

ALGORITMA FUZZY C-MEANS DENGAN METODE ELBOW UNTUK MENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDEKS PEMBANGUNAN GENDER

David Adi Yuliantono¹, Hardian Oktavianto², Habibatul Azizah Al Faruq³

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Jln. Karimata No. 49 Jember Kode Pos 68121

email: davidadiyuliantono25@gmail.com

ABSTRAK

Indeks Pembangunan Gender (IPG) merupakan salah satu alat ukur tingkat keberhasilan capaian pembangunan yang sudah disiapkan untuk persoalan gender. IPG adalah ukuran pembangunan manusia berbasis gender yang dilihat dari tiga dimensi yakni dimensi panjang umur dan hidup sehat, pengetahuan, dan standar hidup layak. Secara umum karakteristik gender laki-laki dan perempuan berbeda secara biologis. Perbedaan tersebut tidak menjadi masalah jika disertai kesetaraan antara laki-laki dan perempuan, ketidakadilan yang terjadi dapat menimbulkan korban bagi laki-laki maupun perempuan. Oleh karena itu, dibutuhkan metode pengelompokan data IPG di Indonesia menggunakan algoritma Fuzzy C-Means. Data yang digunakan yaitu data IPG di Indonesia tahun 2019. Dari pengujian 2 sampai 10 cluster yang telah dilakukan, dihasilkan cluster optimum yang berada pada 2 cluster berdasarkan jarak Sum of Square Error (SSE) pada metode Elbow. Pada cluster 1 terdiri dari 12 anggota provinsi dan cluster 2 terdiri dari 22 anggota provinsi. Berdasarkan karakteristik data tahun 2016 sampai dengan 2019, cluster lebih rendah dibandingkan dengan cluster 2.

Kata Kunci: clustering, fuzzy c-means, elbow, IPG

ABSTRAK

The Gender Development Index (IPG) is one measure of the development achievements success rate of that have been prepared to deal with gender. IPG is a form of human development based on sex seen from three dimensions; such as length and healthy living, knowledge, and a decent standard of living dimensions. In general, gender characteristics of men and women differ biologically. This difference is not a problem when it comes to equality between men and women, the injustices that occur can cause casualties for men and women. Therefore, we need a method for grouping IPG data in Indonesia by using the Fuzzy C-Means algorithm. The data used are IPG data in Indonesia in 2019. From the testing of 2 to 10 clusters that have been carried out, an optimal cluster is generated in 2 clusters based on the distance of Sum from Square Error (SSE) on the Elbow method. Cluster 1 consists of 12 provincial members and Cluster 2 consists of 22 provincial members. Based on the characteristics of the data from 2016 to 2019, cluster is lower than cluster 2.

keywords: clustering, fuzzy c-means, elbow, IPG

1. PENDAHULUAN

Indeks Pembangunan Gender (IPG) merupakan salah satu alat ukur tingkat keberhasilan capaian pembangunan yang sudah disiapkan untuk persoalan gender. IPG adalah ukuran pembangunan manusia berbasis gender yang dilihat dari tiga dimensi yakni dimensi panjang umur dan hidup sehat, pengetahuan, dan standar hidup layak. Secara umum karakteristik gender laki-laki dan perempuan berbeda secara biologis. Perbedaan tersebut tidak menjadi masalah jika disertai kesetaraan antara laki-laki dan perempuan, ketidakadilan yang terjadi dapat menimbulkan korban bagi laki-laki maupun perempuan. Oleh karena itu, kesetaraan merupakan hal yang harus didapatkan pihak laki-laki maupun pihak perempuan agar kedua belah pihak mendapatkan peluang ataupun kesempatan yang sama untuk berperan dalam setiap aspek kehidupan.

Permasalahan ketidaksetaraan gender menjadi permasalahan hampir di setiap negara, termasuk Indonesia. Menurut Fajriyyah & Budiantara (2015) Indonesia memiliki IPG yang rendah jika dibandingkan dengan negara lain seperti Malaysia dan Australia. Oleh karena itu, di Indonesia masih terjadi kesetimpangan yang menandakan belum adanya pemerataan pembangunan yang dirasakan oleh seluruh rakyat. Dilihat dari pemberdayaan di bidang ekonomi perempuan hanya diberi upah kerja sebesar 76,43 % jauh lebih rendah dibandingkan dengan upah laki-laki secara rata-rata (UNDP, 2015). Berdasarkan hasil perhitungan Badan Pusat Statistik (2019), pembangunan laki-laki dan perempuan di Indonesia mengalami peningkatan 9 tahun terakhir. Berdasarkan data Badan Pusat Statistik, jumlah angkatan kerja pada Februari 2019 meningkat sebesar 2,24 juta orang jika dibandingkan dengan Februari 2018 menjadi 138,18 juta orang. Dengan kenaikan jumlah angkatan kerja, tingkat partisipasi angkatan kerja (TPAK) naik 0,12% menjadi 69,32% dibandingkan periode tahun lalu. Hal ini mengindikasikan adanya potensi ekonomi dari sisi pasokan tenaga kerja meningkat dan jika dilihat berdasarkan gender, TPAK masih didominasi laki-laki dengan partisipasi sebesar 83,18% dan TPAK

perempuan hanya sebesar 55,5%. Untuk mengklasterisasi provinsi yang ada di Indonesia yakni untuk meningkatkan pembangunan manusia berdasarkan gender dan kesadaran masyarakat atas adanya ketidaksetaraan gender yang selama ini ada di masyarakat.

Pada penelitian sebelumnya yang dilakukan oleh Azuri, Zulhanif, & Potoh (2016) dengan studi kasus "Pengelompokan Kabupaten/Kota di pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisecting K-means". Pada penelitian tersebut, peneliti menggunakan data IPG tahun 2015 di pulau Jawa. Dari penelitian tersebut menghasilkan data 119 yang dibagi menjadi 3 cluster dan untuk mencari cluster optimum peneliti menggunakan metode Elbow & Silhouette. Penelitian lainnya dilakukan oleh Saputra dan Rikosakomara (2018) dengan penelitian "Implementasi Fuzzy C-Means dan Model RFM untuk Segmentasi Pelanggan (Studi Kasus: PT. XYZ)" menggunakan algoritma Fuzzy C-Means dan metode Elbow untuk mencari cluster terbaik. Karena ketidaksetaraan gender masih menjadi masalah, penulis melakukan penelitian terhadap data IPG dari tahun 2016 sampai dengan tahun 2019 di 34 provinsi di Indonesia. Agar proses pengelompokan data lebih cepat, maka penelitian ini menggunakan Fuzzy C-Means yang merupakan salah satu pengelompokan data yang tiap-tiap titik data dalam suatu cluster ditentukan dengan derajat keanggotaan, dan metode Elbow mudah untuk diimplementasikan dengan cara melihat grafik dari jumlah cluster yang telah diinputkan. Oleh karena itu, penelitian ini dibuat dengan judul "Algoritma Fuzzy C-Means dengan Metode Elbow untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Indeks Pembangunan Gender".

2. TINJAUAN PUSTAKA

2.1 *Data Mining*

Data mining adalah aktivitas yang menggambarkan sebuah proses analisis yang terjadi secara iteratif pada *database* yang besar, dengan tujuan mengekstrak informasi

dan *knowledge* yang akurat dan berpotensi berguna untuk *knowledge workers* yang berhubungan dengan pengambilan keputusan dan pemecahan masalah (Vercellis, 2009 dalam Waworuntu & Amin, 2018).

Menurut Larose & Daniel (dalam Susanto & Sudiyatno, 2014) *data mining* dilakukan dengan *tool* khusus, yang mengeksekusi operasi *data mining* yang telah didefinisikan berdasarkan model analisis. *Data mining* merupakan proses analisis terhadap data dengan penekanan menemukan informasi yang tersembunyi pada jumlah data besar yang disimpan ketika menjalankan bisnis perusahaan. Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor antara lain:

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam *data warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam *database*.
3. Adanya peningkatan akses data melalui navigasi web dan internet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan perkembangan kapasitas media penyimpanan.

2.2 Clustering

Menurut Nugraha, dkk. (2014) *Clustering* merupakan salah satu teknik *data mining* yang digunakan untuk mendapatkan kelompok-kelompok dari objek-objek yang mempunyai karakteristik yang umum di data yang cukup besar. Tujuan utama dari metode

clustering adalah pengelompokan sejumlah data atau objek kedalam *cluster* atau grup sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin. *Clustering* melakukan pengelompokan data yang didasarkan pada kesamaan antar objek, oleh karena itu klusterisasi digolongkan sebagai metode *unsupervised learning*.

Pada dasarnya *clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target *output*. Dalam *data mining* ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierachical clustering* dan *non-hierarchiacal clustering* (Santosa, 2007).

2.3 Fuzzy C-Means

Fuzzy C-Means pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. *Fuzzy C-Means* merupakan metode *clustering* dengan pendekatan *fuzzy*, artinya setiap data yang di *cluster* memungkinkan menjadi anggota lebih dari satu *cluster*. Konsep dasar *Fuzzy C-Means* adalah menentukan pusat *cluster*, pada kondisi awal pusat *cluster* ini masih belum akurat dan setiap data memiliki derajat keanggotaan untuk tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster*.

Ada beberapa algoritma *clustering* data, salah satu diantaranya adalah *Fuzzy C-Means*. *Fuzzy C-Means* adalah salah satu teknik klasterisasi data di mana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. *Output* dari *Fuzzy C-Means* bukan merupakan *fuzzy inference system*, namun merupakan deretan pusat *cluster* dan beberapa derajat keanggotaan untuk tiap-tiap titik data. Informasi ini dapat digunakan untuk membangun suatu *fuzzy inference system* (Kusumadewi & Purnomo, 2010).

Algoritma *Fuzzy C-Means* adalah sebagai berikut:

1. *Input* data yang akan di *cluster* X , berupa matriks berukuran $n \times m$ (n =jumlah sampel data, m =atribut setiap data). X_{ij} =data sampel ke- i ($i=1,2,\dots,n$), atribut ke- j ($1,2,\dots,n$).
2. Menentukan beberapa *input* yang dibutuhkan dalam perhitungan *Fuzzy C-Means*, yaitu:
 - Jumlah *cluster* (c)
 - Pangkat (w)
 - Maksimum iterasi ($MaxIter$)
 - *Error* terkecil (ξ)
 - Fungsi objektif awal ($P_0 = 0$)
 - Iterasi awal ($t = 1$)
3. Membangkitkan bilangan *random* μ_{ik} , $i=1,2,\dots,n$; $k=1,2,\dots,c$; sebagai elemen-elemen matriks partisi awal U . Hitung jumlah setiap kolom

$$Q_i = \sum_{k=1}^c \mu_{ik}$$

Q_i ialah jumlah setiap kolom dari nilai *random* sebuah matrik, jumlah Q tergantung dari beberapa jumlah kriteria penilaian.

4. Menghitung pusat *cluster* ke- k : V_{kj} , dengan $k=1,2,\dots,c$; dan $j=1,2,\dots,m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$

i = iterasi

μ_{ik} = perubahan matriks

X_{ij} = atribut

V_{kj} ialah titik pusat setiap *cluster*, jumlah V_{kj} tergantung dari berapa *cluster* yang akan dibentuk dan n ialah jumlah proposal.

5. Menghitung fungsi objektif pada iterasi ke- t ,

P_t

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right)$$

P_t = fungsi objektif

$\sum_{i=1}^n$ = jumlah data di *cluster*

$\sum_{k=1}^c$ = jumlah perhitungan *cluster* awal

t merupakan iterasi yang dihitung, jika iterasi dimulai dari 1 maka pada awal perhitungan nilai t ialah 1. Iterasi akan berulang sesuai dengan ketentuan iterasi yang sedang berjalan. Hitung perubahan matrik partisi.

$$\mu_{ik(t)} = \frac{\left[\sum_{j=1}^m (x_{ij} - v_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (x_{ij} - v_{kj})^2 \right]^{\frac{-1}{w-1}}}$$

Iterasi akan tetap berulang jika nilai atau kondisi-kondisi tertentu belum tercapai, adapun kondisi tersebut ialah jika $(|P_t - P_{t-1}| < \xi)$ atau $(t > MaxIter)$ maka berhenti yang mana P_t ialah pusat *cluster* iterasi ke t kekurangan dari nilai *error* yang diharapkan atau jika t (jumlah iterasi) sudah lebih besar daripada iterasi maksimum. Namun jika iterasi akan diulang lagi dengan $t+1$ akan mengulang proses yang ke-4 atau menghitung pusat *cluster* lagi (Kusumadewi & Purnomo, 2010).

2.4 Metode *Elbow*

Metode *Elbow* merupakan suatu metode penentuan jumlah *cluster* optimum atau terbaik untuk menghasilkan suatu informasi dengan cara melihat presentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Metode ini memberikan ide atau gagasan dengan cara memilih nilai *cluster* dan kemudian menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan *cluster* terbaik. Selain itu, presentase perhitungan yang dihasilkan menjadi pembanding antara jumlah *cluster* yang ditambah (Merliana, dkk., 2015).

Menurut Kodinariya & Makwana (dalam Alatubir, 2017) untuk mendapatkan perbandingannya adalah dengan menghitung *Sum of Square Error* (SSE) dari masing-masing nilai *cluster*. Karena semakin besar jumlah *cluster* K maka nilai SSE akan semakin kecil.

Rumus SSE sebagai berikut :

$$SSE = \sum_{k=1}^k \sum_{xi \in Sk} \|Xi - Ck\|_2^2$$

Keterangan:

X_i = fitur atau atribut dari data ke - i

C_k = fitur atau atribut pusat *cluster* ke i

Algoritma metode *Elbow* dalam menentukan nilai K pada *K-Means* (Bholowalia, dkk. 2014 dalam Alatubir, 2017):

1. Inisialisasi awal nilai K
2. Naikkan nilai K
3. Hitung hasil *sum of square error* dari tiap nilai K
4. Lihat hasil *sum of square error* dari nilai K yang turun secara drastis
5. Tetapkan nilai K yang berbentuk siku

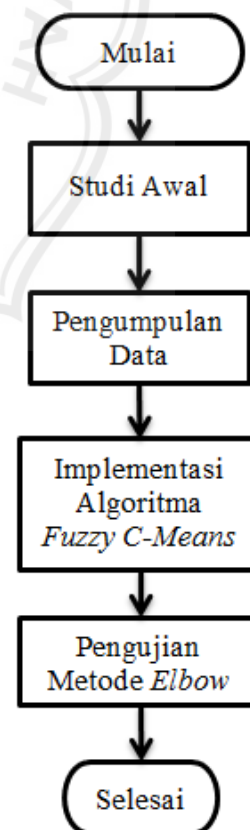
2.5 RStudio

Rstudio adalah aplikasi yang digunakan untuk mempermudah pengolahan bahasa R. Bahasa R yaitu bahasa pemrograman untuk komputasi statistika dan grafik. Rstudio didirikan oleh JJ Allaire pada tahun 2008. Rstudio memiliki dua versi, yaitu *open source* (gratis) dan *commercial edition* (berbayar). Rstudio juga tidak hanya terbatas dalam bentuk aplikasi dekstop (*Windows, MacOS, dan linux*), melainkan terdapat versi Rstudio server, yaitu Rstudio yang dapat diakses melalui *browser* yang terhubung dengan suatu jaringan komputer.

3. METODOLOGI

3.1 Tahapan Penelitian

Tahapan penelitian yang dilakukan dalam menggunakan algoritma *Fuzzy C-Means* untuk mengelompokkan provinsi di Indonesia berdasarkan indek pembangunan *gender*, memiliki tahapan sebagai berikut:



Gambar 1. Diagram Alur Penelitian

3.2 Studi Awal

Studi awal pada penelitian ini adalah mencari dan mempelajari masalah kesetaraan *gender* yang selama ini ada di masyarakat, kemudian menentukan ruang lingkup masalah, latar belakang, serta melakukan observasi dengan cara membaca beberapa artikel dan berita yang berkaitan dengan permasalahan ketidaksetaraan *gender* di masyarakat. Untuk mencapai tujuan, penulis mencari dan mempelajari metode pengelompokan data dengan membaca beberapa paper, artikel ilmiah dan makalah, kemudian peneliti menggunakan metode pengelompokan data sebagai salah satu solusi permasalahan *gender* yang selama ini ada di masyarakat.

3.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari situs resmi Badan Pusat Statistik yang berjudul IPG 2010-2019 yakni <https://www.bps.go.id/dynamictable/2018/08/15/1569/-ipg-indeks-pembangunan-gender-ipg-2010-2017.html>. Data yang digunakan dalam penelitian ini adalah data Indeks Pembangunan Manusia di setiap Provinsi di Indonesia pada tahun 2016 sampai 2019 yang terdiri dari 34 provinsi.

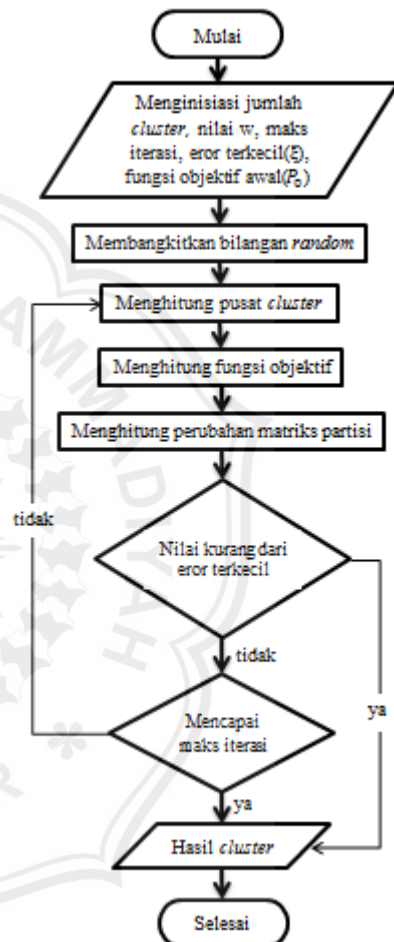
3.4 Dataset

Dataset pada penelitian ini adalah data IPG mulai tahun 2016 sampai dengan tahun 2019 yang dimana terdiri dari 136 data dari 34 provinsi di Indonesia. Data ini dikelompokkan menggunakan algoritma *Fuzzy C-Means* dan menggunakan metode *Elbow* sebagai metode optimasi pengelompokan guna mengetahui jumlah *cluster* optimum yang akan digunakan. Pada bab ini, data yang digunakan hanya 15

data sampel untuk mewakili data yang dihitung.

3.5 Implementasi Algoritma *Fuzzy C-Means*

Diagram alur algoritma *Fuzzy C-Means* yang digunakan berdasarkan data IPG di Indonesia, pada umumnya alur diagram *Fuzzy C-Means* adalah sebagai berikut:



Gambar 2. Flowchat *Fuzzy C-Means*

4. HASIL DAN PEMBAHASAN

4.1 Data Pengujian

Berdasarkan hasil penelitian yang telah didapat dari perhitungan yang telah dilakukan, data IPG yang telah dikumpulkan akan di *cluster* menggunakan algoritma *Fuzzy C-Means* kemudian diolah untuk mendapatkan hasil *cluster* optimum atau *cluster* terbaik dengan menggunakan metode *Elbow*. Data

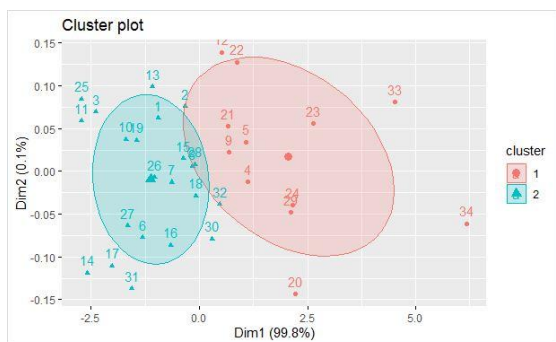
yang digunakan adalah data IPG pada tahun 2016 sampai tahun 2019 yang terdiri dari 136 data yang terdiri dari 34 provinsi.

No	Provinsi	Tahun			
		2016	2017	2018	2019
1	ACEH	91.89	91.67	91.67	91.84
2	SUMATERA UTARA	90.82	90.65	90.66	90.71
3	SUMATERA BARAT	94.42	94.16	94.17	94.09
4	RIAU	88.04	88.17	88.37	88.43
5	JAMBI	88.29	88.13	88.44	88.44
6	SUMATERA SELATAN	92.08	92.43	92.62	92.40
7	BENGGULU	91.06	91.34	91.37	91.19
8	LAMPUNG	90.30	90.49	90.57	90.39
9	KEP. BANGKA BELITUNG	88.90	88.93	89.15	89.00
10	KEPULAUAN RIAU	93.13	92.96	92.97	93.10
11	DKI JAKARTA	94.98	94.70	94.70	94.71
12	JAWA BARAT	89.56	89.18	89.19	89.26
13	JAWA TENGAH	92.22	91.94	91.95	91.89
14	D I YOGYAKARTA	94.27	94.39	94.73	94.77
15	JAWA TIMUR	90.72	90.76	90.77	90.91
16	BANTEN	90.97	91.14	91.30	91.67
17	BALI	93.20	93.70	93.71	93.72
18	NUSA TENGGARA BARAT	90.05	90.36	90.37	90.40
19	NUSA TENGGARA TIMUR	92.72	92.44	92.57	92.72
20	KALIMANTAN BARAT	85.77	86.28	86.74	86.81
21	KALIMANTAN TENGAH	89.07	88.91	89.13	89.09
...
34	PAPUA	79.09	79.38	80.11	80.05

Tabel 1. Data indeks Pembangunan Gender

4.2 FuzzyC-Means Pada Rstudio

Data IPG diolah menggunakan RStudio yang di *cluster* menggunakan algoritma *Fuzzy C-Means* dari 2 sampai 10 *cluster*. Hasil *cluster* yang didapat dari perintah pada RStudio adalah Jumlah Iterasi, Pusat *cluster*, fungsi objektif, dan derajat keanggotaan setiap objek terhadap tiap *cluster*. Hasil dari perintah *Fuzzy C-Means* pada RStudio ditampilkan ke dalam *plot* dari setiap *cluster* yang terbentuk. Berikut adalah contoh *plot* pada 2 *cluster* hasil perintah *Fuzzy C-Means* di RStudio:



Gambar 3. Plot 2 cluster Fuzzy C-Means

Keterangan:

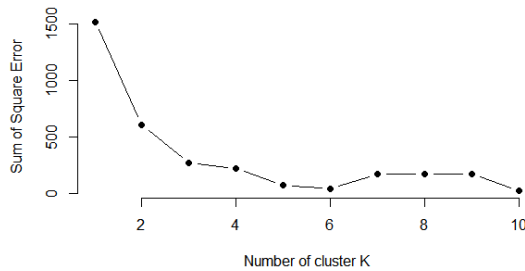
- Pada *cluster* 1 ditandai dengan warna merah terdapat 12 provinsi yaitu Riau, Jambi, Kep. Bangka Belitung, Jawa Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Gorontalo, Papua Barat, dan Papua.
- Pada *cluster* 2 ditandai dengan warna biru terdapat 22 provinsi yaitu Aceh, Sumatra Utara, Sumatra Barat, Sumatra Selatan, Bengkulu, Lampung, Kepulauan Riau, DKI Jakarta, Jawa Tengah, DI Yogyakarta Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Maluku, dan Maluku Utara.

4.3 Penentuan Jumlah Cluster Optimum

Setelah melakukan proses klasterisasi menggunakan *Fuzzy C-Means*, kemudian diproses menggunakan metode *Elbow* untuk menentukan *cluster* optimum atau *cluster* terbaik. Hasil perhitungan metode *Elbow* pada RStudio ditampilkan dalam nilai *Sum of Squares Error (SSE)* dan grafik yang terdiri dari sumbu x dan sumbu y. Nilai pada sumbu x adalah *Sum of Square Error (SSE)* yang dihasilkan dari pusat setiap *cluster*. Berikut ini adalah hasil metode *Elbow*:

C	SSE	Jarak	Keterangan
1	1520,73284	-	-
2	60739005	-60737484,3	Jarak C1 ke C2
3	27047936	33691069	Jarak C2 ke C3
4	22384586	4663350	Jarak C3 ke C4
5	7861512	14523074	Jarak C4 ke C5
6	4276195	3585317	Jarak C5 ke C6
7	17747404	-13471209	Jarak C6 ke C7
8	17556150	191254	Jarak C7 ke C8
9	17495459	60691	Jarak C8 ke C9
10	2594858	14900601	Jarak C9 ke C10

Tabel 2. Hasil Nilai Elbow



Gambar 4. Plot Metode *Elbow*

4.4 Profiling *Cluster* Optimum

Cluster optimum yang dapat dilihat pada gambar 4.2 nilai *cluster* yang diambil sebagai *cluster* optimum atau *cluster* terbaik adalah titik yang membentuk siku. Titik yang membentuk siku adalah dimana titik yang terjadi penurunan yang signifikan antara 2 titik *cluster* dan kemudian diikuti oleh nilai yang relatif konstan. Pada tabel 4.2 menunjukkan nilai *Sum of Square Error* (SSE) dengan jarak paling signifikan atau paling besar terdapat pada titik *cluster* 2 dengan jarak 1 *cluster* ke 2 *cluster* yaitu -607374484,3. Nilai jarak 1 *cluster* ke 2 *cluster* merupakan nilai jarak yang mengalami penurunan paling signifikan atau paling besar dan kemudian diikuti dengan nilai jarak yang relatif konstan, sehingga dapat disimpulkan bahwa 2 *cluster* merupakan *cluster* optimum atau *cluster* terbaik. Dapat dilihat juga pada grafik pada gambar 4.2 bahwa penurunan nilai yang signifikan berada pada titik 2 *cluster* dibandingkan dengan yang lain. Kemudian dari titik 2 *cluster* ke titik selanjutnya diikuti nilai yang relatif konstan. Jadi, *cluster* optimum terdapat pada 2 *cluster*.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan permasalahan pada bab sebelumnya, maka

penerapan *Fuzzy C-Means* pada data Indeks Pembangunan *Gender* (IPG) untuk mengelompokkan provinsi di Indonesia menghasilkan 2 *cluster* optimum dengan jarak *Sum of Square Error* (SSE) antara titik 1 *cluster* ke titik 2 *cluster* pada metode *Elbow* yaitu -607374484,3. Jarak 1 *cluster* ke 2 *cluster* tersebut merupakan nilai jarak yang mengalami penurunan paling signifikan dibanding dengan yang lain dan diikuti nilai jarak yang relatif konstan.

Hasil pengelompokan yang didapat adalah *cluster* 1 terdapat 12 provinsi yaitu Riau, Jambi, Kep. Bangka Belitung, Jawa Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Gorontalo, Papua Barat, Papua. Pada *cluster* 2 terdapat 22 provinsi yaitu Aceh, Sumatra Utara, Sumatra Barat, Sumatra Selatan, Bengkulu, Lampung, Kepulauan Riau, DKI Jakarta, Jawa Tengah, Jawa Timur, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Maluku, Maluku Utara. Pada hasil karakteristik data tahun 2016 sampai dengan tahun 2019, *cluster* 1 memiliki jumlah ketidakadilan pencapaian laki-laki dan perempuan lebih rendah dibandingkan dengan *cluster* 2.

5.2 Saran

Saran pada penelitian ini adalah:

1. Perhitungan *manual cluster Fuzzy C-Means* pada *Microsoft Excel* dapat menggunakan bilangan *random* yang diambil dari RStudio untuk menentukan *cluster* awal supaya menghasilkan pengelompokan yang serupa dengan RStudio.

2. Untuk mencari *cluster* terbaik, bisa menggunakan metode alternatif lain seperti *Silhouette*, *Gap Statistic*, *Davies Bouildin Index*, dll.

6. DAFTAR PUSTAKA

- Anonim. Badan Pusat Statistik. 2019. *Pembangunan Manusia Berbasis Gender 2019*. Jakarta: Badan Pusat Statistik.
- Anonim. Sirusa Badan Pusat Statistik. 2010. *Indeks Pembangunan Gender (IPG)*. Jakarta: Badan Pusat Statistik.
- Anonim. Kementrian Pemberdayaan Perempuan Dan Perlindungan Anak. 2013. *Pembangunan Manusia Berbasis Gender*. Jakarta: Badan Pusat Statistik.
- Anonim, "Tingkat Partisipasi Angkatan Kerja Menurut Jenis Kelamin", 7 Mei 2019, <<https://databoks.katadata.co.id/datapublish/2019/05/07/tingkat-partisipasi-tenaga-kerja-masih-didominasi-laki-laki>>. Diakses pada 20 Maret 2019.
- Alatubir, D. A. V. 2017. *Penerapan Algoritma K-Means untuk Pengelompokan Sekolah Menengah Atas Di Provinsi Daerah Istimewa Yogyakarta Berdasarkan Nilai Daya Serap Ujian Nasional Bahasa Indonesia*. Yogyakarta: Fakultas Sains Dan Teknologi Universitas Sanata Dharma..
- Azuri, D. F., Zulhanif & Pontoh, R. S. 2016. *Pengelompokan Kabupaten/Kota Dipulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Besecting K-Means*. Jatinagor: Prosiding Seminar Nasional FMIPA Universitas Padjadjaran.
- Bwolowalia, P. & Kumar, A. 2014. RBK-Means: *A Clustering Techiniques based on Elbow Method and K-Means in WSN. International Journal of Computer Aplication (0975-8887), IX(105), 17-24.*
- Fajriyyah, N., & Budiantara, I Nyoman. 2015. *Permodelan Indeks Pembangunan Gender dengan Pendekatan Regresi Nonparametrik Spline di Indonesia*. Surabaya: Jurnal Sains Dan Seni ITS Vol. 4, No.2, (2015) 2337-3520 (2301-928XPrint).
- Kodinariya, T. M. & Makwana, P. R. 2013. *Riview on determining number of cluster in K-Means Clustering. Internasional Journalof Advance Research in Computer Science and Management Studies,1(6), 90-95.*
- Kusumadewi, S., & Purnomo, H. 2010. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan edisi 2*. Yogyakarta: Graha Ilmu.
- Larose, & Daniel, T. 2005. *Discovering Knowledge In Data: An Introduce to Data Mining*. Canada: John Willey and Sons, Inc.
- Merliana, N. P. E., Ermawati, & Santoso, A. J. 2015. *Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering*. Yogyakarta: Program Studi Magister Teknik Informatika Universitas Atma Jaya.
- Nugraha, D. D. C., Naimah, Z., Fahmi, M. & Setiani, N. 2014. *Klasterisasi Judul Buku dengan Menggunakan Metode K-Means*. Seminar Nasional Aplikasi Teknologi Informasi. ISSN: 1907-5022. G-2.
- Santosa, B. 2007. *Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.

Saputra & Rikosakomara. 2018. *Implementasi Fuzzy C-Means dan Model RFM untuk Segmentasi Pelanggan (Studi Kasus: PT. XYZ)*. Surabaya: Jurnal Teknik ITS Vol 7, No. 1 (2018) 2337-3520 (2301-928X Print).

Waworuntu, M. N. V. & Amin M. F. 2018. *Penerapan Metode K-Means Pemetaan Calon Penerima Jamkesda*. Banjarbaru: Kumpulan jurnal Ilmu Komputer (KLIK) Volume 05, No.02 September 2018. ISSN: 2406-7857.

