

**LAPORAN PROJECT AKHIR DATA SCIENCE
CLUSTERING TINGKAT PERKEMBANGAN DESA/KELURAHAN DI
PROVINSI LAMPUNG MENGGUNAKAN
ALGORITMA K-MEANS**



DISUSUN OLEH:

ANDITA AYU SAFITRI

123200118

MAULANA ARYA WISNU W.

123200138

**PROGRAM STUDI INFORMATIKA
JURUSAN INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”
YOGYAKARTA
2022**

**HALAMAN PENGESAHAN
LAPORAN PROYEK AKHIR**

Disusun oleh :

Andita Ayu Safitri

123200118

Maulana Arya Wisnu W.

123200138

Telah Diperiksa dan Disetujui oleh Asisten Praktikum Data Science

Pada Tanggal : 7 Desember 2022.

Asisten Praktikum

Asisten Praktikum



Dhea Anggita
123190046



Rico Aminanda
123190076

DAFTAR ISI

HALAMAN PENGESAHAN LAPORAN PROYEK AKHIR	1
DAFTAR ISI	2
1. PENDAHULUAN	3
2. METODE	3
2.1 Data Selection	4
2.2 Data Integration	4
2.3 Data Transformation	4
2.4 Data Mining	4
2.5 Pattern Evaluation	5
2.6 Knowledge Presentation	5
3. HASIL DAN PEMBAHASAN	5
3.1 Code Program	5
Poin 1. Tahap Data Selection	5
Poin 2. Tahap Pembersihan Data	7
Poin 3. Tahap Data Integration	8
Poin 4. Tahap Data Transformation	8
Poin 5. Tahap Data Mining	9
Poin 6. Tahap Pattern Evaluation	12
Poin 7. Tahap Knowledge Presentation	15
3.2 Tampilan Gui	16
3.3 Hasil Proyek	18
4. KESIMPULAN	18

1. PENDAHULUAN

Desa/kelurahan merupakan elemen terkecil namun diakui keberadaannya oleh suatu Pemerintah Negara Kesatuan Republik Indonesia. Desa/kelurahan dapat didefinisikan sebagai kesatuan masyarakat hukum yang memiliki batas wilayah yang berwenang untuk mengatur dan mengurus urusan pemerintahan, yang mana pemerintah sebagai pemangku kepentingan masyarakat setempat berdasarkan prakarsa masyarakat, bahkan hak asal usul dan atau hak tradisionalnya tetap diakui dan dihormati dalam sistem pemerintahan tersebut.

Pemerintahan dalam lingkup desa/kelurahan ini dipimpin oleh seorang kepala desa yang memiliki wewenang dalam membuat ketetapan dalam mengatur kehidupan masyarakatnya. Dengan catatan, ketetapan tersebut mampu menunjang kemajuan desa yang dipimpinnya. Tidak hanya itu, kepala desa juga dituntut memiliki suatu program yang berorientasi kepada peningkatan ekonomi dan kesejahteraan masyarakat desa/kelurahan itu sendiri. Dalam pelaksanaannya, kepala desa dibantu dengan perangkat desa.

Suatu desa pasti memiliki banyak potensi-potensi yang dapat dikembangkan. Baik dalam bidang pemerintah, bidang kewilayahan, maupun bidang kemasyarakatannya. Tingkat kemajuan perkembangan masing-masing desa/kelurahan pun berbeda-beda, tergantung bagaimana pemerintah desa/kelurahan beserta masyarakatnya mengelola sumber daya dan fasilitas yang ada dengan baik dan benar agar dapat memaksimalkan hasil yang kemudian dapat mencapai suatu kemajuan dan kesejahteraan seperti yang diharapkan.

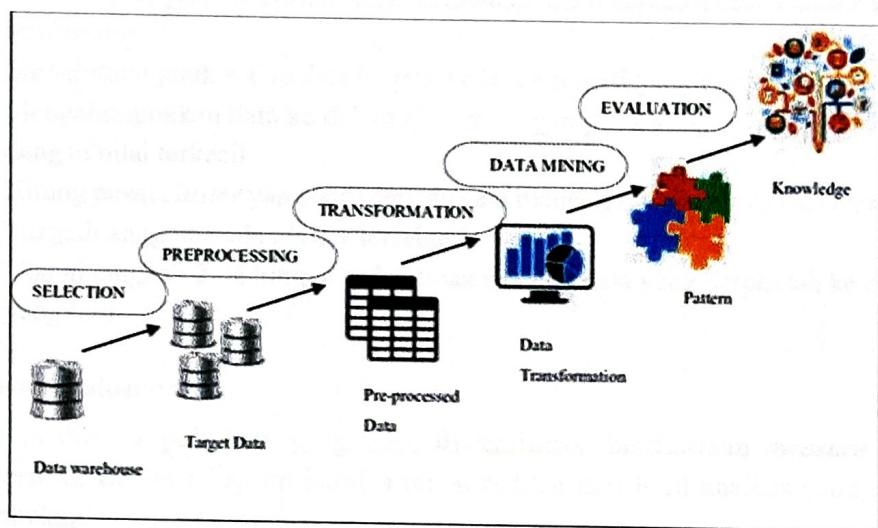
Provinsi Lampung merupakan suatu wilayah yang memiliki 1594 desa/kelurahan di dalamnya berdasarkan data administratif yang diperoleh dari portal database Kementerian Dalam Negeri Republik Indonesia. Provinsi lampung ini memiliki banyak kekayaan alam dan berbagai potensi lainnya. Dibutuhkan strategi dan manajemen pengelolaan pemerintah daerah yang baik untuk mengelola begitu banyak desa/kelurahan yang memiliki sekitar 9 - 10 juta penduduk per tahun 2022.

Untuk mengetahui tingkat perkembangan desa/kelurahan di Provinsi Lampung, maka dibutuhkan suatu algoritma untuk melakukan clusterisasi pada data desa/kelurahat tersebut. Untuk itu kami membuat project data science ini dengan judul “Clustering Tingkat Perkembangan Desa/Kelurahan di Provinsi Lampung menggunakan Algoritma K-Means”.

2. METODE

Metode yang akan kami gunakan dalam menyelesaikan project data science ini yaitu Metode KDD (*Knowledge Discovery In Databases*). Metode ini merupakan proses data mining dalam upaya menyelesaikan suatu persoalan atau permasalahan yang ada, dimana keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. Dalam hal ini, kami akan mengclusteringkan tingkat perkembangan desa/kelurahan dengan memanfaatkan algoritma K-Means. Algoritma

K-Means adalah algoritma yang digunakan untuk clustering data berdasarkan pusat data(*centroid*) atau yang biasa disebut Mean (nilai rata-rata).



Gambar 2.1. Metode KDD (*Knowledge Discovery In Databases*)

2.1 Data Selection

Tahapan untuk membersihkan data dengan cara mengolah lalu memilih data yang dianggap bisa dipakai dan relevan dengan masalah yang akan diselesaikan. Dalam hal ini, kami tidak menggunakan seluruh data yang ada, kami akan melakukan penyaringan yaitu hanya menggunakan data dari provinsi lampung dan akan mengambil data yang dapat diolah menggunakan algoritma K-Means yaitu kolom yang memiliki struktur data berupa numerik.

2.2 Data Integration

Proses menggabungkan data yang dianggap berulang serta data-data yang sudah bersih menjadi satu agar dapat dengan mudah dilakukan analisis pada tahap selanjutnya. Dalam hal ini kami hanya menggunakan tiga kolom dari dataset tingkat perkembangan desa/kelurahan yaitu bidang pemerintahan, bidang kewilayahan, dan bidang kemasyarakatan.

2.3 Data Transformation

Proses transformasi data terpilih ke dalam bentuk mining procedure agar didapatkan data sesuai apa yang diharapkan dalam menunjang analisis data pada proses clustering tersebut. Dalam hal ini kami melakukan transformasi data menggunakan Metode *Scaling Data*.

2.4 Data Mining

Proses dimana dilakukan berbagai teknik untuk mengekstrak pola-pola potensial menghasilkan data yang berguna. Dalam hal ini, kami menggunakan Algoritma K-Means yang memanfaatkan tiga metode untuk menentukan jumlah cluster yaitu

Metode *Elbow* atau *WSS*, Metode *Silhouette*, dan Metode *Gap Statistic*. Langkah-langkah penyelesaian Algoritma K-Means sebagai berikut:

- a) Menentukan jumlah *cluster* (*K*), kemudian menentukan pusat *cluster* secara sembarang
- b) Menghitung jarak setiap data ke pusat *cluster* seperti yang ditentukan
- c) Mengelompokkan data ke dalam *cluster* dengan jarak yang paling pendek atau dengan nilai terkecil
- d) Hitung pusat *cluster* yang baru dengan cara mencari nilai rata-rata dari data yang menjadi anggota pada *cluster* tersebut
- e) Ulangi langkah 2 - 4 hingga sudah tidak ada lagi data yang berpindah ke *cluster* yang lain.

2.5 Pattern Evaluation

Proses dimana pola-pola yang telah diidentifikasi berdasarkan *measure* yang diberikan. Dalam tahap ini barulah mulai terlihat dari hasil analisis yang sudah dilakukan.

2.6 Knowledge Presentation

Proses terakhir yaitu data-data yang sudah diproses akan disosialisasikan agar lebih mudah dipahami oleh pengguna dan diharapkan bisa didapatkan hasil analisis berdasarkan proses yang sudah dilakukan.

3. HASIL DAN PEMBAHASAN

3.1 Code Program

Poin 1. Tahap Data Selection

Pertama yaitu memilih dan menampilkan dataset yang akan digunakan, kemudian mengidentifikasi *library*. Dalam hal ini, kami menggunakan dataset tingkat perkembangan desa melalui portal Kementerian Dalam Negeri Republik Indonesia dalam bentuk format excel.

Dataset yang digunakan dalam penelitian ini adalah dataset tingkat perkembangan desa yang merupakan data yang diambil dari portal Kementerian Dalam Negeri Republik Indonesia. Data yang digunakan dalam penelitian ini merupakan data yang masih dalam bentuk raw data yang belum diproses.



```

```{r}
library(tidyverse) #data tidying
library(dplyr) #untuk memanipulasi data
library(here) #untuk menyimpan dataset
library(ggplot2) #untuk memvisualisasikan data
library(cluster) #untuk menganalisis kluster
library(factoextra) #mengekstrak dan memvisualisasikan hasil
analisis
library(shiny) #untuk membuat tampilan GUI
library(kernlab)
library(readxl) #untuk membaca file excel
#menampilkan dataset dengan variabel objek data
here()
data1<-read_excel(here("tingkatperkembangandesa.xlsx"))
data1
```

```

Listing 3.1 Library

The screenshot shows a data frame titled "tingkatperkembangandesa.xlsx" with 24,581 rows and 11 columns. The columns are labeled: TAHUN PROVINSI, KAB/KOTA, KECAMATAN, DESA, and STATUS. The data consists of 2021 entries for ACEH, with various sub-districts and names listed under each category.

| TAHUN PROVINSI | KAB/KOTA | KECAMATAN | DESA | STATUS |
|----------------|------------------------|--------------------|------------------|--------|
| 2021 ACEH | KABUPATEN ACEH SELATAN | KLUET UTARA | PULO IE I | DESA |
| 2021 ACEH | KABUPATEN ACEH SELATAN | KLUET UTARA | KRUENG BATU | DESA |
| 2021 ACEH | KABUPATEN ACEH SELATAN | KLUET UTARA | KAMPUNG TIMANGGI | DESA |
| 2021 ACEH | KABUPATEN ACEH SELATAN | KLUET UTARA | SUAQ CEURINGCENG | DESA |
| 2021 ACEH | KABUPATEN ACEH SELATAN | KLUET SELATAN | PASI MEURAPAT | DESA |
| 2021 ACEH | KABUPATEN ACEH SELATAN | LABUHAN HAJI | PASAR LAMA | DESA |
| 2021 ACEH | KABUPATEN ACEH SELATAN | MEUKEX | ALUE MEUTUAH | DESA |
| 2021 ACEH | KABUPATEN ACEH SELATAN | TRUMON | UJONG TANOH | DESA |
| 2021 ACEH | KABUPATEN ACEH SELATAN | PASI RAJA | UJONG BATEE | DESA |
| 2021 ACEH | KABUPATEN ACEH SELATAN | LABUHAN HAJI TIMUR | KEUMUMU HILIR | DESA |

Gambar 3.1 Dataset sebelum difilter

Namun, dari banyaknya data yang kami dapatkan. Hal pertama yang harus dilakukan yaitu memilih data yang akan diolah dengan jangkauan yang lebih sempit. Kami akan mengolah dataset tersebut untuk mendapatkan clustering tingkat perkembangan dari desa/kelurahan di Provinsi Lampung saja.

```

```{r}
#menampilkan struktur dataset
str(data1)
#menampilkan ringkasan deskriptif data
summary(data1)
#mempersempit jangkauan data analisis dengan menyaring data
data2 <- data1 %>% filter(data1$PROVINSI == "LAMPUNG")
data2
```

```

Listing 3.2 Menampilkan Data Set

The screenshot shows the RStudio interface with the 'R Console' tab selected. The output window displays the following information:

```

tbl_df [4,581 x 11] (S3:tbl_df/tbl/data.frame)
  $ TAHUN : num [1:4581] 2021 2021 2021 2021 2021 ...
  $ PROVINSI : chr [1:4581] "ACEH" "ACEH" "ACEH" ...
  $ KAB_KOTA : chr [1:4581] "KABUPATEN ACEH SELATAN" "KABUPATEN ACEH SELATAN" "KABUPATEN ACEH SELATAN" ...
  $ KECAMATAN : chr [1:4581] "KLUET UTARA" "KLUET UTARA" "KLUET UTARA" ...
  $ DESA : chr [1:4581] "TULU IE II" "KRILING BATU" "KAMPUNG TENGGE" "SUAR GEURINGGENG" ...
  $ STATUS : num [1:4581] 0 0 0 0 0 ...
  $ BIDANG PEMERINTAHAN : num [1:4581] 160 193 199 715 214 217 190 0 212 49 ...
  $ BIDANG KEWILAYAHAN : num [1:4581] 0 0 85 75 86 76 58 0 67 0 ...
  $ BIDANG KEMASYARAKATAN : num [1:4581] 0 0 193 187 190 182 191 0 194 0 ...
  $ NILAI : chr [1:4581] "160" "193" "197" "715" "214" ...
  $ TINGKAT PEREMBANGAN : chr [1:4581] "Kurang Berkembang" "Kurang Berkembang" "Kurang Berkembang" ...
  $ TAHUN PROVINSI KAB/KOTA KECAMATAN DESA STATUS BIDANG PEMERINTAHAN BIDANG KEWILAYAHAN BIDANG KEMASYARAKATAN
  Min. :2021 Length:24581 Min. :0 Length:24581 Min. :0 Length:24581 Min. :0 Length:24581
  1st Qu.:2021 Class :character Class :character Class :character Class :character Class :character
  Median :2021 Mode :character Mode :character Mode :character Mode :character Mode :character
  Mean :2021
  3rd Qu.:2021
  Max. :2021
  
```

Below this, there is a summary of the 'BIDANG PEMERINTAHAN' column:

| | BIDANG PEMERINTAHAN | BIDANG KEWILAYAHAN | BIDANG KEMASYARAKATAN | NILAI | TINGKAT PEREMBANGAN |
|---------|---------------------|--------------------|-----------------------|-------|---------------------|
| Min. | 0 | 0 | 0 | 0 | Kurang Berkembang |
| 1st Qu. | 59.00 | 59.00 | 59.00 | 59.00 | Kurang Berkembang |
| Median | 69.00 | 69.00 | 69.00 | 69.00 | Kurang Berkembang |
| Mean | 179.8 | 64.62 | 64.62 | 179.8 | Kurang Berkembang |
| 3rd Qu. | 81.00 | 81.00 | 81.00 | 81.00 | Kurang Berkembang |
| Max. | 224 | 224 | 224 | 224 | Kurang Berkembang |

Gambar 3.2 Struktur dan ringkasan dataset

The screenshot shows the RStudio interface with the 'R Console' tab selected. The output window displays the following table:

| TAHUN | PROVINSI | KAB/KOTA | KECAMATAN | DESA | STATUS | BIDANG PEMERINTAHAN | BIDANG KEWILAYAHAN | BIDANG KEMASYARAKATAN |
|-------|----------|---------------------------|-----------|--------------|--------|---------------------|--------------------|-----------------------|
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | HAJIRENA | DESA | 225 | 90 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | SIDOSARI | DESA | 185 | 67 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | PEMANGLILAN | DESA | 178 | 65 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | NATAR | DESA | 174 | 57 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | MJARA PUTIH | DESA | 211 | 80 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | NEGARA RATU | DESA | 208 | 69 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | MERAK BATIN | DESA | 207 | 90 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | TANJUNG SARU | DESA | 190 | 61 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | BUMI SARI | DESA | 215 | 78 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | CANDI MAS | DESA | 209 | 80 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | BRANTI RAYA | DESA | 210 | 83 | |
| 2021 | LAMPUNG | KABUPATEN LAMPUNG SELATAN | NATAR | HADUYANG | DESA | 189 | 58 | |

Gambar 3.2.1 Dataset setelah di filter

Poin 2. Tahap Pembersihan Data

Kemudian, membersihkan dataset desa/kelurahan di seluruh Indonesia yang hanya menyisakan kolom penilaian dalam bidang pemerintahan, bidang kewilayahana, serta bidang kemasyarakatan.

```

```{r}
#menghapus data yang tidak digunakan dan mengubah nama kolom
databaru <- data2 %>% select(-1:-6,-10) %>%
 rename(bidang_pemerintahan = 'BIDANG PEMERINTAHAN',
 bidang_kewilayahana = 'BIDANG KEWILAYAHAN',
 bidang_kemasyarakatan = 'BIDANG KEMASYARAKATAN')
```

```

Listing 3.3 Menghapus Data Tidak Digunakan

Poin 3. Tahap Data Integration

Kami menjadikan satu tiga kolom yang akan digunakan untuk memudahkan dalam melakukan analisis pada proses selanjutnya.

```
#menampilkan kolom yang digunakan saja dalam bentuk data frame  
datadesa = data.frame(databaru)  
datadesa
```

Listing 3.4 Menampilkan Kolom Data Frame

| bidang_pemerintahan | bidang_kewilayahann | bidang_kemasyarakatan |
|---------------------|-------------------------|-----------------------|
| 225 | 90 | 197 |
| 185 | 67 | 174 |
| 178 | 65 | 169 |
| 174 | 57 | 175 |
| 211 | 80 | 179 |
| 208 | 69 | 197 |
| 207 | 90 | 181 |
| 190 | 61 | 164 |
| 216 | 78 | 175 |
| 209 | 80 | 184 |
| 210 | 83 | 191 |
| 1-11 of 1,594 rows | | |
| | 1 2 3 4 5 6 ... 97 Next | |

Gambar 3.3 Data pilihan yang akan diolah

Poin 4. Tahap Data Transformation

Scaling data yaitu berdasarkan ketiga data di atas kemudian akan dilakukan pengurangan nilai tersebut dengan pusat cluster(centroid/mean), lalu dibagi dengan standar deviasinya.

```
```{r}  
#melakukan transformasi data menggunakan scaling data : nilai -
mean / standar deviasi
datafix <- scale(datadesa)
datafix

#visualisasi data transformasi scaling
boxplot(datafix)

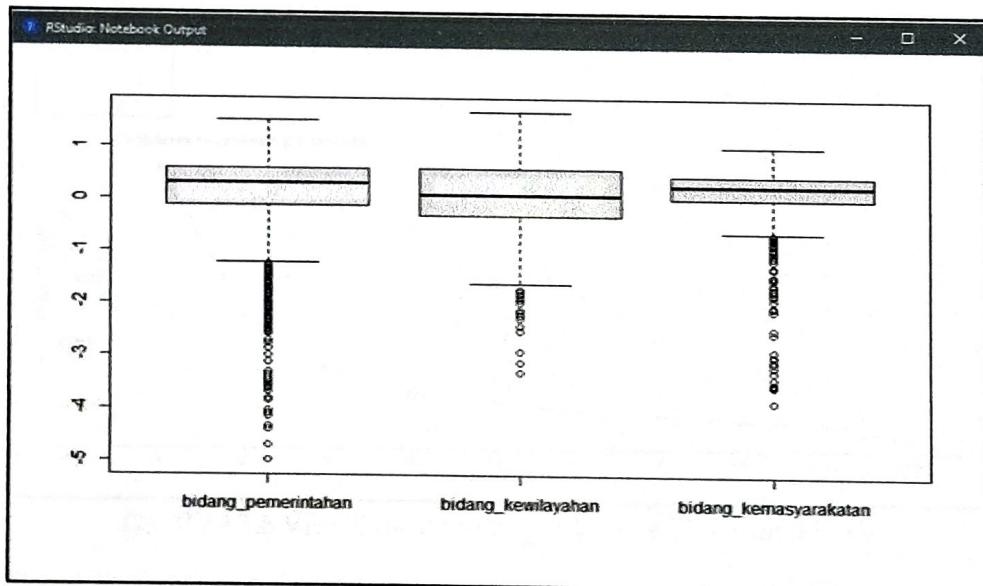
#menampilkan data baru dalam bentuk data frame
view(datadesa)
```
```

Listing 3.5 Transformasi Data

R Console

```
[312,] -0.360177617 1.07132992 -0.073489526
[313,] -1.606621998 -0.17676975 0.565012422
[314,] -2.013071253 -0.02699779 0.679030627
[315,] -2.256940806 0.27254613 -0.096293167
[316,] 0.235947957 0.67193803 0.268565089
[317,] -1.308559211 1.37087384 -1.646940755
[318,] -1.335655828 0.47224208 -1.692548037
[319,] -1.769201700 -0.97555354 -0.529562346
[320,] -1.877588168 0.07285018 -0.734795115
[321,] 1.319812637 1.67041776 0.907067038
[322,] 0.019175021 0.17269816 -2.034602652
[323,] -0.793723488 -1.02547753 -1.510118909
[324,] -1.145979509 -0.92562955 -0.301525936
[325,] -0.549853935 1.27102587 -1.304686140
[326,] -0.929206573 -0.42638968 -1.031242448
[327,] -0.929206573 1.17117790 0.245761448
[328,] 1.211426169 1.57056979 0.975477961
[329,] -1.823394934 0.57209005 -0.233115013
[330,] -1.687911849 -0.37646570 0.063332320
[331,] -1.010496424 0.97148195 -0.233115013
[332,] -1.742105083 -0.92562955 0.268565089
[333,] -1.091786275 -0.02699779 -0.529562346
[ reached getOption("max.print") -- omitted 1261 rows ]
attr(,"scaled:center")
  bidang_pemerintahan      bidang_kewilayahann      bidang_kemasyarakatan
       184.29235           66.54078          169.22271
attr(,"scaled:scale")
  bidang_pemerintahan      bidang_kewilayahann      bidang_kemasyarakatan
        36.90498            20.03045            43.85265
```

Gambar 3.4 Data yang sudah ditransformasi



Gambar 3.5 Visualisasi data transformasi dengan boxplot

Poin 5. Tahap Data Mining

Setelah dilakukan transformasi data, kemudian yang perlu dilakukan yaitu pemodelan dalam bentuk data mining menggunakan 3 metode yang telah ditentukan sebelumnya.

```

```{r}
memperkirakan jumlah cluster yang optimal dalam bentuk
visualisasi

METODE ELBOW atau WSS
dilihat dari grafik yang mulai landai dari yang pertama
setelah curam atau titik siku, hasil wss yaitu k = 3
fviz_nbclust(datafix, kmeans, method = "wss")

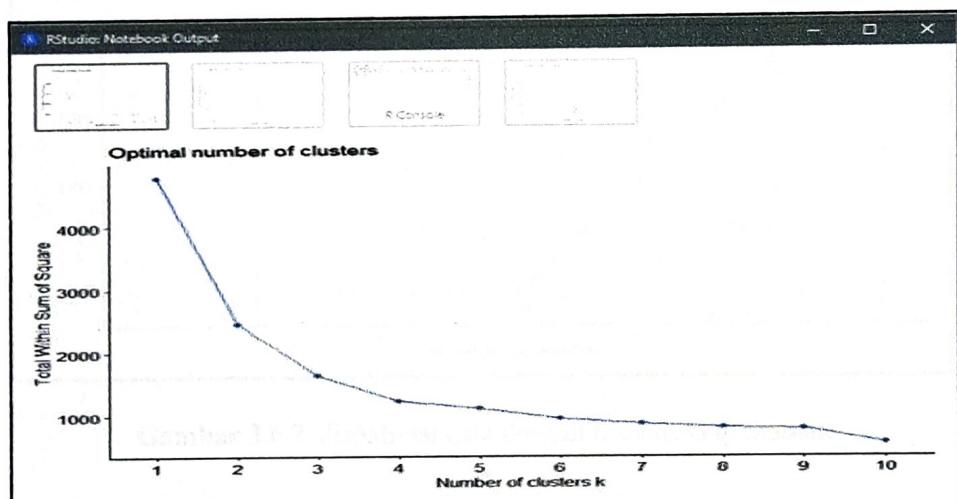
METODE SILHOUETTE
dilihat dari garis tertinggi yaitu K = 2
fviz_nbclust(datafix, kmeans, method = "silhouette")

METODE GAP STATISTIC
dilihat dari titik tertinggi yaitu K = 3
dim(datadesa)
set.seed(9999) #Mengunci data, nilainya ditentukan sendiri

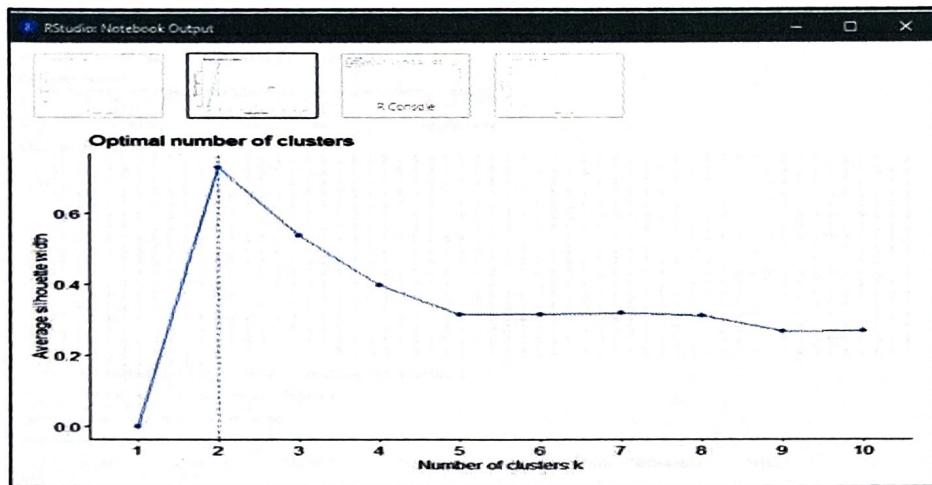
gap_stat <- clusGap(datafix, FUN=kmeans, nstart=300, K.max=3, B=
1594)
fviz_gap_stat(gap_stat)
```

```

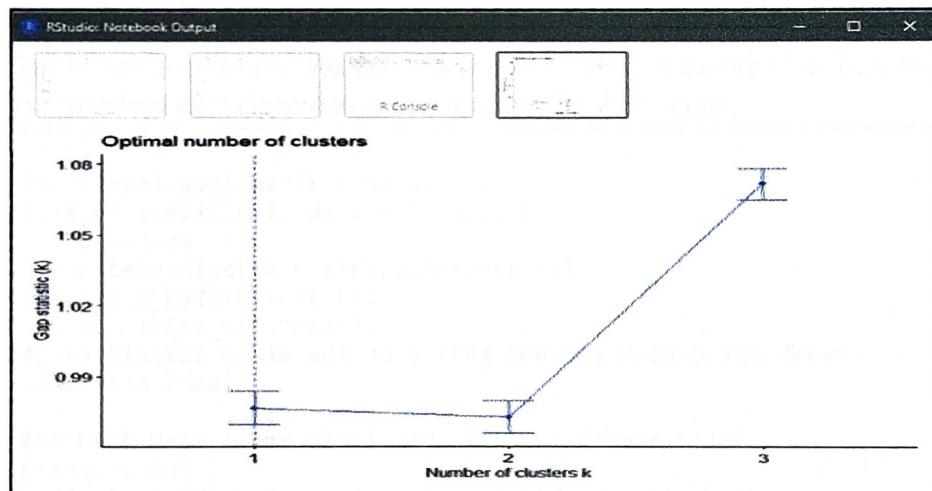
Listing 3.6 Memperkirakan Jumlah Cluster dalam Visual



Gambar 3.6 Visualisasi data dengan Metode Elbow atau WSS



Gambar 3.6.1 Visualisasi data dengan Metode Silhouette



Gambar 3.6.2 Visualisasi data dengan Metode Gap Statistic

Jadi dari 3 metode menghasilkan jumlah cluster dominan menuju nilai 3, sehingga K yang digunakan yaitu 3. Selanjutnya yaitu memproses data mining dengan Algoritma K-Means berdasarkan 3 cluster yang didapatkan dari perhitungan sebelumnya.

```
```{r}
final <- kmeans(datafix, 3)
final
```
```

Listing 3.7 Final Analisis Kmeans

Gambar 3.7 Hasil analisis menggunakan metode K-Means

Poin 6. Tahap Pattern Evaluation

Dari tahap-tahap sebelumnya didapatkan pola pola dalam bentuk visualisasi data yang nantinya akan digunakan dalamvisualisasi hasil clustering.

```

```{r}
#memvisualisasi hasil klasterisasi
fviz_cluster(final, data = datafix)
 datadesa %>%
 mutate(Cluster = final$cluster) %>%
 group_by(Cluster) %>%
 summarise_all("mean")
#jika kluster masih ada yg saling menimpa maka belum dapat
dikatakan ideal

#membuat data frame hasil clusterisasi dengan nilai
transformasi
 finaldata=data.frame(datafix, final$cluster)
View(finaldata)

#visualisasi lain dalam bentuk persebaran
pairs(datafix,col=c(1:3)[final$cluster],pch=16)

```

### **Listing 3.8 Visualisasi Hasil Klasterisasi**

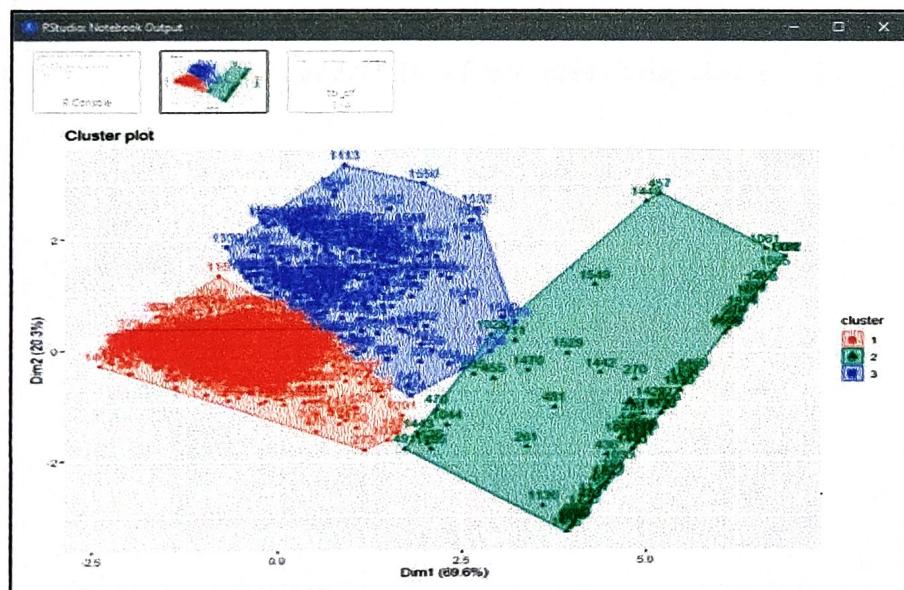
Project Prak Data Science.Rmd\* finaldata datadesa

Filter

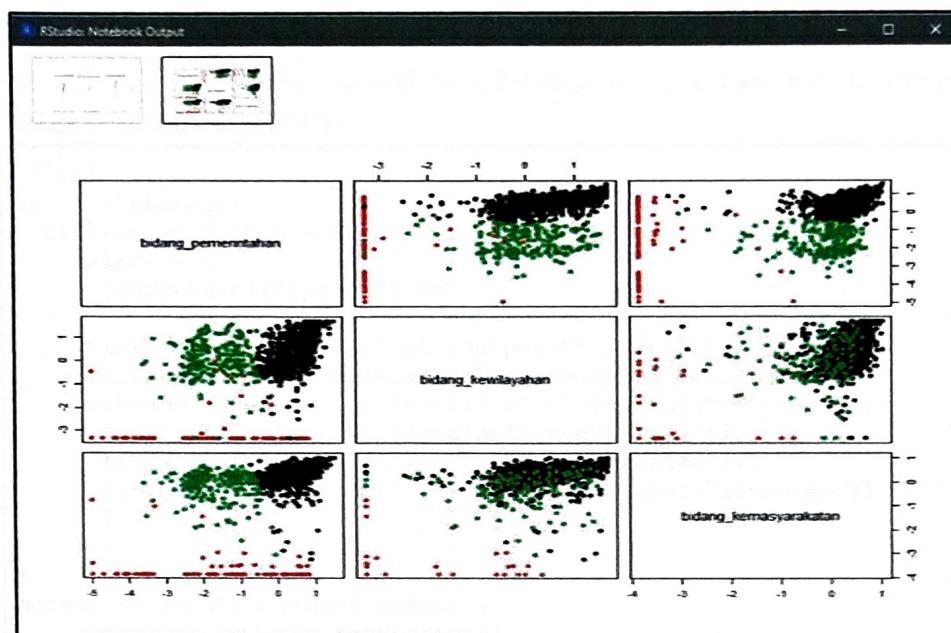
	bidang_pemerintahan	bidang_kewilayahann	bidang_kemasyarakatan	final.cluster
1	1.103039701	1.17117790	0.633423345	1
2	0.019175021	0.02292620	0.103939602	1
3	-0.170501296	-0.07692178	-0.005078603	1
4	-0.278687766	-0.47631367	0.131743243	1
5	0.723687063	0.67193803	0.222957807	1
6	0.642397212	0.12277417	0.633423345	1
7	0.615300595	1.17117790	0.268565069	1
8	0.154658106	-0.27661772	-0.119096808	1
9	0.859170148	0.57209005	0.131743243	1
10	0.669493829	0.67193803	0.336976012	1
11	0.696590446	0.82170999	0.496601499	1
12	0.127561489	-0.42638968	-0.050685863	1
13	-0.007921596	-0.17676975	0.496601499	1
14	0.235947957	-0.07692178	0.587816063	1
15	0.506914127	0.17269916	0.610619704	1
16	1.130136318	1.37087384	0.359779653	1
17	0.344334425	1.22110168	0.359779653	1
18	0.235947957	0.27254613	0.519405140	1
19	0.804976914	0.77178600	0.565012422	1
20	0.696590446	0.27254613	0.542208781	1
21	0.615300595	0.62201404	0.200154166	1
22	0.534010744	0.82170999	0.382583284	1
23	0.750783680	0.37239411	0.359779653	1

Showing 1 to 24 of 1,594 entries, 4 total columns

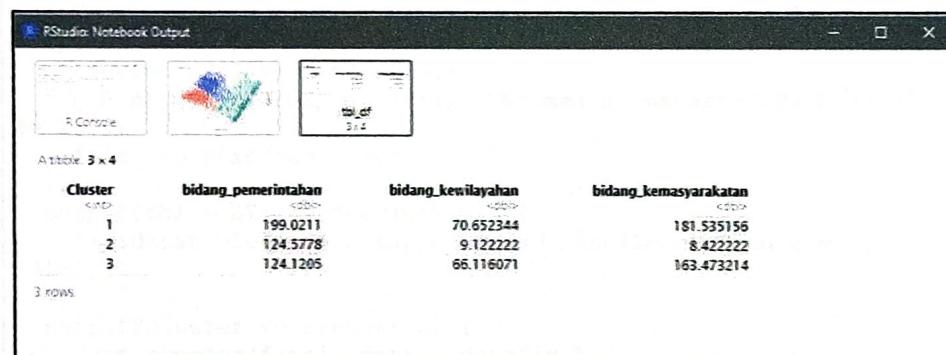
Gambar 3.8 Data frame hasil dari clustering



Gambar 3.9 Visualisasi data dari hasil clustering



**Gambar 3.9.1** Visualisasi data cluster dalam bentuk persebaran



**Gambar 3.9.2** Data frame rata-rata tiap cluster

### **Poin 7. Tahap Knowledge Presentation**

Visualisasi data yang akan memudahkan dalam penyampaian ke orang lain yaitu dalam bentuk tampilan GUI

```
```{r}
ui <- fluidPage(
  titlePanel("HASIL ANALISIS CLUSTERING"),
  mainPanel(
    tabsetPanel(type = "tabs",
      tabPanel("Metode WSS" plotOutput("metode1")),
      tabPanel("Metode Silhouette",plotOutput("metode2")),
      tabPanel("Metode Gap Statistic" plotOutput("metode3")),
      tabPanel("Data", DT::dataTableOutput('tbl')),
      tabPanel("Clustering", plotOutput("cluster")),
      tabPanel("Finalisasi",DT::dataTableOutput("simpulan"))
    )
  )
)
server <- function(input,output){
  output$metode1 <- renderPlot({
    fviz_nbclust(datafix, kmeans, method = "wss")
  })
  output$metode2 <- renderPlot({
    fviz_nbclust(datafix, kmeans, method = "silhouette")
  })
  output$metode3 <- renderPlot({
    gap_stat <-clusGap(datafix, FUN=kmeans, nstart=300, K.max=3,
B= 1594)
    fviz_gap_stat(gap_stat)
  })
  output$tbl = DT::renderDataTable({
    DT::datatable(finaldata, options= list(lengthChange =
FALSE))
  })
  output$cluster <- renderPlot({
    fviz_cluster(final, data = datafix )
  })
  output$simpulan <- DT::renderDataTable({
    d <- datadesa %>%
      mutate(Cluster = final$cluster) %>%
      group_by(Cluster) %>%
      summarise_all("mean")

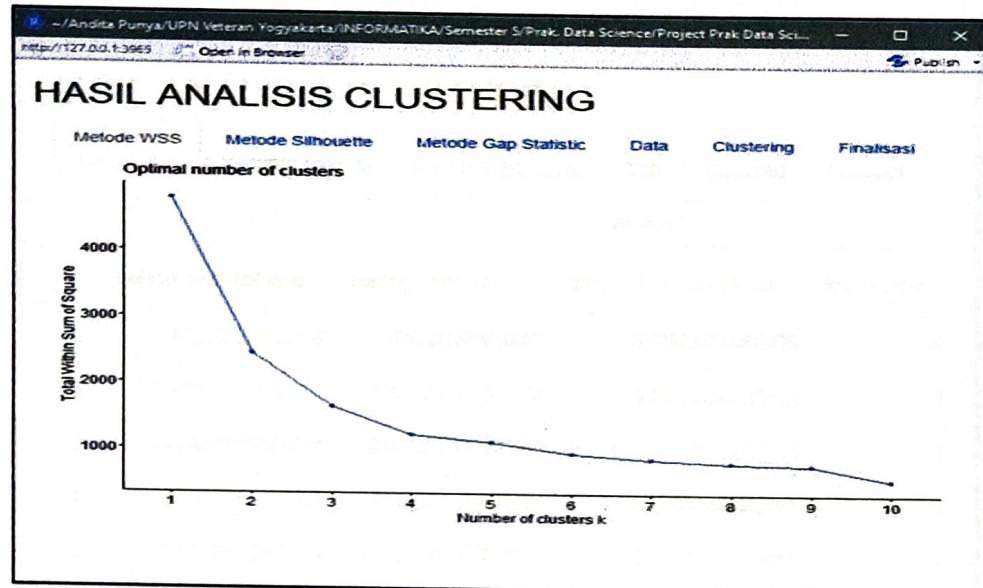
    DT::datatable(d, options= list(lengthChange = FALSE))
  })

}
shinyApp(ui = ui, server = server)
```

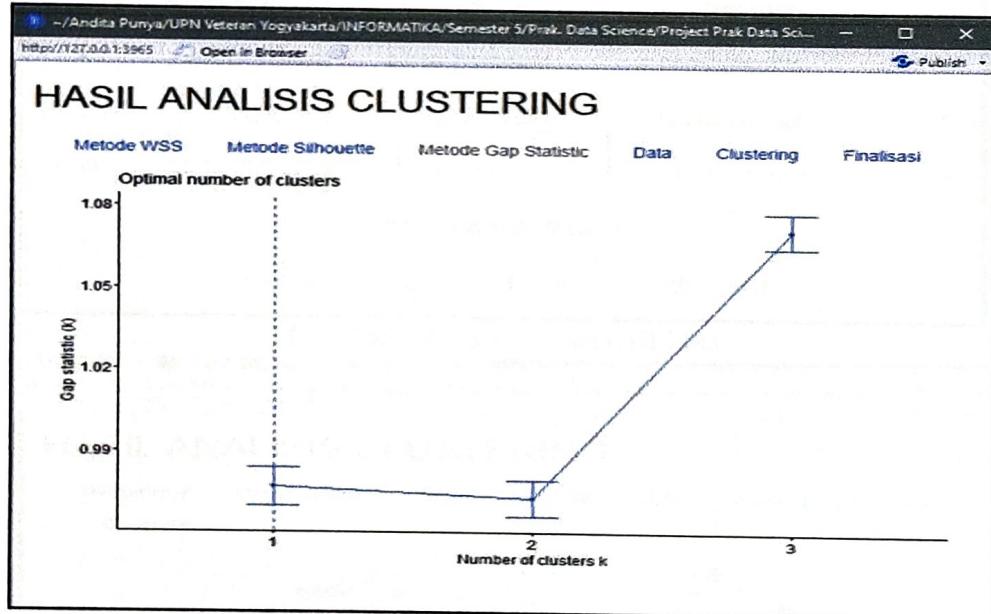
```

**Listing 3.9 Shiny**

### 3.2 Tampilan Gui



Gambar 3.2 Tampilan GUI Metode WSS



Gambar 3.2 (a) Tampilan GUI Metode Gap Statistic

~ /Andita Punya/UPN Veteran Yogyakarta/INFORMATIKA/Semester 5/Prak. Data Science/Project Prak Data Sci... http://127.0.0.1:3965 Open in Browser Publish

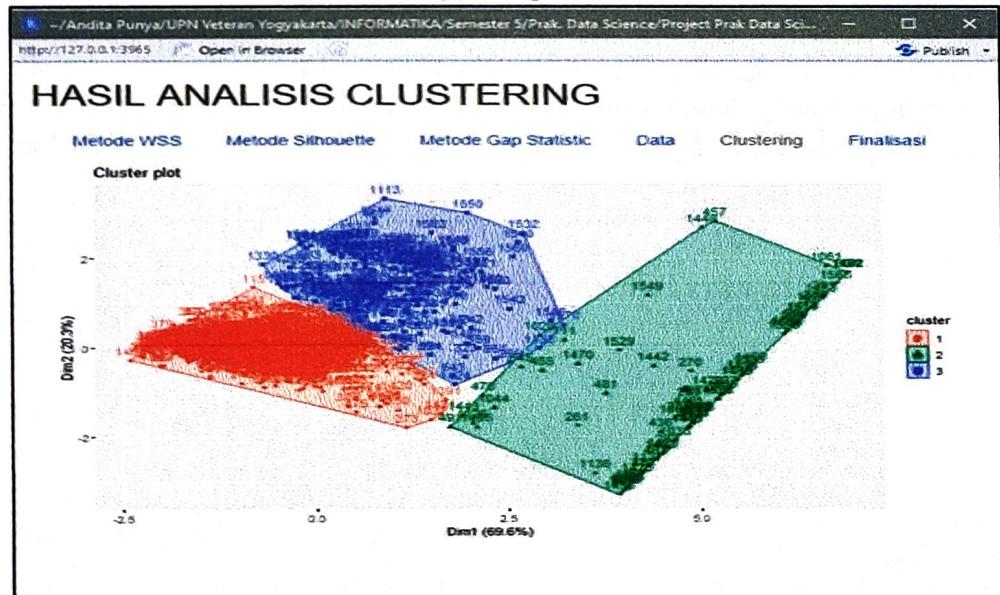
## HASIL ANALISIS CLUSTERING

|                              | Metode WSS          | Metode Silhouette   | Metode Gap Statistic  | Data | Clustering    | Finalisasi |
|------------------------------|---------------------|---------------------|-----------------------|------|---------------|------------|
| Search: <input type="text"/> |                     |                     |                       |      |               |            |
|                              | bidang_pemerintahan | bidang_kewilayahana | bidang_kemasyarakatan |      | final.cluster |            |
| 1                            | 1.10303970080905    | 1.17117789550052    | 0.633423345483136     |      | 1             |            |
| 2                            | 0.0191750213061367  | 0.0229261972376954  | 0.106939602408792     |      | 1             |            |
| 3                            | -0.170501297606874  | -0.0769217765242892 | -0.00507860260736954  |      | 1             |            |
| 4                            | -0.278887765557166  | -0.476313671572227  | 0.131743243412024     |      | 1             |            |
| 5                            | 0.723687062983034   | 0.671938026690595   | 0.222957807424954     |      | 1             |            |
| 6                            | 0.642397212020315   | 0.12277417099968    | 0.633423345483136     |      | 1             |            |
| 7                            | 0.615300595032742   | 1.17117789550052    | 0.268565089431418     |      | 1             |            |
| 8                            | 0.154658106244002   | -0.276617724046258  | -0.119096807623531    |      | 1             |            |
| 9                            | 0.859170147920898   | 0.57209005292861    | 0.131743243412024     |      | 1             |            |
| 10                           | 0.669493829007888   | 0.671938026690595   | 0.336976012441115     |      | 1             |            |

Showing 1 to 10 of 1,594 entries

Previous 1 2 3 4 5 ... 160 Next

Gambar 3.2 (b) Tampilan GUI Data



~ /Andita Purnya/UPN Veteran Yogyakarta/INFORMATIKA/Semester 5/Prak. Data Science/Project Prak Data Sci... - Open in Browser Publish

### HASIL ANALISIS CLUSTERING

| Metode WSS | Metode Silhouette   | Metode Gap Statistic | Data                   | Clustering | Finalisasi |
|------------|---------------------|----------------------|------------------------|------------|------------|
|            |                     |                      |                        | Search:    |            |
| Cluster    | bidang_pemerintahan | bidang_kewilayah     | bidang_kemasayarakatan |            |            |
| 1          | 199.02109375        | 70.65234375          | 181.53515625           |            |            |
| 2          | 124.577777777778    | 9.12222222222222     | 8.42222222222222       |            |            |
| 3          | 124.120535714286    | 66.1160714285714     | 163.473214285714       |            |            |

Showing 1 to 3 of 3 entries

Previous  Next

**Gambar 3.2 (d) Tampilan GUI Clustering**

### 3.3 Hasil Proyek

Link folder pembuatan proyek menggunakan bahasa pemrograman R dengan memanfaatkan aplikasi RStudio:

[https://drive.google.com/drive/folders/1P9SnTlojFL\\_BFD0WaJIJ1HdHz6k\\_Ncj7  
?usp=share\\_link](https://drive.google.com/drive/folders/1P9SnTlojFL_BFD0WaJIJ1HdHz6k_Ncj7?usp=share_link)

## 4. KESIMPULAN

Berdasarkan kesimpulan yang dapat dilihat dari data pada **Gambar 3.1.6(d)** yang diperoleh dari hasil analisis untuk menentukan clustering dari tingkat perkembangan desa/kelurahan di Provinsi Lampung didapatkan 3 cluster yaitu kluster 1, 2 dan 3, dimana masing-masing cluster memiliki nilai rata-rata yang berbeda. Dari data tersebut dapat kita simpulkan bahwa golongan cluster 1 memiliki nilai rata-rata paling tinggi pada tiap bidang. Cluster 2 memiliki nilai rata-rata paling rendah tiap bidangnya dibandingkan dengan nilai pada cluster lainnya. Sedangkan, pada cluster 3 tiap bidang memiliki nilai rata-rata sedang.