

## CLUSTERING DAERAH MISKIN DI PROVINSI RIAU MENGUNAKAN METODE K-MEANS

Fatimah Isyarah<sup>1</sup>, Mhd, Arief Hasan<sup>2</sup>, Fana Wiza<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang  
Kuning

<sup>1,2,3</sup>Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015

e-mail: fatimahisyarah@gmail.com<sup>1</sup>, m.arif@unilak.ac.id<sup>2</sup>, fana@unilak.ac.id<sup>3</sup>

### Abstrak

*Analisis Klaster sangat cocok untuk mengelompokkan data-data berukuran besar. Dan penelitian kali ini bertujuan mengelompokkan daerah miskin di Provinsi Riau berdasarkan Umur Harapan Hidup, Jumlah Penduduk, Pengeluaran perKapita, Rata-rata Lama Sekolah, Angkatan Kerja, dan Harapan Lama Sekolah agar tepat sasaran, dan masyarakat mendapat program-program sekaligus bantuan dari pemerintah, dengan menggunakan K-means sebagai metode peng-klasteran. K-means menjabarkan algoritma untuk menentukan sebuah obyek kedalam klaster tertentu berdasarkan rata-rata terdekat dan mudah di implementasikan. Dari hasil penelitian, klaster pertama memiliki 5 anggota, klaster kedua memiliki 6 anggota, dan klaster 3 yang memiliki 1 anggota.*

**Kata Kunci:** Clustering, K-Means, Penduduk

### Abstract

*Cluster Analysis is perfect for grouping large data. And this research is trying to classify poor areas in Riau Province based on Life Expectancy, Population Amount, Per capita Expenditures, Old Average Schools, Labor Force, and Old Hope Schools to be on target, and the community gets government assistance programs, with using K-means as a Clustering method. K-means describing the algorithm to determine the object to a certain Cluster based on the nearest average and easy to implement. From the results of the study, the first Cluster has 5 members, the second Cluster has 6 members, and Cluster 3 has 1 member.*

**Keywords:** Clustering, K-Means, Population

### 1. PENDAHULUAN

Pada era yang serba modern ini, kita dapat menikmati semua fasilitas dan kemajuan teknologi yang membuat semuanya dapat diakses dengan mudah dan cepat. Teknologi sudah digunakan di hampir segala bidang untuk mempermudah pekerjaan kita. Namun sayangnya, tidak semua orang dapat menikmati kenyamanan yang tersedia saat ini. Kemiskinan masih saja menjadi masalah bagi negara berkembang seperti Indonesia. Badan Pusat Statistik (BPS) mencatat jumlah penduduk miskin Di provinsi Riau pada bulan Maret 2018 dengan 500,44 ribu. Jumlah penduduk miskin meningkat 4,05 ribu orang dibandingkan dengan september 2017, sedangkan maret 2017 mereka turun sebanyak 14,18 ribu orang. Dan jika di hitung kembali, penduduk miskin di provinsi Riau turun menjadi 7,39 persen.

Saat penyaluran bantuan, terkadang tak semua masyarakat miskin mendapatkan bantuan yang memang seharusnya mereka dapatkan. Alasannya karena hingga saat ini pun banyak terjadi kesalahan dalam pembagian bantuan. Pemerintah terkadang tak terlalu memperhatikan masyarakat yang mendapatkan bantuan adalah masyarakat yang benar-benar membutuhkan bantuan atau tidak. Di sekitar kita masih sering terjadi kesalahan pemberian bantuan berupa orang yang tak berhak mendapatkan bantuan malah

mendapatkannya, sedangkan yang membutuhkan malah tidak mendapatkan bantuan. Hal ini terjadi karena kurangnya ketelitian pemerintah dalam mendata masyarakat miskin.

Pada penelitian ini peneliti menggunakan aplikasi *Rapid Miner* sebagai pengujian untuk perhitungan *K-means Cluster*. Sedangkan untuk penelitian terdahulu, beberapa peneliti menggunakan berbagai pengujian yang berbeda, yakni [8] Penelitian ini memiliki beberapa asumsi dalam pengujian *Cluster K-Means*. Yaitu Uji Multikolenieritas dan Uji Diskriminan, adapun uji Diskriminan terdiri dari : Uji *Normal Multivariate*, Uji Kehomogenan Matriks Kovarian Kedua Kelompok, Uji Perbedaan Vektor Rata-rata, dan Uji *Variance*. Peneliti [12] menggunakan Davies Bouldin Index untuk memaksimalkan jarak *Inter-Cluster* supaya memiliki skema *Clustering* yang optimal. Selain itu mereka juga menggunakan *Min-max Normalization* yang merupakan salah satu dari metode normalisasi data yang mengubah nilai pada dataset dalam skala yang lebih kecil dan bernilai 0 hingga 1. Sedangkan untuk penelitian [15] peneliti menggunakan pendeteksian Multikolinearitas yang memiliki artian adanya hubungan yang linier di antara variabel penelitian. Lalu dilanjutkan dengan memilih ukuran kemiripan, yaitu untuk mengetahui seberapa mirip objek-objek yang diteliti butuhkan untuk ukuran jarak. Lalu selanjutnya standarisasi data, serta penentuan jumlah klaster.

Lalu pada penelitian [1] mereka melakukan pengujian asumsi mengenai multikolinearitas menggunakan korelasi *Pearson*. Mengatasi multikolinearitas dalam data dapat mengekstraksi faktor, yang mengakibatkan faktor tak dapat berkorelasi dengan sesamanya. Maka dari itu, sebelum melakukan analisis faktor maka harus dilakukan pengujian data menggunakan uji *Bartlett* dan KMO. Uji *Bartlett* menggunakan statistik *Chi-square*, dengan tujuan untuk melihat apakah variabel berkorelasi dengan variabel lainnya. Sedangkan KMO itu sendiri digunakan untuk meneliti ketepatan analisis faktor. Pada penelitian [2] sebelum memulai proses perhitungan, sebelum itu mereka melakukan salah satu proses yang dilakukan adalah membersihkan data sebelum di proses, atau disebut juga dengan Preproses data. Teknik dalam preproses data itu sendiri diantaranya data *Cleaning*, yang berarti mengisi *Missing Value* dan membuang *Outlier*. Lalu data Integrasi yang berarti mengumpulkan data dari berbagai sumber, data transformasi, data reduksi, dan data diskretisasi.

Berdasarkan penelitian yang telah dilakukan, dengan menerapkan Metode *K-means* dapat mengelompokkan daerah miskin menjadi tiga klaster, yaitu Mampu, Menengah, dan Kurang mampu. Berdasarkan masalah di atas, maka dilakukan penelitian yang membahas seputar "*Clustering* Pengelompokan Daerah Miskin Di Provinsi Riau Menggunakan Metode *K-means*".

## 2. METODE PENELITIAN

### 2.1 Praproses Data

Praproses data adalah mengumpulkan data mentah dari Badan Pusat Statistik Provinsi Riau berupa data yang tersimpan dalam format PDF, yang nantinya akan dijadikan sebuah data baru yang terdiri dari data Jumlah Penduduk, Angkatan Kerja, Umur Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, serta Pengeluaran Perkapita se-Provinsi Riau yang akan dibuat dalam format *(xls)* [8].

Langkah ini menjelaskan bagaimana tahapan mengubah data mentah menjadi data yang bisa digunakan untuk proses klasterisasi. Tahapan dalam praproses data yaitu mengintegrasikan data, membersihkan, dan mentransformasi.

#### a. Pembersihan Data

1. Menghilangkan Duplikat Data
2. Menghapus Data yang tidak Diperlukan

Tahapan ini melakukan proses penghapusan data yang tidak digunakan, untuk memperkecil jumlah data sesuai dengan kebutuhan, akan tetapi tetap mempertahankan

---

nilai dari data asli. Data yang digunakan kali ini adalah data dari tahun 2018, yang mencakup Umur Harapan Hidup, Jumlah Penduduk, Pengeluaran perKapita, Rata-rata Lama Sekolah, Angkatan Kerja, dan Harapan Lama Sekolah. Selain dari data yang tadi disebutkan, maka semuanya akan dihapus.

| Kabupaten/Kota    | Penduduk Kabupaten/Kota (Jiwa) |         |         |         |         |         |         |         |         |
|-------------------|--------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|
|                   | 2010                           | 2011    | 2012    | 2013    | 2014    | 2015    | 2016    | 2017    | 2018    |
| Kuantan Singingi  | 293314                         | 294468  | 302631  | 306718  | 310619  | 314276  | 317935  | 321216  | 324413  |
| Indragiri Hulu    | 365421                         | 372074  | 383814  | 392354  | 400901  | 409431  | 417733  | 425897  | 433934  |
| Indragiri Hilir   | 670499                         | 675898  | 676419  | 685530  | 694614  | 703734  | 713034  | 722234  | 731396  |
| Pelalawan         | 304597                         | 329539  | 339869  | 358210  | 377221  | 396990  | 417498  | 438788  | 460780  |
| Siak              | 379089                         | 391760  | 404093  | 416298  | 428499  | 440841  | 453052  | 465414  | 477670  |
| Kampar            | 692179                         | 711236  | 733506  | 753376  | 773171  | 793005  | 812702  | 832387  | 851837  |
| Rokan Hulu        | 478496                         | 507079  | 523024  | 545483  | 568576  | 592278  | 616466  | 641208  | 666410  |
| Bengkalis         | 500635                         | 503604  | 519389  | 527918  | 536138  | 543987  | 551683  | 559081  | 566228  |
| Rokan Hilir       | 556575                         | 574419  | 592403  | 609779  | 627233  | 644680  | 662242  | 679663  | 697218  |
| Kepulauan Meranti | 175989                         | 177004  | 177587  | 178839  | 179894  | 181095  | 182152  | 183297  | 184372  |
| Pekanbaru         | 903038                         | 929247  | 958352  | 984674  | 1011467 | 1038118 | 1064566 | 1091088 | 1117359 |
| Dumai             | 255096                         | 259913  | 268022  | 274089  | 280109  | 285967  | 291908  | 297638  | 303292  |
| RIAU              | 5574928                        | 5726241 | 5879109 | 6033268 | 6188442 | 6344402 | 6500971 | 6657911 | 6814909 |

**Gambar 1.** Pembersihan Data

#### b. Integrasi Data

Penggabungan data dari beberapa sumber, yang berguna untuk untuk mengolah data menggunakan aplikasi *RapidMiner*, yaitu menggabungkan data Kabupaten/kota, Umur Harapan Hidup, Jumlah Penduduk, Pengeluaran perKapita, Rata-rata Lama Sekolah, Angkatan Kerja, serta Harapan Lama Sekolah menjadi sebuah data baru.

**Tabel 1.** Integrasi Data

| Nama Kabupaten   | Umur Harapan Hidup | Harapan Lama Sekolah | Rata-rata Lama Sekolah | Jumlah Penduduk | Angkatan Kerja | Pengeluaran Perkapita |
|------------------|--------------------|----------------------|------------------------|-----------------|----------------|-----------------------|
| Kuantan Singingi | 68,17              | 13,27                | 8,31                   | 324,41          | 147,906        | 10,476                |
| Indragiri Hulu   | 69,77              | 12,32                | 8,16                   | 433,93          | 208,558        | 10,481                |
| Indragiri Hilir  | 67,32              | 11,89                | 7,19                   | 731,40          | 330,363        | 10,254                |
| Pelalawan        | 70,74              | 12,16                | 8,44                   | 460,78          | 219,362        | 11,894                |
| Siak             | 70,79              | 12,73                | 9,64                   | 477,67          | 214,41         | 12,119                |
| Kampar           | 70,35              | 13,21                | 9,10                   | 851,54          | 372,775        | 11,128                |
| Rokan Hulu       | 69,55              | 12,82                | 8,37                   | 666,41          | 293,214        | 9,608                 |
| Bengkalis        | 70,85              | 12,83                | 9,21                   | 566,23          | 259,127        | 11,64                 |
| Rokan Hilir      | 69,87              | 12,63                | 8,15                   | 697,22          | 295,682        | 9,316                 |
| Kep. Meranti     | 67,21              | 12,78                | 7,48                   | 184,37          | 91,058         | 7,978                 |
| Pekanbaru        | 71,94              | 15,34                | 11,22                  | 1117,36         | 539,831        | 14,778                |
| Kota Dumai       | 70,55              | 12,98                | 9,84                   | 303,29          | 136,112        | 12,063                |

#### Data Hasil Praproses

Data hasil praproses, terdiri dari 12 buah data sampel, serta terdiri dari enam buah atribut yang akan diklusterisasikan, yaitu Umur Harapan Hidup, Jumlah Penduduk,

Pengeluaran perKapita, Rata-rata Lama Sekolah, Angkatan Kerja, dan Harapan Lama Sekolah

**Tabel 2.** Data Hasil Praproses

| Nama Kabupaten   | Umur Harapan Hidup | Harapan Lama Sekolah | Rata-rata Lama Sekolah | Jumlah Penduduk | Angkatan Kerja | Pengeluaran Perkapita |
|------------------|--------------------|----------------------|------------------------|-----------------|----------------|-----------------------|
| Kuantan Singingi | 68,17              | 13,27                | 8,31                   | 324,41          | 147,906        | 10,476                |
| Indragiri Hulu   | 69,77              | 12,32                | 8,16                   | 433,93          | 208,558        | 10,481                |
| Indragiri Hilir  | 67,32              | 11,89                | 7,19                   | 731,40          | 330,363        | 10,254                |
| Pelalawan        | 70,74              | 12,16                | 8,44                   | 460,78          | 219,362        | 11,894                |
| Siak             | 70,79              | 12,73                | 9,64                   | 477,67          | 214,41         | 12,119                |
| Kampar           | 70,35              | 13,21                | 9,10                   | 851,54          | 372,775        | 11,128                |
| Rokan Hulu       | 69,55              | 12,82                | 8,37                   | 666,41          | 293,214        | 9,608                 |
| Bengkalis        | 70,85              | 12,83                | 9,21                   | 566,23          | 259,127        | 11,64                 |
| Rokan Hilir      | 69,87              | 12,63                | 8,15                   | 697,22          | 295,682        | 9,316                 |
| Kep. Meranti     | 67,21              | 12,78                | 7,48                   | 184,37          | 91,058         | 7,978                 |
| Pekanbaru        | 71,94              | 15,34                | 11,22                  | 1117,36         | 539,831        | 14,778                |
| Kota Dumai       | 70,55              | 12,98                | 9,84                   | 303,29          | 136,112        | 12,063                |

#### c. Analisa Pengelompokan Data

Pada penelitian ini akan di lakukan pengklasteran sebanyak tiga buah klaster, yakni Menengah, Mampu, dan Tidak Mampu. Jadi pada penelitian ini akan dibentuk sebanyak tiga kelompok atan nilai  $k=3$ . Selain itu akan digunakan enam buah atribut, yakni Umur Harapan Hidup, Jumlah Penduduk, Pengeluaran perKapita, Rata-rata Lama Sekolah, Angkatan Kerja, serta Harapan Lama Sekolah.

#### 2.2 K-means Clustering

*K-means Clustering* adalah analisis kelompok non hirarki yang paling banyak di gunakan. Berikut langkah dalam metode *K-means*:

- Menentukan titik pusat secara acak.
- Hitunglah jarak setiap dengan setiap titik pusat menggunakan rumus *Euclidian Distance*.
- Menentukan keseluruhan objek ke klaster terdekat.
- Kembali menghitung pusat klaster berdasarkan anggota klaster yang baru.
- Menugaskan setiap objek agar menggunakan pusat klaster yang baru.

#### 2.3 Euclidian Distance

Penulis menggunakan rumus *Euclidean Distance* sebagai cara menghitung setiap jarak data yang ada pada pusat klaster.

$$d(x,y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; I = 1,2,3,....n ....$$

Di mana :

$x_i$  : Obyek x ke- $i$

$y_i$  : Daya y ke- $i$

$n$  : banyaknya obyek

**Gambar 2.** Rumus *Euclidean Distance*

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Perhitungan Manual

Tahapan ini melakukan proses utama yakni mengelompokkan data tingkat menggunakan metode algoritma *K-means*. Atribut per Provinsi Riau digunakan dalam proses analisa klustering algoritma *K-means* dan menghitung secara manual menggunakan aturan *K-means* yang sudah di tetapkan. Percobaan dilakukan menggunakan ukuran-ukuran di bawah ini :

- Jumlah klaster: 3
- Jumlah data : 12
- Jumlah variabel : 6

#### Iterasi Pertama

##### 1. Menentukan Pusat *Cluster* Awal

Tentukan centroid awal secara acak dari obyek yang tersedia sejumlah klaster *k*. Nilai untuk C1 di ambil dari baris PCDK12, C2 di ambil dari baris PCDK6, C3 di ambil dari baris data PCDK11. Di bawah ini merupakan nilai *Centroid* awal penelitian, C adalah *Cluster*:

C1 = (70,35 ; 13,21 ; 9,10 ; 851,84 ; 372,775 ; 11,128)  
C2 = (70,55 ; 12,98 ; 9,84 ; 303,29 ; 136,112 ; 12,063)  
C3 = (71,94 ; 15,34 ; 11,22 ; 1117,36 ; 539,831 ; 14,778)

##### 2. Menghitung Jarak dengan Pusat *Cluster*

Proses menghitung jarak dengan Euclidean Distance untuk Iterasi satu dan Centroid satu :

$$\begin{aligned} K01 &= \sqrt{(68,17 - 70,35)^2 + (13,27 - 13,21)^2 + (8,31 - 9,1)^2 + (324,41 - 851,84)^2 + (147,906 - 372,775)^2 + (10,476 - 11,128)^2} = 573,370 \\ K02 &= \sqrt{(69,97 - 70,35)^2 + (12,32 - 13,21)^2 + (8,16 - 9,1)^2 + (433,93 - 851,84)^2 + (208,558 - 372,775)^2 + (10,481 - 11,128)^2} = 449,019 \\ K03 &= \sqrt{(67,32 - 70,35)^2 + (11,89 - 13,21)^2 + (7,19 - 9,1)^2 + (731,40 - 851,84)^2 + (330,363 - 372,775)^2 + (10,254 - 11,128)^2} = 127,749 \\ K04 &= \sqrt{(70,74 - 70,35)^2 + (12,16 - 13,21)^2 + (8,44 - 9,1)^2 + (460,78 - 851,84)^2 + (219,363 - 372,775)^2 + (11,894 - 11,128)^2} = 420,078 \\ K05 &= \sqrt{(70,79 - 70,35)^2 + (12,73 - 13,21)^2 + (9,64 - 9,1)^2 + (477,67 - 851,84)^2 + (214,41 - 372,775)^2 + (12,119 - 11,128)^2} = 406,305 \\ &\dots\dots \\ K12 &= \sqrt{(70,55 - 70,35)^2 + (12,98 - 13,21)^2 + (9,84 - 9,1)^2 + (303,29 - 851,84)^2 + (136,112 - 372,775)^2 + (12,063 - 11,128)^2} = 597,426 \end{aligned}$$

Setelah menghitung jarak dengan *Euclidean Distance* untuk Iterasi satu dan *Centroid* satu dilakukan, maka tahapan selanjutnya melakukan perhitungan dengan *Centroid* dua, dan tiga dengan rumus yang sama dengan perhitungan diatas.

##### 3. Pengelompokan Data

Alokasikan masing-masing data ke Centroid terdekat. Mengalokasikan data ke masing-masing klaster di dasarkan perbandingan jarak antar data dan Centroid di tiap kluser yang ada, alokasikan data dengan tegas dalam klaster yang memiliki jarak terdekat ke Centroid dengan data tersebut.

**Tabel 3.** Hasil Perhitungan Iterasi Satu

| Kode  | C1      | C2      | C3      | C1 | C2  | C3 |
|-------|---------|---------|---------|----|-----|----|
| PCDK1 | 573,370 | 24,408  | 884,545 |    | □ □ |    |
| PCDK2 | 449,019 | 149,403 | 759,512 |    | □ □ |    |
| PCDK3 | 127,749 | 470,142 | 439,217 | □  |     |    |
| PCDK4 | 420,078 | 178,147 | 730,633 |    | □ □ |    |
| PCDK5 | 406,305 | 191,152 | 717,718 |    | □ □ |    |

| Kode   | C1      | C2      | C3       |
|--------|---------|---------|----------|
| PCDK6  | 0       | 597,426 | 313,740  |
| PCDK7  | 201,786 | 395,659 | 514,026  |
| PCDK8  | 307,391 | 290,294 | 618,514  |
| PCDK9  | 172,787 | 425,034 | 485,980  |
| PCDK10 | 724,502 | 127,299 | 1035,353 |
| PCDK11 | 313,741 | 908,689 | 0        |
| PCDK12 | 597,426 | 0       | 908,689  |

| C1 | C2  | C3    |
|----|-----|-------|
| □  |     |       |
| □  |     |       |
|    | □ □ |       |
| □  |     |       |
|    | □ □ |       |
|    |     | □ □ □ |
|    | □ □ |       |

### Iterasi Kedua

#### 1. Penentuan Pusat Cluster Baru

Menghitung nilai rata-rata dari rata-rata pada Centroid yang sama adalah cara menentukan Centroid baru.

$$\text{Pusat Cluster baru} = \frac{\sum_{i=1}^n x_i}{n} ; i = 1, 2, 3, \dots, n \dots$$

Dimana :

$x_i$  : Obyek ke- $i$  (anggota Cluster)

$n$  : Banyaknya obyek/jumlah obyek yang menjadi anggota Cluster

Gambar 3. Menentukan Centroid Baru

$$C1 = \left( \frac{67,32 + 70,35 + 69,55 + 69,87}{4} ; \frac{11,89 + 13,21 + 12,82 + 12,63}{4} ; \frac{7,19 + 9,1 + 8,37 + 8,15}{4} ; \frac{731,4 + 851,84 + 666,41 + 697,22}{4} ; \frac{330,363 + 372,775 + 293,214 + 295,682}{4} ; \frac{10,254 + 11,128 + 9,608 + 9,316}{4} \right) = (69,27; 12,63; 8,20; 736,717; 323,008; 10,076)$$

$$C2 = \left( \frac{68,17 + 69,97 + 70,74 + 70,79 + 70,85 + 67,21 + 70,55}{7} ; \frac{13,27 + 12,32 + 12,16 + 12,73 + 12,83 + 12,78 + 12,98}{7} ; \frac{8,31 + 8,16 + 8,44 + 9,64 + 9,21 + 7,48 + 9,84}{7} ; \frac{324,41 + 433,93 + 460,78 + 477,67 + 566,23 + 184,37 + 303,29}{7} ; \frac{147,906 + 208,558 + 219,362 + 214,41 + 259,127 + 91,058 + 136,112}{7} ; \frac{10,476 + 10,481 + 11,894 + 12,119 + 11,64 + 7,978 + 12,063}{7} \right) = (69,75; 12,72; 8,72; 392,954; 182,361; 10,950)$$

$$C3 = \left( \frac{71,94 + 15,34 + 11,22 + 1117,36 + 539,831 + 14,778}{1} \right) = (71,94; 15,34; 11,22; 1117,36; 539,831 + 14,778)$$

#### 2. Menghitung Jarak dengan Pusat Cluster yang Baru

Begitu mengetahui perhitungan dari pusat Cluster baru, lakukan perhitungan jarak dari masing-masing data pada pusat Cluster menggunakan *Euclidean Distance*. Di bawah ini adalah menghitung jarak dengan *Euclidean Distance* untuk Iterasi dua dan Centroid satu (pusat Cluster baru) :

$$\begin{aligned}
 K01 &= \sqrt{(68,17 - 69,27)^2 + (13,27 - 12,63)^2 + (8,31 - 8,20)^2 + (324,41 - 736,717)^2 + (147,906 - 323,008)^2 + (10,476 - 10,076)^2} = 477,951 \\
 K02 &= \sqrt{(69,97 - 69,27)^2 + (12,32 - 12,63)^2 + (8,16 - 8,20)^2 + (433,93 - 736,717)^2 + (208,558 - 323,008)^2 + (10,481 - 10,076)^2} = 323,697 \\
 K03 &= \sqrt{(67,32 - 69,27)^2 + (11,89 - 12,63)^2 + (7,19 - 8,20)^2 + (731,40 - 736,717)^2 + (330,363 - 323,008)^2 + (10,254 - 10,076)^2} = 9,369 \\
 K04 &= \sqrt{(70,74 - 69,27)^2 + (12,16 - 12,63)^2 + (8,44 - 8,20)^2 + (460,78 - 736,717)^2 + (219,363 - 323,008)^2 + (11,894 - 10,076)^2} = 294,770 \\
 K05 &= \sqrt{(70,79 - 69,27)^2 + (12,73 - 12,63)^2 + (9,64 - 8,20)^2 + (477,67 - 736,717)^2 + (214,41 - 323,008)^2 + (12,119 - 10,076)^2} = 280,905 \\
 &\dots\dots \\
 K12 &= \sqrt{(70,55 - 69,27)^2 + (12,98 - 12,63)^2 + (9,84 - 8,20)^2 + (303,29 - 736,717)^2 + (136,112 - 323,008)^2 + (12,063 - 10,076)^2} = 472,014
 \end{aligned}$$

Setelah menghitung jarak dengan *Euclidean Distance* untuk Iterasi dua dan *Centroid* satu (pusat *Cluster* baru) dilakukan, maka tahapan selanjutnya melakukan perhitungan dengan *Centroid* dua dan tiga dengan rumus yang sama dengan perhitungan diatas.

### 3. Pengelompokan Data

Setelah diketahui jarak dari masing-masing data terhadap pusat *Cluster* menggunakan *Euclidean Distance*, maka tahap selanjutnya adalah melakukan pengelompokan data tersebut, dimana *Cluster* dengan jarak terdekat diberi nilai (I,II,III), sedangkan untuk jarak yang jauh dikosongkan. Berikut hasil pengelompokan data untuk Iterasi dua :

**Tabel 4.** Hasil Pengelompokan Data Iterasi Dua

| Kode   | C1      | C2      | C3       |  | C1 | C2  | C3    |
|--------|---------|---------|----------|--|----|-----|-------|
| PCDK1  | 447,951 | 76,738  | 884,545  |  |    | □ □ |       |
| PCDK2  | 323,697 | 48,641  | 759,512  |  |    | □ □ |       |
| Kode   | C1      | C2      | C3       |  | C1 | C2  | C3    |
| PCDK3  | 9,369   | 369,403 | 439,217  |  | □  |     |       |
| PCDK4  | 294,770 | 77,276  | 730,633  |  |    | □ □ |       |
| PCDK5  | 280,905 | 90,593  | 717,718  |  |    | □ □ |       |
| PCDK6  | 125,432 | 496,823 | 313,741  |  | □  |     |       |
| PCDK7  | 76,362  | 295,073 | 514,026  |  | □  |     |       |
| PCDK8  | 182,079 | 189,523 | 618,514  |  | □  |     |       |
| PCDK9  | 48,038  | 324,687 | 485,980  |  | □  |     |       |
| PCDK10 | 599,080 | 227,729 | 1035,353 |  |    | □ □ |       |
| PCDK11 | 438,116 | 807,824 | 0        |  |    |     | □ □ □ |
| PCDK12 | 472,014 | 100,905 | 908,689  |  | □  |     |       |

### 4. Perbandingan Hasil Pengelompokan Iterasi

Sedangkan berikut tabel perbandingan hasil perhitungan Iterasi ke Satu dan ke Dua :

**Tabel 5.** Hasil Perhitungan Iterasi Satu dan Dua

| Iterasi ke-Satu |     |    | Kode  | Iterasi ke-Dua |     |    |
|-----------------|-----|----|-------|----------------|-----|----|
| C1              | C2  | C3 |       | C1             | C2  | C3 |
|                 | □ □ |    | PCDK1 |                | □ □ |    |



| C1 | C2  | C3    |        | C1 | C2  | C3    |
|----|-----|-------|--------|----|-----|-------|
|    | □ □ |       | PCDK2  |    | □ □ |       |
| □  |     |       | PCDK3  | □  |     |       |
|    | □ □ |       | PCDK4  |    | □ □ |       |
|    | □ □ |       | PCDK5  |    | □ □ |       |
| □  |     |       | PCDK6  | □  |     |       |
| □  |     |       | PCDK7  | □  |     |       |
|    | □ □ |       | PCDK8  | □  |     |       |
| □  |     |       | PCDK9  | □  |     |       |
|    | □ □ |       | PCDK10 |    | □ □ |       |
|    |     | □ □ □ | PCDK11 |    |     | □ □ □ |
|    | □ □ |       | PCDK12 |    | □ □ |       |

Karena di Iterasi dua posisi kluster masih terjadi perubahan atau posisi tidak sama dengan posisi kluster pada Iterasi satu maka proses berlanjut ke Iterasi selanjutnya.

### Iterasi ke-Tiga

Proses pada Iterasi ke-tiga dan seterusnya melalui tahapan yang sama dengan Iterasi ke-dua, mulai dari penentuan pusat *Cluster* baru, serta menghitung jarak dengan pusat *Cluster* yang baru. Yang berbeda adalah hasil dari pengelompokkan data.

#### 1. Pengelompokkan Data

**Tabel 6.** Hasil Pengelompokkan Data Iterasi Tiga

| Kode   | C1      | C2      | C3       |
|--------|---------|---------|----------|
| PCDK1  | 411,576 | 45,222  | 884,545  |
| PCDK2  | 287,284 | 80,003  | 759,512  |
| Kode   | C1      | C2      | C3       |
| PCDK3  | 35,224  | 400,987 | 439,217  |
| PCDK4  | 258,356 | 108,784 | 730,633  |
| PCDK5  | 244,520 | 122,142 | 717,718  |
| PCDK6  | 161,802 | 528,402 | 313,740  |
| PCDK7  | 40,017  | 326,644 | 514,026  |
| PCDK8  | 145,663 | 211,111 | 618,514  |
| PCDK9  | 15,561  | 356,220 | 485,980  |
| PCDK10 | 562,701 | 196,144 | 1035,353 |
| PCDK11 | 474,093 | 839,385 | 0        |
| PCDK12 | 435,646 | 69,412  | 908,689  |

| C1 | C2  | C3    |
|----|-----|-------|
|    | □ □ |       |
|    | □ □ |       |
| C1 | C2  | C3    |
| □  |     |       |
|    | □ □ |       |
|    | □ □ |       |
| □  |     |       |
| □  |     |       |
| □  |     |       |
|    | □ □ |       |
|    |     | □ □ □ |
|    | □ □ |       |



## 2. Perbandingan Hasil Pengelompokan Iterasi

Sedangkan berikut tabel perbandingan hasil pengelompokan data Iterasi ke-Dua dan Iterasi ke-Tiga :

**Tabel 7.** Hasil Pengelompokan Data Iterasi Dua dan Iterasi Tiga

| Iterasi ke-Dua |     |       | Kode   | Iterasi ke-Tiga |     |       |
|----------------|-----|-------|--------|-----------------|-----|-------|
| C1             | C2  | C3    |        | C1              | C2  | C3    |
|                | □ □ |       | PCDK1  |                 | □ □ |       |
|                | □ □ |       | PCDK2  |                 | □ □ |       |
| □              |     |       | PCDK3  | □               |     |       |
|                | □ □ |       | PCDK4  |                 | □ □ |       |
|                | □ □ |       | PCDK5  |                 | □ □ |       |
| □              |     |       | PCDK6  | □               |     |       |
| □              |     |       | PCDK7  | □               |     |       |
| □              |     |       | PCDK8  | □               |     |       |
| □              |     |       | PCDK9  | □               |     |       |
|                | □ □ |       | PCDK10 |                 | □ □ |       |
|                |     | □ □ □ | PCDK11 |                 |     | □ □ □ |
|                | □ □ |       | PCDK12 |                 | □ □ |       |

Pada tahap ini, posisi klaster Iterasi tiga tidak berubah dan sama seperti posisi pada Iterasi dua, maka proses Iterasi tidak perlu di lanjutkan. Jadi kesimpulannya proses Iterasi akan di hentikan di Iterasi tiga dengan nilai :

- Anggota Cluster 1 : {PCDK3, PCDK6, PCDK7, PCDK8, PCDK9}= 5 data
- Anggota Cluster 2 : { PCDK1, PCDK2, PCDK4, PCDK5, PCDK10, PCDK12} = 6 data
- Anggota Cluster 3 (C3) : { PCDK11} = 1 data

Berdasarkan proses pengelompokan di atas, maka di hasilkan pengelompokan pengelompokan daerah miskin di Provinsi Riau berdasarkan setiap atribut yang ada. Nilai atau hasil penelitian di uraikan menjadi tiga bagian, yakni :

### 1. Klaster Satu

Berdasarkan tabel di bawah, dapat di lihat jika atribut klaster satu yang mempunyai hasil rata-rata Umur Harapan Hidup 69,58, Angkatan Kerja 310,232, dan Jumlah Penduduknya 702,62 memiliki nilai yang lebih dominan dibandingkan dengan klaster dua yang memiliki Umur Harapan Hidup 69,57, Angkatan Kerja 169,567, dengan Jumlah Penduduk 364,075. Dan jika semua nilai atribut dihitung, maka nilai rata-rata klaster satu lebih dominan di bandingkan klaster dua. Adapun anggota dari klaster satu adalah sebanyak lima anggota klaster, yaitu Rokan Hilir, Bengkalis, Rokan Hulu, Kampar, dan Indragiri Hilir.

### 2. Klaster Dua

Klaster dua memiliki nilai rata-rata atribut Pengeluaran Perkapita sebesar 10,835, Rata-rata Lama Sekolah sebesar 8,64, dan Harapan Lama Sekolah sebesar 12,7, nilai-nilai ini lebih dominan daripada Pengeluaran Perkapita, Rata-rata Lama Sekolah, dan Harapan Lama Sekolah , dan yang ada pada klaster satu. Klaster dua merupakan klaster dengan nilai yang lebih rendah dari pada klaster satu dan klaster tiga. Klaster dua memiliki enam anggota klaster, yaitu Kuantan Singingi, Indragiri Hulu, Pelalawan, Siak, Kepulauan Meranti, dan Kota Dumai.

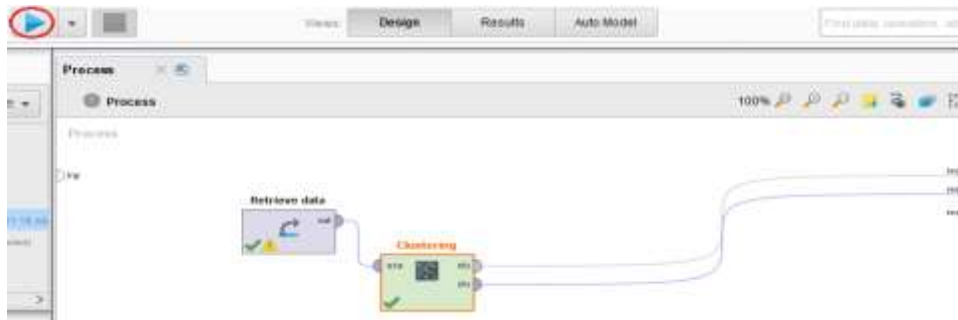
### 3. Kluster Tiga

Sedangkan untuk kluster tiga, mempunyai hasil yang tinggi di setiap atributnya dari pada nilai di kluster satu dan dua. Dan mempunyai nilai rata-rata yang jauh di atas kluster satu dan kluster dua. Kluster tiga hanya memiliki satu anggota kluster saja, yaitu Kota Pekanbaru.

## 3.2 Perhitungan Menggunakan Rapid Miner

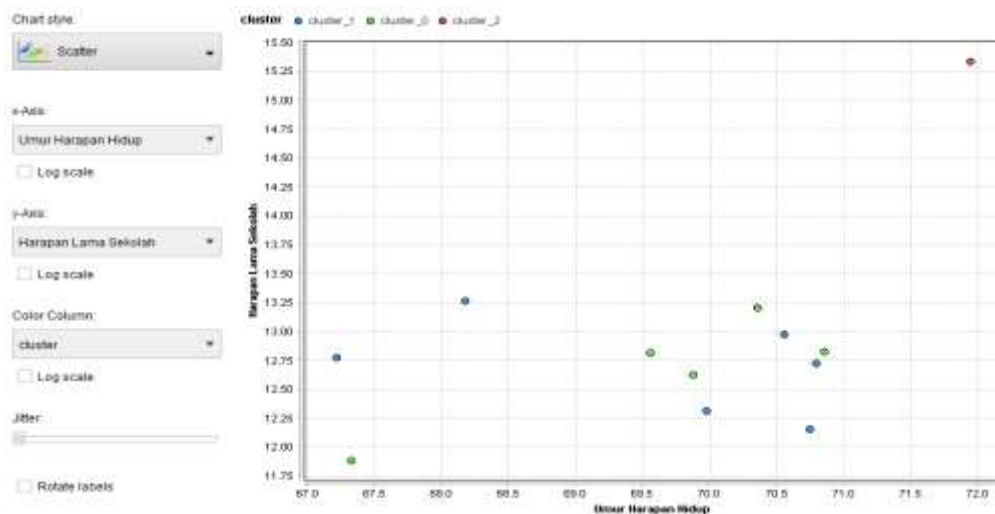
### 1. Konfigurasi Rapid Miner

Hubungkan Data yang kita miliki dengan metode *Clustering* dan sambungkan pada garis *Res* dan *Clu*, agar aplikasi dapat memulai perhitungan.



Gambar 4. Menyambungkan garis *Res* dan *Clu*

### 2. Hasil dari perhitungan



Gambar 5. Hasil Perhitungan

## Cluster Model

```
Cluster 0: 5 items
Cluster 1: 6 items
Cluster 2: 1 items
Total number of items: 12
```

Gambar 6. Hasil Perhitungan (1)

### 3.3 Perbandingan antara Perhitungan Manual dan Software

**Tabel 8.** Perbandingan Perhitungan Manual dan *Software*

| Perhitungan Manual |                                            |                | Software <i>RapidMiner</i> |                                            |                |
|--------------------|--------------------------------------------|----------------|----------------------------|--------------------------------------------|----------------|
| Cluster            | Anggota Kelompok                           | Jumlah Anggota | Cluster                    | Anggota Kelompok                           | Jumlah Anggota |
| 1                  | PCDK3, PCDK6, PCDK7, PCDK8, PCDK9          | 6              | 0                          | PCDK3, PCDK6, PCDK7, PCDK8, PCDK9          | 6              |
| 2                  | PCDK1, PCDK2, PCDK4, PCDK5, PCDK10, PCDK12 | 5              | 1                          | PCDK1, PCDK2, PCDK4, PCDK5, PCDK10, PCDK12 | 5              |
| 3                  | PCDK11                                     | 1              | 2                          | PCDK11                                     | 1              |

### 4. KESIMPULAN

Adapun kesimpulan yang bisa di ambil dari penelitian *Clustering* Pengelompokan Daerah Miskin di Provinsi Riau Menggunakan Metode *K-means*, yakni :

1. Berdasarkan hasil dari perhitungan manual dan juga pengujian dengan perangkat lunak *RapidMiner*, menunjukkan bahwa hasil pengujian bernilai sama. Yang artinya hasil pengujian bernilai sangat baik, sehingga dapat dijadikan sebagai referensi atau sebagai salah satu bahan pertimbangan bagi pemerintah untuk menyalurkan bantuan ataupun program-program bantuan pemerintah lainnya.
2. Berdasarkan pengelompokan 12 Kabupaten di Provinsi Riau, dapat dibentuk tiga buah *Cluster*. *Cluster* pertama termasuk dalam wilayah menengah, yang terdiri dari lima Kabupaten/kota, yakni : Rokan Hilir, Bengkalis, Rokan Hulu, Kampar, dan Indragiri Hilir. Sedangkan untuk *Cluster* kedua termasuk dalam wilayah kurang mampu, yang terdiri dari enam Kabupaten/kota, yakni : Siak, Pelalawan, Indragiri Hulu, Kuantan Singingi, Kep.Meranti, dan Kota Dumai. Lalu untuk *Cluster* Tiga termasuk dalam wilayah mampu, dan terdiri dari satu anggota saja, yakni : Kota Pekanbaru.
3. Dari hasil perhitungan rata-rata menunjukkan bahwa, *Cluster* Tiga dan *Cluster* Satu lebih dominan dibandingkan dengan *Cluster* Dua. Yang artinya, *Cluster* Dua seharusnya lebih di prioritaskan oleh pemerintah dalam pemberian bantuan atau program-program pemerintah lainnya dibandingkan dengan *Cluster* Tiga dan Juga *Cluster* Satu.

### DAFTAR PUSTAKA

- [1] Dita Prima Tri Hapsari, Edy Widodo, "Pengelompokan Daerah Rawan Kriminalitas di Indonesia Menggunakan Analisis *K-means Clustering*" Yogyakarta, Vol 1 No.1, 149, 2017.
- [2] C.T.I., B. R., Gafar, A. A., Fajriani, N., Ramdanif, U., Uyun, F. R., P, Y. P., & Ransi, N, "Implemetasi *K-means Clustering* pada rapidminer untuk analisis daerah rawan kecelakaan" *Seminar Nasional Riset Kuantitatif Terapan 2017, April*, 58–60, 2017.
- [3] Retno Tri Vlandari. "Data Mining (Teori dan Aplikasi *RapidMiner*)" Surakarta, Gava Media, 53, 2017.
- [4] Badan Pusat Statistik Provinsi DIY, "Berita Resmi Statistik" *DIY.Bps.Go.Id*, 27, 1–16. <https://jakarta.bps.go.id/pressrelease/2019/11/01/375/tingkat-penghunian-kamar-tpk--hotel--berbintang-dki-jakarta-pada-bulan-september-2019-mencapai-58-97-persen.html>, 2018.
- [5] Priambodo, Y. A., & Prasetyo, S. Y. J, "Pemetaan Penyebaran Guru di Provinsi Banten

- dengan Menggunakan Metode Spatial *Clustering K-means* (Studi kasus : Wilayah Provinsi Banten)” *Indonesian Journal of Computing and Modeling*, 1(1), 18–27, 2018. <https://doi.org/10.24246/j.icm.2018.v1.i1.p18-27>
- [6] Sormartini, Supartini Enny, “Analisis *K-means* Untuk Pengelompokan Kabupaten/kota di Jawa Barat Berdasarkan Indikator Masyarakat” Bandung, 146, 2017.
  - [7] Yohanes Aji Priambodo, Sri Yulianto Joko Prasetyo, “Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Methoded Spatial *Clustering K-means*” Salatiga, e-ISSN : 2598-9421, 2018.
  - [8] Fitri Larasati Sibuea, Andy Sapta, “Pemetaan Siswa Berprestasi Menggunakan Metode *K-means Clustering*” Vol IV, No.1, <http://jurnal.stmikroyal.ac.id/index/php/jurteks>, 2017.
  - [9] Aldi Nurzahputra, Much Aziz Muslim, Miranita Khusniati, “Penerapan Algoritma *K-means* untuk *Clustering* Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa” Semarang, Vol. 16 No.1, 2017.
  - [10] Siti Syahidatul Helma, Mustakim, Risma Rustiyan R, Eva Normala, “*Clustering* Pada Data Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma *K-means*” Pekanbaru, 133, 2017.
  - [11] S., Ramadhani. I., Ambarita. A.M.H., Pardede, “Metode *K-means* untuk Pengelompokan Masyarakat Miskin Dengan Menggunakan Jarak Kedekatan Manhattan City dan Euclidian (Studi Kasus Kota Binjai.)” Binjai, Vol 4, No.2 Juli 2019.
  - [12] Eric Fammaido, Lukman Hakim, “Penerapan Algoritma *K-means Clustering* untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar” Jakarta. ISSN : 2407-3911, Vol V, No. 1. 15 Desember 2018.
  - [13] Fitra Ramdhani, Abdul Hoyyi, Moch. Abdul Mukid, “Pengelompokan Provinsi di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode *K-means Cluster* Jurnal *Gaussian*” Vol 4, No.4, 876, 2015.



**Prosiding- SEMASTER:** Seminar Nasional Teknologi Informasi & Ilmu Komputer is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)

---