

ANALISIS KLASTERING MENGGUNAKAN METODE K-MEANS dan PCA PENGELUARAN PER KAPITA di INDONESIA TAHUN 2022-2023 MENGGUNAKAN RSTUDIO

Cahya Rahmawati

Universitas Koperasi Indonesia

Email : cahyarahma170@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk menganalisis data pengeluaran perkapita di Indonesia selama periode tahun 2022-2023 menggunakan metode klustering K-Means di Rstudio. Pengeluaran per kapita merupakan indikator penting dalam menilai tingkat kesejahteraan ekonomi individu dan rumah tangga. Data yang digunakan dalam penelitian ini mencakup pengeluaran per kapita dari berbagai wilayah di Indonesia, diukur dalam ribu rupiah per orang per tahun. Dengan memanfaatkan perangkat lunak Rstudio, penelitian ini menerapkan algoritma K-Means untuk mengelompokkan data pengeluaran per kapita ke dalam beberapa kluster yang homogen.

Langkah-langkah penelitian mencakup pra-pemrosesan data untuk memastikan kualitas dan konsistensi, penentuan jumlah kluster optima menggunakan metode Elbow, dan evaluasi hasil klustering menggunakan metrik validasi seperti Silhouette Score. Hasil analisis menunjukkan adanya segmentasi yang signifikan dalam pengeluaran per kapita di Indonesia, dengan kluster-kluster yang mengindikasikan perbedaan regional dan ekonomi yang mencolok. Penemuan ini memberikan wawasan berharga bagi pembuat kebijakan dan peneliti ekonomi dalam merancang strategi yang lebih tepat sasaran untuk meningkatkan kesejahteraan ekonomi dan mengurangi disparitas regional.

Penelitian ini tidak hanya berkontribusi pada literatur ekonomi dan sosial di Indonesia, tetapi juga menunjukkan potensi aplikasi teknik klustering K-Means dalam analisis data ekonomi yang kompleks. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi dasar bagi pengambilan keputusan yang lebih berbasis data dan efektif di masa depan.

1. PENDAHULUAN

Pengeluaran per kapita merupakan salah satu indikator kunci dalam mengukur tingkat kesejahteraan ekonomi suatu populasi. Di Indonesia, analisis pengeluaran per kapita menjadi penting untuk memahami distribusi kesejahteraan ekonomi di berbagai wilayah dan kelompok masyarakat. Penelitian mengenai pola pengeluaran per kapita dapat memberikan wawasan mendalam tentang dinamika ekonomi regional serta faktor-faktor yang mempengaruhi distribusi pendapatan dan konsumsi. Metode klustering K-Means adalah salah satu teknik analisis data yang efektif untuk mengelompokkan data berdasarkan kemiripan tertentu.

Dalam konteks analisis pengeluaran per kapita, metode ini dapat digunakan untuk mengidentifikasi kelompok-kelompok homogen yang memiliki pola pengeluaran serupa. Pengeluaran algoritma K-Means di Rstudio memungkinkan untuk memanfaatkan kapabilitas komputasi dan visualisasi data yang kuat, sehingga hasil analisis dapat lebih mudah diinterpretasikan dan diimplementasikan. Penelitian ini bertujuan untuk mengaplikasikan metode klustering K-Means dalam menganalisis data pengeluaran per kapita di Indonesia selama periode 2022-2023. Dengan mengelompokkan data ke dalam kluster-kluster yang representatif, penelitian ini berusaha untuk mengidentifikasi perbedaan regional dan tren ekonomi yang ada.

Pendekatan berbasis data ini tidak hanya akan memperkaya literatur ekonomi dan sosial di Indonesia, tetapi juga menunjukkan bagaimana teknik data modern dapat diterapkan untuk menangani masalah-masalah ekonomi yang kompleks. Dengan demikian, analisis ini memiliki potensi untuk memberikan kontribusi signifikan dalam pengambilan keputusan yang lebih berbasis data di masa depan.

2. TINJAUAN PUSTAKA

Pengeluaran perkapita adalah ukuran yang dapat digunakan untuk menilai rata-rata pengeluaran individu dalam suatu populasi dan mencerminkan tingkat kesejahteraan ekonomi. Menurut Deaton (1997), pengeluaran per kapita memberikan wawasan tentang pola konsumsi dan distribusi pendapatan dalam masyarakat. Metode klustering, termasuk K-Means klustering, adalah teknik yang digunakan untuk mengelompokkan data ke dalam kluster-kluster berdasarkan kemiripan. K-Means klustering adalah algoritma yang efektif untuk mengelompokkan data besar dari beragam ke dalam kelompok-kelompok homogen. Jain (2010) menyatakan bahwa K-Means klustering adalah alat yang berguna dalam berbagai aplikasi, termasuk analisis data ekonomi. Sumber artikel ini ditemukan di [IEE Xplore] | (<https://ieeexplore.ieee.org/document/5288816>).

Clustering adalah teknik yang digunakan untuk mengelompokkan data dalam kelompok-kelompok berdasarkan kemiripan karakteristik. Clustering PCA menggabungkan kedua teknik tersebut, seperti K-Means. PCA juga merupakan teknik reduksi dimensi dimana berguna sebelum menerapkan kluster ataupun regresi berganda untuk menganalisis data yang kompleks.

Beberapa studi telah menggunakan K-Means klustering untuk berbagai analisis ekonomi. Misalnya, Kassambara (2017) menunjukkan penggunaan K-means klustering untuk mengelompokkan data pengeluaran berdasarkan pola konsumsi. Penelitian ini memberikan wawasan tentang bagaimana metode ini dapat diterapkan dalam konteks pengeluaran per kapita. Rstudio adalah perangkat lunak yang kuat untuk analisis statistik dan data. Menurut Kabacoff (2011), Rstudio menyediakan berbagai alat dan paket untuk analisis data yang kompleks, termasuk metode klustering. Panduan dan artikel ini diakses di [R-bloggers] | (<https://www.r-bloggers.com/2011/08/the-r-guide/>).

Studi mengenai pengeluaran per kapita di Indonesia telah dilakukan oleh beberapa peneliti. Pratama dan Hartono (2019) menganalisis pola pengeluaran per kapita di berbagai provinsi di Indonesia dan menemukan perbedaan signifikan antar provinsi yang dipengaruhi oleh faktor-faktor ekonomi dan sosial. Artikel ini diakses di [Jurnal Ekonomi dan Studi Pembangunan] | (<http://journal.unnes.ac.id/nju.index.php/DP/article/view/22591>).

Tinjauan pustaka ini menunjukkan bahwa pengeluaran perkapita adalah indikator penting dalam memahami kesejahteraan ekonomi, dan metode klustering seperti K-Means klustering adalah lat efektif untuk menganalisis data tersebut. Dengan menggunakan Rstudio, penelitian ini diharapkan dapat memberikan wawasan berharga tentang pola pengeluaran per kapita di Indonesia selama periode 2020-2023. Studi-studi terdahulu dan literature yang ada menyediakan dasar yang kuat untuk penelitian ini, dan penggunaan sumber-sumber terbuka memastikan aksesibilitas informasi bagi penelilti lainnya.

3. METODE PENELITIAN

Penelitian ini menggunakan metode Clustering untuk menganalisis segmentasi pengeluaran per kapita berdasarkan jenis kelamin di Indonesia selama periode 2022-2023. Data diperoleh dari <https://www.bps.go.id/id/statistics-table/2/NDYxlzl%3D/pengeluaran-per-kapita-yang-disesuaikan-menurut-jenis-kelamin.html> Badan Pusat Statistik (BPS).

1. Pengumpulan Data :
Data pengeluaran per kapita yang disesuaikan menurut jenis kelamin (dalam ribu rupiah per orang per tahun) diperoleh dari sumber Badan Pusat Statistik (BPS).
2. Pemrosesan Data :
Kemudian data akan dimuat dan diproses menggunakan aplikasi RStudio. Langkah ini mencakup pengimporan data dan persiapan data untuk analisis clustering.
3. Pemilihan Metode Clustering :
Satu metode clustering utama yang digunakan dalam penelitian ini adalah K-Means. Pemilihan metode didasarkan pada kenanpuan masing-masing algoritma utama mengelompokan data dengan efisien dan memberikan interpretasi yang jelas.
4. Implementasi Clustering di RStudio :
Data dianalisis menggunakan perangkat lunak RStudio. Algoritma K-Means diterapkan dengan menentukan jumlah klaster yang optimal menggunakan metode Elbow.
5. Evaluasi Hasil Clustering :
Hasil clustering di evaluasi dengan menggunakan metode metrik validasi internal seperti Silhouette Score untuk menilai kualitas klaster yang terbentuk.
6. Interpretasi :
Hasil Clustering diinterpretasikan untuk memahami pola pengeluaran per kapita berdasarkan jenis kelamin.

4. HASIL PENELITIAN DAN PEMBAHASAN

4.1 Statistika Deskriptif

```
> statdes<-summary(data1)
> statdes
Provinsi.Kabupaten.Kota laki2022 laki2023
Length:548 Min. : 4146 Min. : 4321
Class :character 1st Qu.:12458 1st Qu.:12953
Mode :character Median :14720 Median :15254
Mean :14818 Mean :15307
3rd Qu.:17043 3rd Qu.:17481
Max. :28077 Max. :28976

Perempuan2022 Perempuan2023
Min. : 1582 Min. : 1630
1st Qu.: 6440 1st Qu.: 6678
Median : 8418 Median : 8639
Mean : 8640 Mean : 8897
3rd Qu.:10277 3rd Qu.:10526
Max. :23626 Max. :23967
```

4.2 Pengumpulan Data

Data di ambil dari Badan Pusat Statistik (BPS). <https://www.bps.go.id/id/statistics-table/2/NDYxlzl%3D/pengeluaran-per-kapita-yang-disesuaikan-menurut-jenis-kelamin.html>

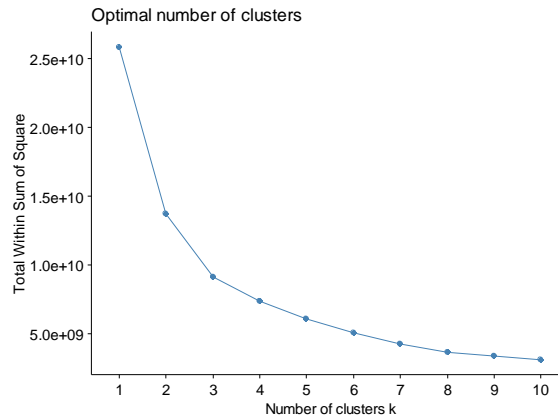
```
16 #import data excell RStudio
17 data1=read.delim("clipboard")
18 #melihat data setelah diimport
19 data1
20 options(max.print = 1000000)
21 data1
22 #summary of data
```

4.3 Library yang dibutuhkan

```
#install package excel untuk membaca data
> library(readxl)
> #install package Tidyverse untuk clustering
> library(tidyverse)
> #install package ggplot untuk membuat grafik dan visualisasi data
> library(ggplot2)
> #install package dplyr untuk menyederhanakan proses manipulasi
> library(dplyr)
> #install package cluster untuk clustering
> library(cluster)
> #install package factroextra untuk clustering dan visualisasi
> library(factoextra)
> #untuk analisis statistik dan psikometrik
> library(psych)
```

4.3 Implementasi Clustering K-Means di RStudio

```
> fviz_nbclust(dataku, kmeans, method = "wss")
```

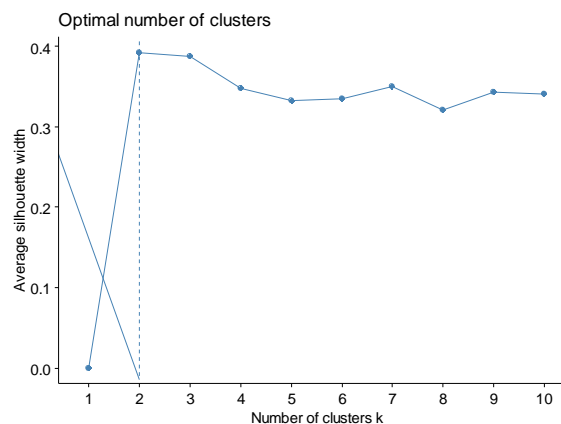


Pada hasil Implementasi Clustering K-Means dilakukan menggunakan Rstudio. Langkah pertama adalah memuat library yang dibutuhkan seperti pada bagian 4.2 untuk menganalisis data pengeluaran per kapita kemudian diimpor dan diproses analisis clustering. Metode Elbow digunakan untuk menentukan jumlah kluster yang optimal, ditunjukkan oleh grafik Total Within Sum of Square (WSS) terhadap jumlah kluster K.

Hasilnya menunjukkan titik optimal di mana penambahan kluster tidak secara signifikan mengurangi WSS lagi.

4.4 Evaluasi Hasil Clustering

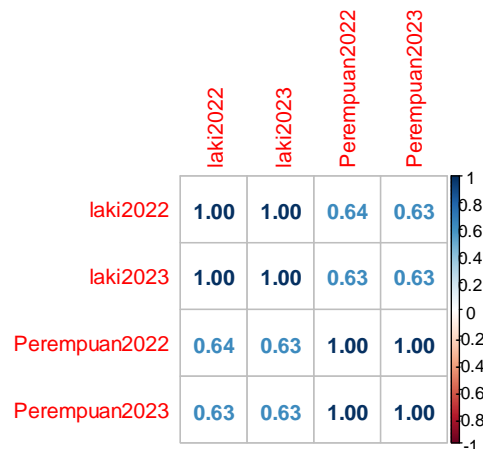
```
> fviz_nbclust(dataku, kmeans, method = "silhouette")
```



Evaluasi hasil clustering dilakukan menggunakan metode Silhouette Width terhadap jumlah cluster k, membantu dalam menentukan seberapa baik setiap data cocok dengan klasternya dibandingkan dengan kluster lain. Kluster yang baik akan memiliki silhouette.

4.5 Analisis Data Eksplanatori

4.5.1 Ekspolerasi Kolerasi



- Analisis korelasi dilakukan untuk mengidentifikasi hubungan antara variabel pengeluaran per kapita laki-laki dan perempuan.
- Korelasi antar variabel dihitung menggunakan metode pearson untuk memahami tingkat hubungan linear antara mereka.

4.6 Uji Asumsi

4.6.1 Uji Sampel Representatif

```
#uji asumsi
> library(psych)
> kmo <- KMO(data[,2:3:4:5])
warning messages:
1: In 2:3:4 : numerical expression has 2 elements: only the first used
2: In 2:3:4:5 : numerical expression has 3 elements: only the first used
> kmo
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = data[, 2:3:4:5])
Overall MSA = 0.48
MSA for each item =
Laki-laki      ...3 Perempuan      ...5
0.48          0.48          0.48          0.48
```

Uji KMO (Kaiser-Meyer-Olkin) dilakukan untuk memastikan kecukupan sampel. Nilai MSA (Measure of Sampling Adequacy) untuk setiap item adalah 0.48, menunjukkan kecukupan sampel yang tidak memadai.

4.6.2 Uji Non-Multikolinearitas

```
> korelasi <- cor(data[,2:3:4:5], method='pearson')
warning messages:
1: In 2:3:4 : numerical expression has 2 elements: only the first used
2: In 2:3:4:5 : numerical expression has 3 elements: only the first used
> korelasi
Laki-laki      ...3 Perempuan      ...5
Laki-laki 1.0000000 0.9991939 0.6403805 0.6386654
...3      0.9991939 1.0000000 0.6361847 0.6352928
```

```
Perempuan 0.6403805 0.6361847 1.0000000 0.9995568
...5      0.6386654 0.6352928 0.9995568 1.0000000
```

- Korelasi antar variabel diuji untuk mendeteksi multikolinearitas.
- Hasil menunjukkan korelasi tinggi antara beberapa variabel (seperti, pengeluaran laki-laki 2022 dengan 2023 yaitu, 0.999), menunjukkan kemungkinan adanya multikolinearitas.

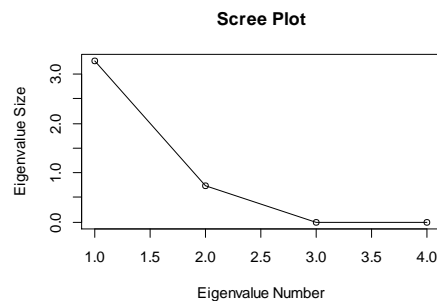
4.7 PCA

4.7.1 Dekomposisi Eigen(Matriks Kovarians)

```
> #standarisasi data
> sc <- scale(data[,2:3:4:5])
warning messages:
1: In 2:3:4 : numerical expression has 2 elements: only the first used
2: In 2:3:4:5 : numerical expression has 3 elements: only the first used
> sc
      laki2022      laki2023 Perempuan2022 Perempuan2023
[1,] -0.221572439 -0.2239579476 -0.045521907 -0.0454749590
[2,] -0.868761365 -0.8581896932 -1.598647945 -1.6153271911
[3,]  0.147571159  0.1651480644 -0.811227206 -0.8168812627
[4,] -0.978302842 -0.9714947109 -0.423496450 -0.4035168856
[5,] -0.728760652 -0.7227763793 -0.199733003 -0.1884430894
[6,] -0.168478356 -0.1714507443 -0.850566771 -0.8666093081
[7,] -0.855907009 -0.8700729024  0.344726580  0.3486198005
[8,] -0.082130611 -0.0863338042 -0.877317676 -0.8948921339
[9,] -0.558021259 -0.5580695731  0.177926824  0.1938412593
[10,] -0.438419851 -0.4431064332  0.155581951  0.1677340355
[11,] -0.599378755 -0.6149984357  0.142363857  0.1472212168
[12,] -0.938622001 -0.9325288390 -0.399263278 -0.3966792794
[13,]  0.017909819  0.0662134392 -0.480145424 -0.4569745344
[14,] -0.269915999 -0.2786759806 -0.242219734 -0.2319551291
[15,]  0.435676419  0.3992196498 -1.183222135 -1.1954360081
[16,] -0.455745289 -0.4229326129 -0.460947717 -0.4162596973
[17,]  0.244817163  0.2853619247 -0.288168346 -0.2673863614
[18,] -0.264327148 -0.2698326621  0.811765900  0.8284954383
[19,] -0.183568253 -0.2026787126  0.543312706  0.5444239791
[20,]  1.489733686  1.4374805074  2.160011481  2.1295054252
[21,] -0.090513887 -0.1194962484  0.712944912  0.7228233419
[22,]  0.351005329  0.3102337579  0.813024766  0.8107798221
[23,]  0.124097985  0.1369599869  0.764873138  0.7961722088
[24,] -0.680696535 -0.6213545708 -1.071183053 -1.0484274740
```

```
attr(,"scaled:center")
      laki2022      laki2023 Perempuan2022 Perempuan2023
      14817.909      15307.403         8639.644         8897.316
attr(,"scaled:scale")
      laki2022      laki2023 Perempuan2022 Perempuan2023
      3578.553      3618.551         3177.463         3217.500
```

Hasil di atas menunjukkan nilai rata-rata yang telah dinormalisasi dari masing-masing variabel. Nilai ini mengindikasikan bahwa rata-rata pengeluaran per kapita laki-laki dan perempuan mengalami peningkatan dari tahun 2022 ke 2023. Laki-laki memiliki pengeluaran per kapita yang lebih tinggi dibandingkan dengan perempuan di tahun 2022 dan 2023.

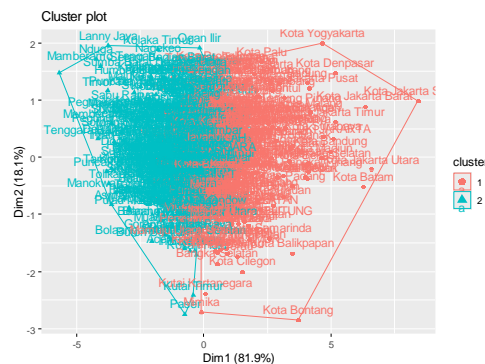


- Komponen pertama memiliki eigenvalue tinggi, menjelaskan sebagian besar variabilitas data.
- Eigenvalue menurun tajam setelah komponen tambahan memberikan kontribusi kecil terhadap variabilitas.

Ini berarti sebagian besar informasi data dapat dijelaskan dengan hanya beberapa komponen utama pertama.

4.8 Hasil Clustering

```
> fviz_cluster(kmeans_clustering, data = dataku)
```



Hasil clustering divisualisasikan menggunakan fungsi 'fviz_cluster' yang menunjukkan pengelompokan data pengeluaran per kapita berdasarkan kluster yang terbentuk. Visualisasi ini memberikan gambaran yang jelas tentang bagaimana data tersegmentasi dalam berbagai kluster, membantu hasil dari interpretasi yang sudah di analisis

Hasil clustering 1 dan 2, sebagai berikut :

1. Clustering 1

Kluster bagian 1 ini terdiri dari wilayah-wilayah dengan pengeluaran per kapita berdasarkan jeni kelamin periode 2022-2023 yang relative rendah.

Wilayah-wilayah dalam kluster ini menunjukkan pola pengeluaran yang homogeny, dengan nilai pengeluaran per kapita yang cenderung berada dibawah rata-rata nasional. Kluster ini mencakup daerah-daerah dengan tingkat kesejahteraan ekonomi yang lebih rendah.

2. Clustering 2

Kluster ini terdiri dari wilayah-wilayah dengan pengeluaran per kapita yang lebih tinggi dari pada Clustering pertama.

Wilayah-wilayah dalam kluster ini menunjukkan pola pengeluaran yang homogen, dengan nilai pengeluaran per kapita yang berada diatas rata-rata nasional. Kluster ini mencakup daerah-daerah dengan tingkat kesejahteraan ekonomi yang lebih tinggi, ini perkotaan atau daerah dengan pengembangan ekonomi yang lebih pesat.

Evaluasi dan hasil dari Clustering ini memberikan wawasan penting tentang distribusi pengeluaran per kapita di Indonesia, membantu pembuat kebijakan dalam merancang startegi yang lebih tepat untuk meningkatkan kesejahteraan ekonomi dan mengurangi kesenjangan regional.

5 KESIMPULAN

Kesimpulan dari analisis ini adalah bahwa metode Clustering K-Means dan analisis komponen utama (PCA) berhasil mengelompokkan dan mengidentifikasi pola pengeluaran per kapita di Indonesia selama 2022-2023. Pengelompokan K-Means menunjukkan adanya segmentasi yang signifikan dalam pengeluaran per kapita berdasarkan wilayah.

Wilayah dengan pengeluaran per kapita yang lebih tinggi cenderung berada di atas rata-rata nasional, mencakup daerah perkotaan atau perkembangan ekonomi yang lebih pesat, sementara wilayah dengan pengeluaran per kapita yang lebih rendah berada dibawah rata-rata nasional, menunjukkan tingkat kesejahteraan ekonomi yang rendah.

Analisis PCA digunakan untuk mengurangi dimensi data dan mengidentifikasi faktor-faktor utama yang mempengaruhi pengeluaran per kapita. Hasil PCA menunjukkan bahwa sebagian besar variasi dalam data dapat dijelaskan oleh beberapa komponen utama, yang membantu dalam memahami struktur data yang kompleks dan mengarahkan analisis lebih lanjut.

Hasil analisis ini memberikan wawasan yang berharga bagi pembuat kebijakan dan peneliti ekonomi dalam merancang strategi yang lebih tepat untuk meningkatkan kesejahteraan ekonomi dan mengurangi kesenjangan regional di Indonesia. Penelitian ini juga memperlihatkan potensi aplikasi Clustering K-Means dan PCA dalam analisis data ekonomi yang kompleks, yang dapat digunakan sebagai dasar ekonomi dasar untuk pengambilan keputusan yang lebih berbasis data.

6 DAFTAR PUSTAKA

Deaton, Angus S. | 10 Juli 1997 | The Analysis of Household Surveys : A Microeconomic Approach to Development Policy (English). Washington, D.C. :World Bank Group. | <https://documents.worldbank.org/curated/en/5938777303124/The-Analysis-of-Household-Surveys-A-Microeconomic-Approach-to-Development-Policy>

Kabacof, R.I | 2011 | R in Action : Data Analysis and Graphic with R | http://www.cs.uni.edu/~jacobson/4772/week11/R_in_Action.pdf .

Pratama, A. dan Hartono, D. | 2019 | Analisis Pola Pengeluaran Per Kapita di Indonesia | jurnal ekonomi dan studi pembangunan | <http://journal.unnes.ac.id/nju.index.php/DP/article/view/22591>.