KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Vol. 1, No. 4, Oktober (2020), pp. 154-161

# Penerapan Metode Algoritma *K-means* Dalam Pengelompokan Angka Harapan Hidup Saat Lahir Menurut Provinsi

Riska Oktavia[1], Jaya Tata Hardinata[2], Irawan[3]
[1,2]*STIKOM Tunas Bangsa, Pematangsiantar – Indonesia*
[3]*AMIK Tunas Bangsa, Pematangsiantar – Indonesia*
*Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara*
[1]*oktaviariska755@gmail.com*, [2]*jayatatahardinata@gmail.com*,
[3]*irawaniwan56@gmail.com*

***Abstract***

*Life Expectancy (AHH) at birth is an estimate of the average additional age of a person from a mother's womb during the birth process which is expected to be able to live normally and healthy. Based on data obtained from the Government's official website address at https://bps.go.id/ which displays several amounts that vary from 2015 to 2018 according to the Province in Indonesia. For this reason, it is necessary to cluster each number of life expectancy at birth with the number from the lowest to the highest using the Data Mining method with the K-means Clustering Algorithm. In this research technique, the data will be classified based on the name of the Province which has the number of Life Expectancy at birth from 2015 to 2018. That is why the Data Mining method is used to facilitate the grouping of data on the number of Life Expectancy at birth according to the name of the Province in Indonesia. After grouping, the results will be obtained the number of Life Expectancy at birth, grouping starts from the lowest to the highest cluster. In the research that has been carried out, it is expected that the Government will provide solutions of the highest life expectancy at birth that has the highest number, so that in the following year the life expectancy rate will be reduced.*

***Keywords:*** *Grouping, Data Mining, Clustering, K-means, Life Expectancy (AHH)*

***Abstrak***

*Angka Harapan Hidup (AHH) saat lahir adalah perkiraan rata-rata tambahan umur seseorang dari kandungan seorang Ibu saat proses kelahiran yang diharapkan bisa dapat hidup secara normal dan sehat. Berdasarkan data yang diperoleh dari situs resmi Pemerintah yang beralamatkan di https://bps.go.id/ yang menampilkan beberapa jumlah yang bervariasi mulai dari tahun 2015 sampai 2018 menurut Provinsi di Indonesia. Untuk itu perlu di cluster setiap jumlah Angka Harapan Hidup saat lahir dengan jumlah dari mulai terendah sampai tertinggi menggunakan metode Data Mining dengan Algoritma K-means Clustering. Teknik penelitian ini, data akan dikelompokan berdasarkan nama Provinsi yang memiliki jumlah Angka Harapan Hidup saat lahir dari tahun 2015 sampai 2018. Itu sebabnya digunakan metode Data Mining untuk memudahkan pengelompokan data jumlah Angka Harapan Hidup saat lahir menurut nama Provinsi di Indonesia. Setelah dilakukannya pengelompokan maka hasil akan didapatkan jumlah Angka Harapan Hidup saat lahir, pengelompokan dimulai dari cluster terendah sampai tertinggi. Dalam penelitian yang telah dilakukan sangat diharapkan Pemerintah untuk memberikan solusi dari tingkat Angka Harapan Hidup saat lahir yang memiliki jumlah tertinggi, agar ditahun berikutnya tingkat Angka Harapan Hidup semakin diperkecil.*

***Kata Kunci:*** *Pengelompokan, Data Mining, Clustering, K-means, Angka Harapan Hidup (AHH)*

ISSN: 2720-992X

# 1. Pendahuluan

Angka Harapan Hidup didefenisikan sebagai data yang menggambarkan usia kematian pada suatu populasi. Data ini merupakan ringkasan pola usia kematian yang terjadi pada seluruh kelompok usia pada anak-anak. Pada 2016, Organisasi Kesehatan Dunia (*World Health Organization* atau *WHO*), mencatat angka harapan hidup Indonesia rata-rata adalah 69 tahun (71 tahun untuk wanita dan 67 tahun untuk pria). Sementara menurut data Badan Pusat Statistik RI, angka harapan hidup Indonesia pada 2018 lalu meningkat menjadi 71,2 tahun, dengan 69,3 tahun untuk pria dan 73,19 tahun untuk wanita. Angka Harapan Hidup pada dasarnya merupakan gambaran kondisi suatu wilayah secara garis besar. Jika angka kematian bayi tinggi, maka harapan hidup di wilayah tersebut akan rendah, begitu pula sebaliknya jika angka kematian bayi rendah, maka harapan hidup di wilayah tersebut akan tinggi. Adapun faktor yang mempengaruhi angka harapan hidup antara lain yaitu Harapan Subjektif, Demografi, Sosio-Ekonomi, Gaya Hidup, Pisikologis. Dalam upaya mengurangi jumlah Angka Harapan Hidup di Indonesia perlu dilakukan peng*cluster*an untuk mengetahui provinsi mana saja yang memiliki *cluster* terendah dari data angka harapan hidup saat lahir. Dari hasil *cluster* yang diperoleh akan lebih mudah bagi pemerintah daerah untuk menentukan solusi dalam mengurangi angka harapan hidup di beberapa provinsi yang ada di Indonesia yang termasuk dalam *cluster* tinggi.

Dalam permasalahan ini penulis memperoleh data dari website resmi pemerintah *https://.www.bps.go.id/* dari tahun 2015 sampai dengan tahun 2018, untuk itu data tersebut dikelompokan dengan algoritma *K-Mean.* Metode *K-means* adalah salah satu metode pengelompokkan bersifat partitional serta pembelajaran berciri *unsupervised* [1]. Secara prinsip, metode *K-means* bekerja dengan memasukkan *K* sebagai konstanta jumlah *cluster* yang diinginkan. Sedangkan, Means dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai *cluster*[2].Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma *K-means* memiliki kemampuan dalam pengelompokan data dalam jumlah yang cukup besar. Setelah dilakukannya pengelompokan menggunakan Metode *K-means* maka hasil penelitian diharapkan dapat membantu Pemerintah untuk memberikan solusi dari tingkat Angka Harapan Hidup saat lahir yang memiliki jumlah tertinggi, agar ditahun berikutnya tingkat Angka Harapan Hidup semakin diperkecil.

# 2. Metodologi Penelitian

## 2.1. Data Mining

*Data Mining* adalah proses ekstraksi atau penggalian data dan informasi yang besar yang belum diketahui sebelumnya, serta dapat digunakan untuk membuat suatu keputusan yang sangat penting". *Data Mining* juga dikenal dengan istilah *pattern recognition* merupakan suatu metode yang digunakan untuk pengolahan data guna menemukan pola yang tersembunyi dari data yang diolah. Dalam *data mining* juga terdapat metode-metode yang digunakan seperti klasifikasi, *clustering*, regresi, seleksi variable, dan market basket analisis [3].

Secara garis besar,data mining dapat dikelompokkan menjadi 2 kategori utama,yaitu: Deskriptive mining, yaitu proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk descriptive mining adalah *cluster*ing, asosiation, dan sequential mining. Predictive, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variable lain di masa depan.Salah satu teknik yang terdapat dalam predictive mining adalah klasifikasi [4].

### 2.2. Algoritma *K-means*

Algoritma *K-means* merupakan algoritma klasterisasi yang mengelompokkan data berdasarkan titik pusat klaster (*centroid*) terdekat dengan data. Tujuan dari *K-means* adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster. Ukuran kemiripan yang digunakan dalam klaster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap titik *centroid*. [5]. Inti dari Algoritma *K-means* adalah klasterisasi pengelompokan data berdasarkan titik pusat klaster (*centroid*) untuk memaksimalkan kemiripan data yang didapatkan berdasarkan jarak terpendek data terhadap titik *centroid* awal. Perbedaan antara *clustering* dengan klasifikasi yaitu *clustering* untuk pengelompokan sedangkan klasifikasi untuk penggolongan.

Langkah-langkah dalam algoritma *K-means* [6]:

a) Tentukan jumlah *cluster* (k) pada data set
b) Tentukan nilai pusat (*centroid*)
   Penentuan nilai *centroid* pada tahap awal dilakukan secara *random* atau dapat diambil dari nilai maksimum untuk *cluster* tinggi dan nilai minimum untuk *cluster* rendah.
c) Pada masing-masing *record*, hitung jarak terdekat dengan *centroid*. Jarak *centroid* yang digunakan adalah *Euclidean Distance*, dengan rumus seperti dibawah ini:

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^{p} |x_{2j} - x_{1j}|^2} \quad (1)$$

   Keterangan :
   D = *Euclidean Distance*
   x = Banyaknya objek
   Σp = Jumlah data *record*

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (2)$$

   Keterangan :
   De = *Eulidean Distance*
   i = Banyaknya objek [2]
   (x, y)= Koordinat objek
   (s, t) = Koordinat *centroid*
d) Kelompokkan objek berdasarkan jarak ke *centroid* terdekat untuk membuat *centroid* baru. *Centroid* baru diambil dari penjumlahan nilai berdasarkan jarak dari iterasi sebelumnya lalu dibagi dengan jumlah jarak dari tiap *cluster*.
e. Ulangi langkah ke-2, lakukan *iterasi* hingga *centroid* bernilai optimal.

## 3. Hasil dan Pembahasan

Data yang diperoleh dari penelitian ini yaitu dari situs resmi Pemerintah *https://www.bps.go.id/*. Data yang diambil adalah data Angka Harapan Hidup Saat Lahir Menurut Provinsi di Indonesia dari tahun 2015-2018. Di dalam penelitian ini data tersebut di kelompokkan menjadi 2 *cluster* yaitu angka harapan hidup saat lahir tertinggi dan terendah. Dalam melakukan *clustering*, data yang diperoleh akan dihitung terlebih dahulu. Berikut uraian perhitungan manual proses Algoritma *K-means clustering*. Data Angka Harapan Hidup saat Lahir Menurut Provinsi yang akan digunakan dalam proses *clustering* dapat dilihat pada tabel 1 berikut ini:

**Tabel 1.** Tabel Data Angka Harapan Hidup

| No. | Provinsi | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| 1 | Aceh | 69.50 | 69.51 | 69.52 | 69.64 |
| 2 | Sumatera Utara | 68.29 | 68.33 | 68.37 | 68.61 |
| 3 | Sumatera Barat | 68.66 | 68.73 | 68.78 | 69.01 |

| No. | Provinsi | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|
| 4 | Riau | 70.93 | 70.97 | 70.99 | 71.19 |
| 5 | Jambi | 70.56 | 70.71 | 70.76 | 70.89 |
| 6 | Sumatera Selatan | 69.14 | 69.16 | 69.18 | 69.41 |
| 7 | Bengkulu | 68.50 | 68.56 | 68.59 | 68.84 |
| 8 | Lampung | 69.90 | 69.94 | 69.95 | 70.18 |
| 9 | Kep. Bangka Belitung | 69.88 | 69.92 | 69.95 | 70.18 |
| 10 | Kep. Riau | 69.41 | 69.45 | 69.48 | 69.64 |
| 11 | Dki Jakarta | 72.43 | 72.49 | 72.55 | 72.67 |
| 12 | Jawa Barat | 72.41 | 72.44 | 72.47 | 72.66 |
| 13 | Jawa Tengah | 73.96 | 74.02 | 74.08 | 74.18 |
| 14 | Di Yogyakarta | 74.68 | 74.71 | 74.74 | 74.82 |
| 15 | Jawa Timur | 70.68 | 70.74 | 70.80 | 70.97 |
| 16 | Banten | 69.43 | 69.46 | 69.49 | 69.64 |
| 17 | Bali | 71.35 | 71.41 | 71.46 | 71.68 |
| 18 | Nusa Tenggara Barat | 65.38 | 65.48 | 65.55 | 65.87 |
| 19 | Nusa Tenggara Timur | 65.96 | 66.04 | 66.07 | 66.38 |
| 20 | Kalimantan Barat | 69.87 | 69.90 | 69.92 | 70.18 |
| 21 | Kalimantan Tengah | 69.54 | 69.57 | 69.59 | 69.64 |
| 22 | Kalimantan Selatan | 67.80 | 67.92 | 68.02 | 68.23 |
| 23 | Kalimantan Timur | 73.65 | 73.68 | 73.70 | 73.96 |
| 24 | Kalimantan Utara | 72.16 | 72.43 | 72.47 | 72.50 |
| 25 | Sulawesi Utara | 70.99 | 71.02 | 71.04 | 71.26 |
| 26 | Sulawesi Tengah | 67.26 | 67.31 | 67.32 | 67.78 |
| 27 | Sulawesi Selatan | 69.80 | 69.82 | 69.84 | 70.08 |
| 28 | Sulawesi Tenggara | 70.44 | 70.46 | 70.47 | 70.72 |
| 29 | Gorontalo | 67.12 | 67.13 | 67.14 | 67.45 |
| 30 | Sulawesi Barat | 64.22 | 64.31 | 64.34 | 64.58 |
| 31 | Maluku | 65.31 | 65.35 | 65.40 | 65.59 |
| 32 | Maluku Utara | 67.44 | 67.51 | 67.54 | 67.80 |
| 33 | Papua Barat | 65.19 | 65.30 | 65.32 | 65.55 |
| 34 | Papua | 65.09 | 65.12 | 65.14 | 65.36 |

a) Menentukan Nilai k Jumlah *Cluster*
Jumlah *cluster* sebanyak 2 *cluster*. *Cluster* yang dibentuk yaitu *cluster* tinggi (C1) dan *cluster* rendah (C2).

b) Menentukan Nilai *Centroid* (Pusat *Cluster*)
Penentuan pusat *cluster* awal ditentukan secara *random* yang diambil dari data yang ada dalam *range.* Adapun nilai untuk *cluster* tinggi (*cluster* 1) diambil dari nilai tertinggi yang terdapat pada tabel 1 dan nilai untuk *cluster* terendah (*cluster* 2) diambil dari nilai terendah yang terdapat pada tabel 1. Berikut daftar tabel *centroid* data dapat dilihat pada tabel 2.

**Tabel 2**. *Centroid* Data Awal

| | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|
| ***Cluster* 1** | 74.68 | 74.71 | 74.74 | 74.82 |
| ***Cluster* 2** | 64.22 | 64.31 | 64.34 | 64.58 |

c) Menghitung Jarak Setiap Data Terhadap *Centroid* (Pusat *Cluster*)
Setelah data nilai pusat *cluster* awal ditentukan, maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat *cluster*. Proses pencarian jarak terpendek pada iterasi 1 dapat dilihat pada perhitungan dan tabel dibawah ini :

$$D_{Aceh,c1} = \sqrt{\begin{matrix}(69.50-74.68)^2+(69.51-74.71)^2+\\(69.52-74.74)^2+(69.64-74.82)^2\end{matrix}} = 10.39$$

$$D_{Aceh,c2} = \sqrt{\begin{matrix}(69.50-64.22)^2+(69.51-64.31)^2+\\(69.52-64.34)^2+(69.64-64.58)^2\end{matrix}} = 10.36$$

Hasil dari keseluruhan perhitungan dapat dilihat pada tabel 3 sebagai berikut :

**Tabel 3.** Hasil Perhitungan Jarak Pusat *Cluster* Iterasi 1

| No | Provinsi | C1 | C2 | Jarak Terpendek | Kelompok |
|---|---|---|---|---|---|
| 1 | Aceh | 10,39 | 10,36 | 10,36 | 2 |
| 2 | Sumatera Utara | 12,68 | 8,08 | 8,08 | 2 |
| 3 | Sumatera Barat | 11,89 | 8,87 | 8,87 | 2 |
| 4 | Riau | 7,44 | 13,32 | 7,44 | 1 |
| 5 | Jambi | 8,02 | 12,74 | 8,02 | 1 |
| 6 | Sumatera Selatan | 11,03 | 9,72 | 9,72 | 2 |
| 7 | Bengkulu | 12,23 | 8,52 | 8,52 | 2 |
| 8 | Lampung | 9,49 | 11,26 | 9,49 | 1 |
| 9 | Kep. Bangka Belitung | 9,51 | 11,24 | 9,51 | 1 |
| 10 | Kep. Riau | 10,49 | 10,27 | 10,27 | 2 |
| … | … | … | … | … | … |
| 31 | Maluku | 18,65 | 2,10 | 2,10 | 2 |
| 32 | Maluku Utara | 14,33 | 6,42 | 6,42 | 2 |
| 33 | Papua Barat | 18,80 | 1,96 | 1,96 | 2 |
| 34 | Papua | 19,12 | 1,63 | 1,63 | 2 |

Dalam menentukan masing-masing data hasil *cluster* Angka Harapan Hidup saat Lahir berdasarkan jarak minimum data terhadap pusat *cluster*. Berdasarkan hasil dari perhitungan jarak pada *cluster* iterasi 1 diperoleh *cluster* tinggi sebanyak 17 provinsi dan *cluster* rendah sebanyak 17 provinsi.

d) Menghitung *centroid* baru menggunakan hasil pada masing-masing *cluster*.
Setelah didapatkan hasil jarak dari setiap objek pada iterasi ke-1 maka lanjut ke iterasi ke-2 pada perhitungan dan tabel dibawah ini :

$$D_{2015,c1} = \frac{70.93+70.56+69.90+69.88+72.43+72.41+73.96+74.68+70.68+71.35+69.87+69.54+73.65+72.16+70.99+69.80+70.44}{17} = 71.37$$

$$D_{2016,c1} = \frac{70.97+70.71+69.94+69.92+72.49+72.44+74.02+74.71+70.74+71.41+69.90+69.57+73.68+72.43+71.02+69.82+70.46}{17} = 71.43$$

$$D_{2017,c1} = \frac{70.99+70.76+69.94+69.92+72.49+72.44+74.02+74.71+70.74+71.41+69.90+69.57+73.68+72.43+71.02+69.82+70.46}{17} = 71.46$$

$$D_{2018,c1} = \frac{71.19+70.89+70.18+70.18+72.67+72.66+74.18+74.82+70.97+71.68+70.18+69.64+73.96+72.50+71.26+70.08+70.72}{17} = 71.63$$

$$D_{2015,c2} = \frac{69.50+68.29+68.66+69.14+68.50+69.41+69.43+65.38+65.96+67.80+67.26+67.12+64.22+65.31+67.44+65.19+65.06}{17} = 67.28$$

$$D_{2016,c2} = \frac{69.51+68.33+68.73+69.16+68.56+69.45+69.46+65.48+66.04+67.92+67.31+67.13+64.31+65.35+67.51+65.30+65.12}{17} = 67.33$$

$$D_{2017,c2} = \frac{69.52+68.37+68.78+69.18+68.59+69.48+69.49+65.55+66.07+68.02+67.32+67.14+64.34+65.40+67.54+65.32+65.14}{17} = 67.37$$

$$D_{2018,c2} = \frac{69.64+68.61+69.01+69.41+68.84+69.64+69.64+65.87+66.38+68.23+67.78+67.45+64.58+65.59+67.80+65.55+65.36}{17} = 67.61$$

**Tabel 4.** *Centroid* baru iterasi 1

| | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|
| ***Cluster* 1** | 71,37 | 71,43 | 71,46 | 71,63 |
| ***Cluster* 2** | 67,28 | 67,33 | 67,37 | 67,61 |

Selanjutnya dilakukan kembali langkah ke 4 dan 5. Jika nilai *centroid* hasil iterasi dengan nilai *centroid* sebelumnya bernilai sama serta posisi *cluster* tidak mengalami perubahan maka proses iterasi berhenti. Namun jika nilai *centroid* tidak sama serta posisi data masih berubah maka proses iterasi berlanjut pada

iterasi berikutnya. Hasil dari keseluruhan perhitungan dapat diihat pada tabel 5. sebagai berikut :

**Tabel 5.** Hasil Perhitungan Jarak Pusat *Cluster* Iterasi 2

| No. | Provinsi | C1 | C2 | Jarak Terpendek | Kelompok |
|---|---|---|---|---|---|
| 1 | Aceh | 3,86 | 4,29 | 3,86 | 1 |
| 2 | Sumatera Utara | 6,15 | 2,01 | 2,01 | 2 |
| 3 | Sumatera Barat | 5,36 | 2,80 | 2,80 | 2 |
| 4 | Riau | 0,91 | 7,25 | 0,91 | 1 |
| 5 | Jambi | 1,49 | 6,67 | 1,49 | 1 |
| 6 | Sumatera Selatan | 3,65 | 4,50 | 3,65 | 1 |
| 7 | Bengkulu | 5,70 | 2,45 | 2,45 | 2 |
| 8 | Lampung | 2,96 | 5,19 | 2,96 | 1 |
| 9 | Kep. Bangka Belitung | 2,98 | 5,17 | 2,98 | 1 |
| 10 | Kep. Riau | 3,96 | 4,20 | 3,96 | 1 |
| … | … | … | … | … | … |
| 31 | Maluku | 12,12 | 3,97 | 3,97 | 2 |
| 32 | Maluku Utara | 7,80 | 0,35 | 0,35 | 2 |
| 33 | Papua Barat | 12,27 | 4,12 | 4,12 | 2 |
| 34 | Papua | 12,59 | 4,44 | 4,44 | 2 |

Dari hasil perhitungan jarak pusat *cluster* dari iterasi 2 diperoleh 21 provinsi pada *cluster* tinggi dan 13 provinsi pada *cluster* rendah. Berdasarkan hasil yang diperoleh dari iterasi 2 tidak sama dengan hasil dari iterasi 1. Maka perlu dilakukan perulangan hingga diperoleh hasil yang sama dari iterasi sebelumnya. Selanjutnya tentukan pusat *cluster* baru menggunakan cara yang sama seperti langkah ke-4, dan berikut adalah hasil *centroid* baru:

**Tabel 6.** *Centroid* baru

| | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|
| ***Cluster* 1** | 70,99 | 71,04 | 71,07 | 71,24 |
| ***Cluster* 2** | 66,63 | 66,7 | 66,74 | 67 |

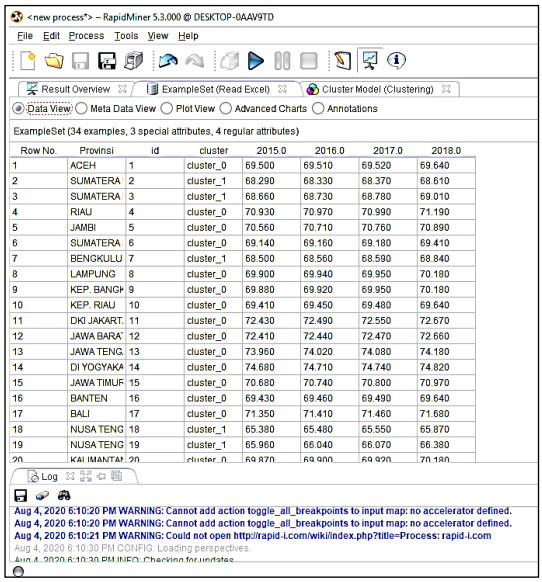
Selanjutnya dilakukan kembali langkah ke 3. Jika nilai *centroid* hasil iterasi dengan nilai *centroid* sebelumnya bernilai sama serta posisi *cluster* tidak mengalami perubahan maka proses iterasi berhenti.

**Tabel 7.** Hasil Perhitungan Jarak Pusat *Cluster* Iterasi 3

| No. | Provinsi | C1 | C2 | Jarak Terpendek | Kelompok |
|---|---|---|---|---|---|
| 1 | Aceh | 3,09 | 5,55 | 3,09 | 1 |
| 2 | Sumatera Utara | 5,37 | 3,27 | 3,27 | 2 |
| 3 | Sumatera Barat | 4,58 | 4,06 | 4,06 | 2 |
| 4 | Riau | 0,13 | 8,51 | 0,13 | 1 |
| 5 | Jambi | 0,72 | 7,93 | 0,72 | 1 |
| 6 | Sumatera Selatan | 3,73 | 4,91 | 3,73 | 1 |
| 7 | Bengkulu | 4,93 | 3,71 | 3,71 | 2 |
| 8 | Lampung | 2,19 | 6,45 | 2,19 | 1 |
| 9 | Kep. Bangka Belitung | 2,21 | 6,43 | 2,21 | 1 |
| 10 | Kep. Riau | 3,18 | 5,46 | 3,18 | 1 |
| … | … | … | … | … | … |
| 31 | Maluku | 11,35 | 2,71 | 2,71 | 2 |
| 32 | Maluku Utara | 7,03 | 1,61 | 1,61 | 2 |
| 33 | Papua Barat | 11,49 | 2,86 | 2,86 | 2 |
| 34 | Papua | 11,82 | 3,18 | 3,18 | 2 |

Perhitungan manual pada data diatas didapatkan hasil akhir yang dimana pada iterasi 2 dan iterasi 3 pengelompokan data yang dilakukan terhadap 2 *cluster* didapatkan hasil yang sama. Hasil dari iterasi kedua benilai C2 = 21 dan C3 = 13 pada posisi data tiap *cluster*. Sehingga posisi *cluster* pada data tersebut tidak mengalami perubahan lagi maka proses iterasi berhenti sampai iterasi 3.

Berdasarkan hasil yang didapat dari *rapidminer* dan *Microsoft excel* menjelaskan bahwa hasil dari perhitungan manual algoritma *k-means* dan *Microsoft excel* data memiliki nilai yang sama yaitu antara beberapa *cluster* yaitu *cluster* tinggi 21 dan rendah 13, serta memasukan perhitungan *Microsoft excel* ke dalam *rapidminer* memiliki nilai yang sama juga. Dapat kita lihat pada gambar 1 berikut.




**Gambar 1.** Tampilan Data Perhitungan *Rapidminer*

## 4. Kesimpulan

Berdasarkan pembahasan dan hasil pada pengelompokkan data Angka Harapan Hidup saat Lahir, maka dapat disimpulkan sebagai berikut :

a) Data yang diolah untuk memperoleh nilai dari Angka Harapan Hidup saat Lahir menerapkan metode *Clustering K-means*. Data tersebut diolah menggunakan *Microsoft Excel* untuk menentukan nilai *centroid* dalam 2 *cluster* yaitu *cluster* tertinggi dan terendah.

b) Hasil yang diperoleh dari metode *K-means clustering* ke dalam *Rapidminer* memiliki nilai yang sama yaitu menghasilkan 2 *cluster*. *Cluster* tertinggi menghasilkan 21 Provinsi diantaranya Aceh, Sumatera Utara, Riau, Jambi, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara. Sedangkan untuk *Cluster* terendah sebanyak 13 yaitu Sumatera Barat, Sumatera Selatan, Bengkulu, NTB, NTT, Kalimantan Selatan, Sulawesi Tengah, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Hasil yang didapat dari penelitian diharapkan dapat menjadi masukan Pemerintah Indonesia agar memberi perhatian lebih pada Provinsi yang memiliki Angka Harapan Hidup saat Lahir terendah berdasarkan *Cluster* yang telah dilakukan.

## Daftar Pustaka

[1] A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Bagian 1: Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 855, 2020, doi: 10.30865/mib.v4i3.2312.

[2] Y. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam," *Y. Darmi, A. Setiawan*, vol. 12, no. 2, pp. 148–157, 2016.

[3] E. G. Sihombing, "Klasifikasi Data Mining Pada Rumah Tangga Menurut Provinsi Dan Status Kepemilikan Rumah Kontrak / Sewa Menggunakan K-Means Clustering Method," Vol. 2, No. 2, Pp. 74–82, 2017.

[4] N. Puspitasari and Haviluddin, "Penerapan Metode K-Means Dalam Pengelompokkan Curah Hujan," *Semin. Nas. Ris. Ilmu Komput. (SNRIK )*, vol. 1, no. March 2017, 2016.

[5] B. R. C.T.I. *et al.*, "Implemetasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan," *Semin. Nas. Ris. Kuantitatif Terap. 2017*, no. April, pp. 58–60, 2017.

[6] M. G. Sadewo, A. P. Windarto, and D. Hartama, "Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, pp. 60–67, 2017.