik.
- Selanjutnya, dilakukan *Exploratory Data Analysis* untuk lebih mengenali apa saja isi dari data yang akan digunakan untuk analisis selanjutnya.
- Tahapan selanjutnya adalah pemilihan fitur yang akan digunakan untuk proses *machine learning*. Setelah itu, data akan di *export* ke dalam bentuk file csv untuk proses visualisasi data.
- Selanjutnya melakukan visualisasi data menggunakan Looker Studio. Pada tahap ini juga akan dilakukan analisis mendalam mengenai data dan pembuatan dashboard visualisasi data.

## VII. HASIL DAN PEMBAHASAN

### A. *Explorasi Data Analisis*

Hal yang dilakukan pada tahap ini adalah memastikan tipe data setiap kolom sesuai dengan isi dan juga konteks datanya.

Data columns (total 28 columns):

#	Column	Non-Null Count	Dtype
0	Country	149 non-null	object
1	Continent	149 non-null	object
2	Area	149 non-null	int64
3	Birth rate	149 non-null	float64
4	Current account balance	149 non-null	object
5	Death rate	149 non-null	float64
6	Electricity consumption	149 non-null	float64
7	Electricity production	149 non-null	float64
8	Exports	149 non-null	float64
9	GDP	149 non-null	float64
10	GDP per capita	149 non-null	int64
11	GDP real growth rate	149 non-null	float64
12	Highways	149 non-null	int64
13	Imports	149 non-null	float64
14	Industrial production growth rate	134 non-null	float64
15	Infant mortality rate	149 non-null	float64
16	Inflation rate	149 non-null	float64
17	Internet users	149 non-null	object
18	Investment	144 non-null	float64
19	Labor force	140 non-null	object
...			
26	Total fertility rate	149 non-null	float64
27	Unemployment rate	127 non-null	float64

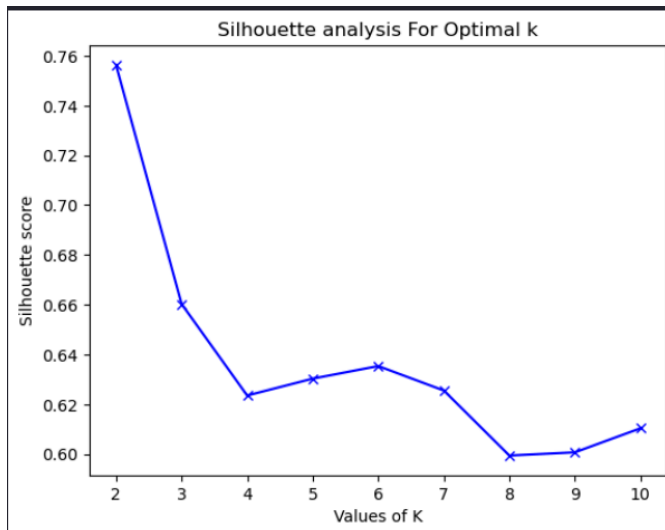
## B. Pemilihan Fitur

Karena fitur pada data terlalu beragam, perlu dilakukan pemilihan fitur yang sesuai dengan tujuan proyek ini, yaitu fitur yang menjadi matriks sebuah negara dapat dikatakan makmur.

```
Index(['Country', 'Continent', 'GDP per capita', 'Death rate', 'Exports',
      'Imports', 'Infant mortality rate', 'Life expectancy at birth',
      'Unemployment rate', 'Birth rate'],
      dtype='object')
```

## C. Membangun Model K-Means

Untuk menentukan nilai K yang optimal, proyek ini menggunakan rata-rata skor silhouette. Nilai silhouette berkisar antara -1 hingga 1. Jika suatu cluster memiliki rata-rata nilai yang semakin mendekati 1, maka semakin baik.



Pada data kami, nilai yang paling baik berada di cluster 2 dengan nilai silhouette adalah 0.76. Maka dari itu bisa dipastikan data akan dikelompokkan dengan jumlah 2 cluster.

```
# Melakukan clustering
n_clusters_gdp = 2 # Jumlah cluster yg optimal pada kurva = 2
kmeans_pendapatan = KMeans(n_clusters=n_clusters_gdp, init='k-means++', random_state=42)
clus["Cluster"] = kmeans_pendapatan.fit_predict(clus[["GDP per capita"]])

clus

✓ 42s
```

	Country	Continent	GDP per capita	Death rate	Exports	Imports	Infant mortality rate	Life expectancy at birth	Unemployment rate	Birth rate	Cluster
0	United States	America	40100	8.25	7.950000e+11	1.470000e+12	6.50	77.71	5.5	1414	1
1	Norway	Europe	40000	9.45	7.664000e+10	4.596000e+10	3.70	79.40	4.3	1167	1
2	Hong Kong	Asia	34200	5.98	2.681000e+11	2.759000e+11	2.97	81.39	6.7	723	1
3	Switzerland	Europe	33800	8.48	1.307000e+11	1.211000e+11	4.39	80.39	3.4	977	1
4	Denmark	Europe	32200	10.43	7.306000e+10	6.345000e+10	4.56	77.62	6.2	1136	1
...	...	...	...	...	...	...	...	...	...	...	...
144	Madagascar	Africa	800	11.35	8.682000e+08	1.147000e+09	76.83	56.95	NaN	41.66	0
145	Yemen	Asia	800	8.53	4.468000e+09	3.749000e+09	61.50	61.75	35.0	43.87	0
146	Tanzania	Africa	700	16.71	1.248000e+09	1.072000e+09	95.54	45.24	NaN	38.16	0
147	Burundi	Africa	600	17.43	3.184000e+07	1.382000e+08	69.29	43.50	NaN	39.66	0
148	Malawi	Africa	600	23.39	5.034000e+08	5.211000e+08	103.32	36.97	NaN	43.95	0

149 rows x 11 columns

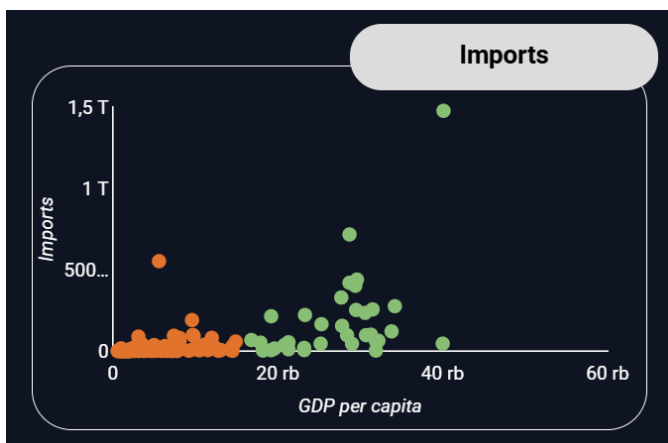
Selanjutnya data akan dikelompokkan berdasarkan nilai clusternya dan disimpan dalam sebuah dataframe baru bernama "clus".

Cluster	GDP per capita	Death rate	Exports	Imports	Infant mortality rate	Life expectancy at birth	Unemployment rate	Birth rate
0	5066.391304	10.386609	2.112294e+10	1.910102e+10	45.498696	63.526000	17.384946	24.850000
1	27102.941176	8.086765	1.853201e+11	1.916401e+11	5.882059	78.450588	6.497059	11.877353

Gambar diatas merupakan informasi rata-rata nilai matriks dari setiap cluster. Dapat dilihat bahwa cluster 0 cenderung memiliki nilai yang rendah dibandingkan dengan cluster 1. Maka dari informasi tersebut dapat disimpulkan bahwa cluster 1 merupakan kelompok negara yang makmur, sedangkan cluster 0 kurang makmur.

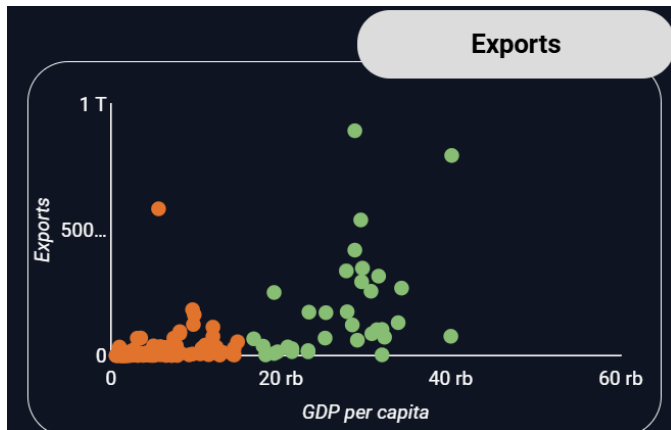
## D. Faktor-faktor yang mempengaruhi

### 1) Import



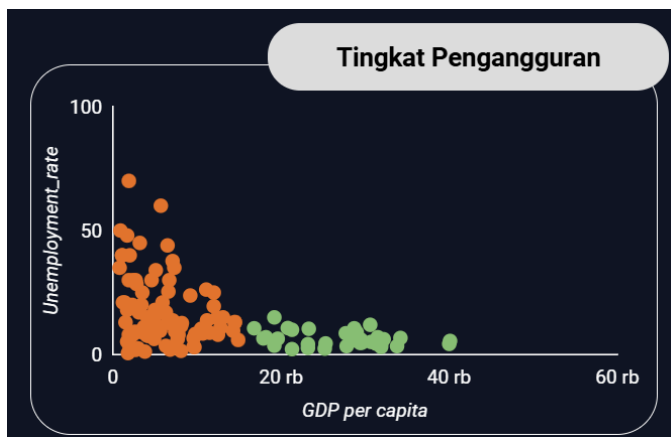
Negara yang makmur umumnya menunjukkan tingkat impor yang tinggi, karena kegiatan tersebut memainkan peran penting dalam mendukung pertumbuhan ekonomi negara.

## 2) Ekspor



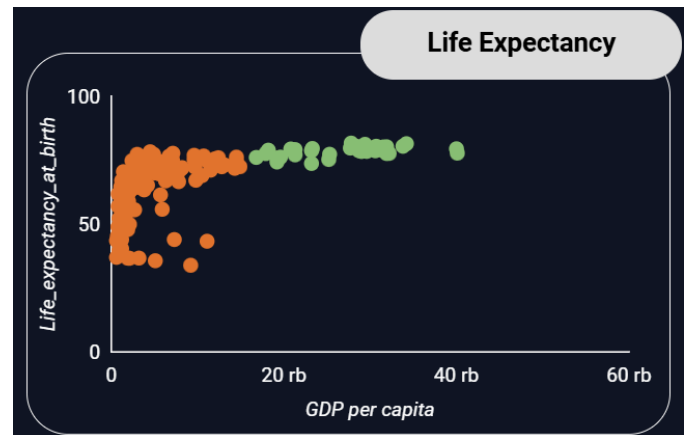
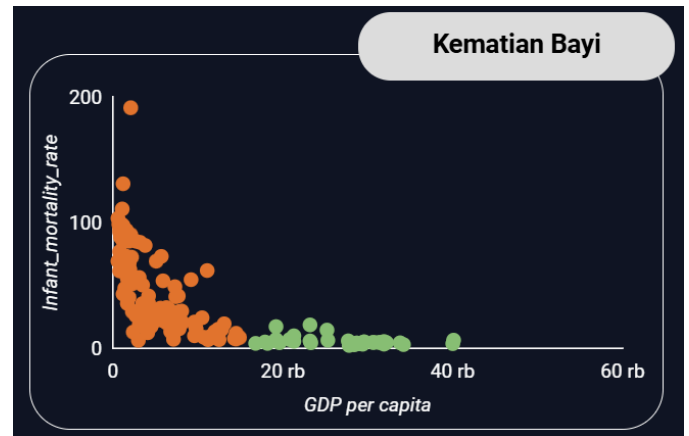
Begitu juga dengan ekspor, negara yang makmur umumnya menunjukkan tingkat ekspor yang tinggi, karena kegiatan tersebut memainkan peran penting dalam mendukung pertumbuhan ekonomi negara.

## 3) Tingkat Pengangguran



Negara yang makmur umumnya menunjukkan tingkat pengangguran yang rendah, karena Pengangguran rendah menandakan ekonomi stabil, dorong daya beli, pajak, kurangi beban sosial, tingkatkan produktivitas, dan dorong pertumbuhan jangka panjang ekonomi.

## 4) Tingkat Kematian Bayi dan Harapan Hidup

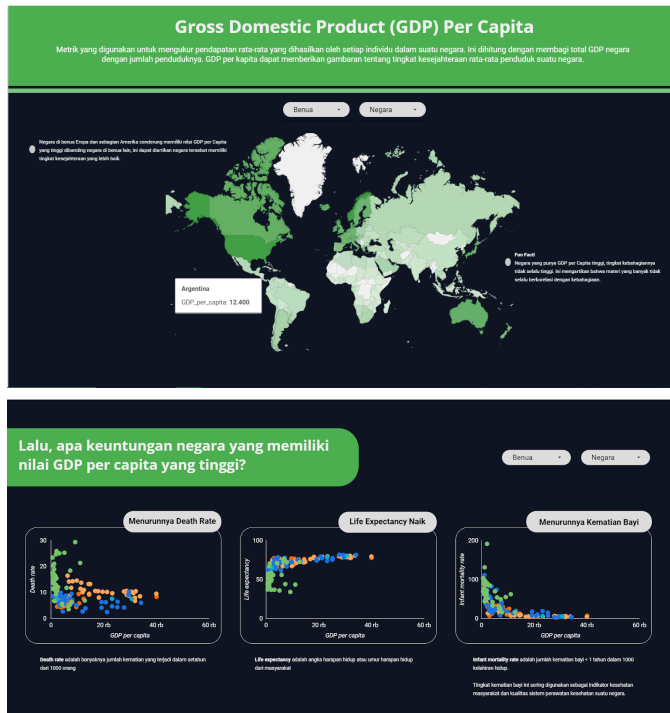


Pada negara makmur, tingkat kematian & kematian bayi cenderung rendah dan harapan hidup tinggi karena akses yang baik terhadap layanan kesehatan dan kondisi hidup yang baik.

Di negara kurang makmur, tingkat kematian & kematian bayi biasanya lebih tinggi dan harapan hidup rendah karena kurangnya akses terhadap perawatan kesehatan dan kondisi hidup yang sulit.

## E. Visualisasi Data

Gambar dibawah ini merupakan hasil dari analisis yang telah dilakukan dan divisualisasikan menggunakan Looker Studio.



## VIII. KESIMPULAN

Negara makmur memiliki impor, ekspor, dan harapan hidup tinggi serta tingkat kematian dan pengangguran yang rendah, sedangkan negara kurang makmur menunjukkan sebaliknya.

Hasil dari Proyek ini diharapkan dapat membantu mengidentifikasi area untuk perbaikan ekonomi dan membantu individu memilih negara dengan kualitas hidup lebih baik.

## REFERENSI

- [1] D. Syaputri, P. H. Noprita, and S. Romelah, "Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 1–6, Mar. 2021, doi: <https://doi.org/10.57152/malcom.v1i1.5>.
- [2] M. Mukarramah and M. Zulkarnain, "Pengaruh Gross Domestic Product dan Kurs Terhadap Ekspor Negara-Negara ASEAN," *Jurnal Samudra Ekonomika*, vol. 7, no. 2, pp. 333–341, Sep. 2023, doi: <https://doi.org/10.33059/jse.v7i2.8528>.
- [3] Team, A., & Bunga. (2022, July 17). 3 hal penting tentang K means clustering. Algoritma. <https://algorit.ma/blog/k-means-clustering-2022/>