

# **ANALISIS CLUSTERING PENGARUH INDEKS KEBAHAGIAAN DENGAN ANGKA HARAPAN HIDUP TERHADAP PERTUMBUHAN PENDUDUK DI BERBAGAI NEGARA TAHUN 2015 DENGAN APLIKASI RSTUDIO.**

**Teni Deinarosa Hermalia**

Prodi Sains Data, Universitas Koperasi Indonesia, Fakultas Sains Dan Teknologi  
E-mail: tenidhermalia.xmma@gmail.com

## **Abstrak**

Penelitian ini bertujuan untuk menganalisis hubungan antara indeks kebahagiaan, angka harapan hidup, dan pertumbuhan penduduk di berbagai negara menggunakan analisis kluster. Dengan memanfaatkan data dari basis data internasional, kami menggunakan RStudio untuk komputasi statistik dan grafik. Analisis ini melibatkan pengelompokan negara berdasarkan indeks kebahagiaan dan angka harapan hidup, mengidentifikasi pola, dan menginterpretasikan dampaknya terhadap pertumbuhan penduduk. Hasil kluster ini membantu dalam memahami pengelompokan negara dengan karakteristik serupa dan mengeksplorasi bagaimana faktor-faktor ini saling mempengaruhi untuk memengaruhi dinamika populasi. Wawasan dari analisis ini dapat membantu pembuat kebijakan dalam merumuskan strategi untuk meningkatkan kesejahteraan dan angka harapan hidup penduduk, sehingga berdampak positif pada pertumbuhan penduduk.

**Kata Kunci :** *Indeks Kebahagiaan, Angka Harapan Hidup, Pertumbuhan Penduduk, Analisis Kluster, RStudio, Komputasi Statistik.*

## **1. PENDAHULUAN**

Indeks kebahagiaan dan angka harapan hidup merupakan dua indikator utama yang sering digunakan untuk menilai kesejahteraan suatu negara. Indeks kebahagiaan mengukur tingkat kepuasan hidup dan kesejahteraan emosional warga negara, sementara angka harapan hidup menggambarkan kesehatan umum dan kualitas hidup penduduk. Kedua indikator ini sering kali digunakan untuk memberikan gambaran tentang kondisi sosial dan ekonomi suatu negara. Hubungan antara indeks kebahagiaan, angka harapan hidup, dan pertumbuhan penduduk menarik untuk diteliti karena ketiga variabel ini saling mempengaruhi. Negara dengan indeks kebahagiaan tinggi dan angka harapan hidup yang tinggi cenderung memiliki pertumbuhan penduduk yang stabil atau meningkat, sementara negara dengan indeks kebahagiaan dan angka harapan hidup yang rendah mungkin menghadapi tantangan dalam pertumbuhan penduduk.

Dari pernyataan diatas maka dibuatlah penelitian dengan memanfaatkan data-data Website Kaagle <https://www.kaggle.com/> seperti ranking kebahagiaan, skor kebahagiaan, GDP per Capita, dan angka harapan hidup. Didalam ilmu computer banyak cabang ilmu yang dapat menyelesaikan masalah yang kompleks. Berdasarkan penjelasan tersebut peneliti menggunakan metode kalster .Penggunaan metode kluster melalui proses normalisasi variable, penentuan jumlah claster optimal menggunakan metode elbow, penerapan algoritma k-means untuk mengelompokan data, dan interpretasi kluster.

Hasil penelitian menunjukkan pola yang menarik, di mana negara-negara dengan ranking dan skor kebahagiaan, GDP per kapita, dan angka harapan hidup yang tinggi cenderung memiliki tingkat pertumbuhan penduduk yang lebih stabil dibandingkan dengan negara-negara dengan nilai yang lebih rendah. Selain itu, ditemukan bahwa faktor ekonomi dan kesehatan berperan penting dalam menentukan kebahagiaan dan pertumbuhan penduduk suatu negara.

## 2. METODOLOGI PENELITIAN

### 2.1 Data Mining

Data mining adalah proses mengekstraksi informasi yang berguna dari kumpulan data besar. Dalam penelitian ini, data mining digunakan untuk mengumpulkan dan mengolah data yang relevan mengenai indeks kebahagiaan, angka harapan hidup, dan pertumbuhan penduduk dari berbagai negara bersumber dari kaagle.

### 2.2 Clustering

Clustering adalah teknik analisis data yang digunakan untuk mengelompokkan objek-objek dalam suatu dataset ke dalam grup-grup (cluster) berdasarkan kemiripan mereka. Dalam penelitian ini, clustering digunakan untuk mengelompokkan negara-negara berdasarkan indeks kebahagiaan dan angka harapan hidup.

### 2.3 K-Means

K-Means adalah algoritma clustering yang bertujuan untuk membagi  $n$  objek ke dalam  $k$  cluster di mana setiap objek termasuk ke dalam cluster dengan centroid terdekat.

Proses K-Means melibatkan beberapa langkah utama:

1. Memilih  $k$  titik sebagai centroid awal dari cluster.
2. Mengalokasikan setiap titik data ke centroid terdekat berdasarkan jarak Euclidean.
3. Menghitung ulang posisi centroid dengan mengambil rata-rata dari semua titik data dalam cluster yang sama.
4. Mengulangi proses penugasan dan pembaruan hingga tidak ada perubahan signifikan dalam posisi centroid atau hingga jumlah iterasi maksimum tercapai .
- 5.

## 3. ANALISA DAN PEMBAHASAN

### 3.1 Pengimplementasian pada RStudio

Untuk melakukan analisis *clustering* dengan *K-Means* di *RStudio*, dengan mengikuti langkah-langkah berikut:

Pertama, memuat data yang telah dikumpulkan ke dalam *RStudio*. Data yang digunakan mencakup ranking kebahagiaan, skor kebahagiaan, GDP per kapita, dan angka harapan hidup.

**# Memuat pustaka yang di perlukan**

```
library(tidyverse)
```

```
library(factoextra)
```

Pustaka *tidyverse* digunakan untuk manipulasi data dan visualisasi, sedangkan *factoextra* digunakan untuk visualisasi analisis *clustering*.

**# Memuat data**

```
data=read.csv("D:\\DATA KULIAH\\SEMESTER 4\\BIG DATA ANALYSTICS\\mentahan tugas  
beesar\\Dataa.csv",header=T,sep=";")  
data
```

**attach(data)**

- *read.csv* digunakan untuk membaca file CSV dengan header dan separator ;.
- `attach(data)` digunakan untuk membuat kolom data dapat diakses secara langsung (misalnya, mengakses *Country* tanpa harus menggunakan `data$Country`).

**#SELEKSI KOLOM RELEVAN**

```
data_selected<- data %>%
```

```
  select(Country,HappinessRank,HappinessScore,GDP,LifeExpectancy)
```

Memilih kolom yang relevan dari data: *Country*, *HappinessRank*, *HappinessScore*, *GDP*, dan *LifeExpectancy*.

**#MENGHAPUS BARIS DENGAN NILAI YANG HILANG**

```
data_clean<- na.omit(data_selected)
```

Menghapus baris yang memiliki nilai yang hilang (NA).

**#MENGHAPUS KOLOM COUNTRY UNTUK ANALISIS CLUSTERING**

```
data_clustering<- data_clean %>%
```

```
  select(-Country)
```

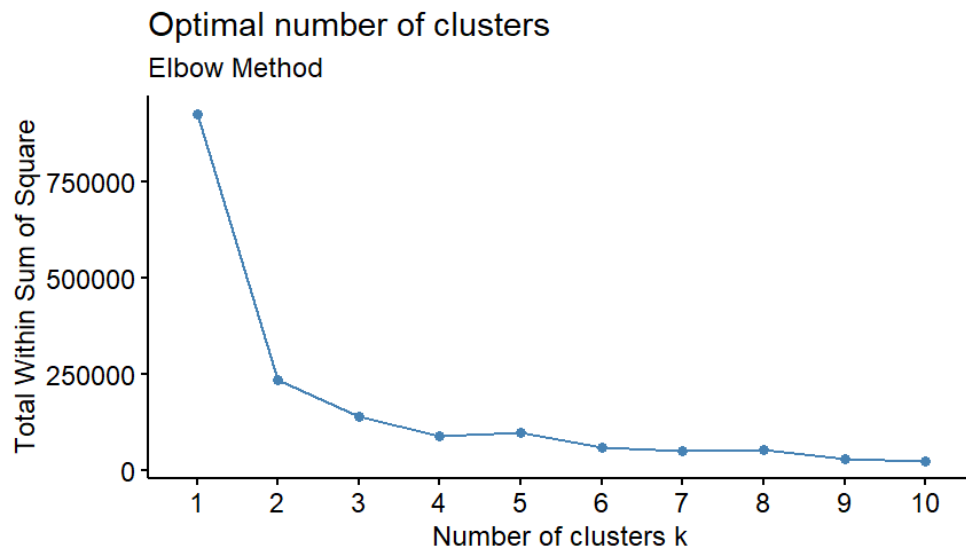
Menghapus kolom *Country* karena tidak diperlukan dalam analisis *clustering*.

**#MENENTUKAN JUMLAH CLUSTER OPTIMAL DENGAN METODE ELBOW**

```
fviz_nbclust(data_clustering, kmeans, method = "wss")+ 
```

```
  labs(subtitle= "elbow Method")
```

- *fviz\_nbclust* digunakan untuk menentukan jumlah *cluster* optimal. Metode *wss* (*within-cluster sum of squares*) digunakan untuk mencari jumlah *cluster* optimal dengan memplot jumlah *cluster* terhadap total *within-cluster sum of squares*.
- *labs* digunakan untuk menambahkan *subtitle* pada plot.



```
set.seed(123) #UNTUK REPLIKASI HASIL
```

```
kmeans_result<- kmeans(data_clustering, centers = 3, nstart = 25)
```

- *set.seed(123)* digunakan untuk memastikan hasil yang replikatif.
- *kmeans* digunakan untuk melakukan *K-Means clustering* dengan jumlah *cluster (centers)* sebanyak 3 dan memulai algoritma dengan 25 inisialisasi yang berbeda (*nstart*).

```
#MENAMBAHKAN HASIL CLUSTER KE DATA ASLI
```

```
data_clean$Cluster <- as.factor(kmeans_result$cluster)
```

Menambahkan hasil *cluster* sebagai kolom baru (*Cluster*) ke data asli. Mengubah hasil *cluster* menjadi faktor

```
#VISUALISASI HASIL CLUSTERING
```

```
fviz_cluster(kmeans_result, data = data_clustering,
```

```
  geom = "point",
```

```
  stand = FALSE,
```

```
  ellipse.type= "norm")+
```

```
labs(title = "K-Means Clustering",
```

```
  x = "GDP per Capita",
```

```
  y = "Life Expectancy")
```

- *fviz\_cluster* digunakan untuk memvisualisasikan hasil clustering.
- *geom = "point"* menentukan bahwa data ditampilkan sebagai titik.
- *stand = FALSE* menunjukkan bahwa data tidak dinormalisasi kembali.
- *ellipse.type = "norm"* menambahkan *elips* untuk menggambarkan distribusi normal dari masing-masing *cluster*.
- *labs* digunakan untuk menambahkan judul dan label sumbu pada plot.



### #MENAMPILKAN RINGKASAN STATISTIK DARI SETIAP CLUSTER

```
data_clean %>%  
  group_by(Cluster) %>%  
  summarise(  
    HappinessRank = mean(HappinessRank),  
    HappinessScore = mean(HappinessScore),  
    GDP = mean(GDP),  
    LifeExpectancy = mean(LifeExpectancy)  
  )
```

- group\_by(Cluster) mengelompokkan data berdasarkan cluster.
- summarise menghitung rata-rata (mean) dari HappinessRank, HappinessScore, GDP, dan LifeExpectancy untuk setiap cluster.

```
# A tibble: 3 × 5  
  Cluster HappinessRank HappinessScore    GDP LifeExpectancy  
  <fct>      <dbl>          <dbl>    <dbl> <dbl>  
1 1         63.1           5.70    1.08  0.640  
2 2         44.6           6.24   123.  4.03  
3 3        128.           4.23    0.628  0.410  
> |
```

Tabel tersebut menyajikan data dalam bentuk "tibble" yang terdiri dari tiga kluster dengan lima variabel, yaitu peringkat kebahagiaan, skor kebahagiaan, PDB, dan harapan hidup. Pada kluster 1, peringkat kebahagiaan adalah 63.1 dengan skor kebahagiaan 5.70, PDB 1.08, dan harapan hidup 0.640. Kluster 2 memiliki peringkat kebahagiaan 44.6 dengan skor kebahagiaan tertinggi di antara ketiga kluster yaitu 6.24, PDB sangat tinggi sebesar 123., dan harapan hidup 4.03. Sementara itu, kluster 3 menunjukkan peringkat kebahagiaan terendah yaitu 128., dengan skor kebahagiaan 4.23, PDB 0.628, dan harapan hidup 0.410. Data ini mengindikasikan bahwa kluster dengan PDB dan harapan hidup yang lebih tinggi cenderung memiliki skor kebahagiaan yang lebih tinggi.

## 3.2 Hasil Clustering

Hasil dari K-Means clustering menunjukkan bahwa data dapat dibagi ke dalam tiga cluster utama. Berikut adalah karakteristik masing-masing cluster:

- **Cluster 1**  
Negara-negara dalam cluster ini memiliki skor kebahagiaan yang tinggi, GDP per kapita yang tinggi, dan angka harapan hidup yang tinggi. Cluster ini umumnya terdiri dari negara-negara maju dengan standar hidup yang tinggi.
- **Cluster 2**  
Negara-negara dalam cluster ini memiliki skor kebahagiaan yang sedang, GDP per kapita yang sedang, dan angka harapan hidup yang sedang. Cluster ini mencakup negara-negara berkembang yang memiliki ekonomi dan kesehatan yang sedang.
- **Cluster 3**  
Negara-negara dalam cluster ini memiliki skor kebahagiaan yang rendah, GDP per kapita yang rendah, dan angka harapan hidup yang rendah. Cluster ini terdiri dari negara-negara yang sedang menghadapi tantangan dalam bidang ekonomi dan kesehatan.

### 3.3 Interpretasi Hasil



Dari hasil clustering, dapat dilihat bahwa terdapat hubungan yang jelas antara kebahagiaan, kesehatan, dan pertumbuhan penduduk:

- **Cluster 1:**  
Negara-negara dalam cluster ini menunjukkan bahwa kebahagiaan dan kesehatan yang tinggi berkorelasi dengan pertumbuhan penduduk yang stabil. Kualitas hidup yang tinggi menarik untuk penduduk untuk menetap dan membangun keluarga.
- **Cluster 2:**  
Negara-negara dalam cluster ini berada pada posisi menengah. Meskipun tidak sebaik cluster 1, negara-negara ini menunjukkan tren peningkatan dalam bidang ekonomi dan kesehatan yang dapat berkontribusi pada pertumbuhan penduduk yang positif.
- **Cluster 3:**  
Negara-negara dalam cluster ini menghadapi tantangan besar dalam meningkatkan kebahagiaan dan harapan hidup penduduknya. Pertumbuhan penduduk di negara-negara ini lebih bervariasi dan cenderung tidak stabil.

## 4. KESIMPULAN

Penelitian ini menganalisis pengaruh indeks kebahagiaan dan angka harapan hidup terhadap pertumbuhan penduduk di berbagai negara pada tahun 2015 menggunakan metode clustering dengan RStudio. Hasilnya menunjukkan hubungan signifikan antara ketiga variabel tersebut, di mana negara dengan kebahagiaan dan harapan hidup tinggi cenderung memiliki pertumbuhan penduduk yang stabil, sedangkan negara dengan nilai rendah menghadapi tantangan dalam kesejahteraan dan kesehatan. Hasil clustering mengelompokkan negara-negara ke dalam tiga cluster berdasarkan tingkat kebahagiaan, GDP per kapita, dan harapan hidup, memberikan panduan bagi pembuat kebijakan untuk merumuskan strategi peningkatan kualitas hidup yang lebih tepat sasaran. Penelitian ini menekankan pentingnya kebahagiaan dan kesehatan dalam mempengaruhi dinamika populasi, sehingga kebijakan yang fokus pada peningkatan kualitas hidup dapat berdampak positif pada stabilitas dan pertumbuhan penduduk.

## REFERENCES

- [1] World Happiness Report 2015. Retrieved from World Happiness Report
- [2] World Bank Data 2015. Retrieved from [World Bank](#)
- [3] Satopaa, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. Proceedings of the 31st International Conference on Distributed Computing Systems Workshops.
- [4] Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory.
- [5] MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.
- [6] Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis. Wiley.