

Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi

Irmanita Nasution*, Agus Perdana Windarto, M. Fauzan

Prodi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: ^{1,*}Irmanitanst30@gmail.com, ²agusperdana@amiktunasbangsa.ac.id, ³m.fauzan@stikomtb.ac.id

Abstrak

Kemiskinan merupakan salah satu masalah yang menghambat pertumbuhan nasional maupun regional. Pada penelitian ini data yang digunakan bersumber dari Badan Pusat Statistik tahun 2012-2018. Penelitian ini menggunakan teknik data mining. Dalam proses pengolahan data dengan metode K-means. Metode K-means merupakan suatu metode pengelompokan data yang ada ke dalam beberapa kelompok dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan memiliki karakteristik yang berbeda dengan data yang ada dalam kelompok yang lain. Jumlah record yang digunakan sebanyak 34 provinsi yang dibagi menjadi 2 cluster yakni cluster tinggi dan cluster rendah. Tujuan dari penelitian ini terbagi menjadi 2 bagian, yaitu kelompok provinsi dengan tingkat kemiskinan tertinggi dan kelompok provinsi dengan tingkat kemiskinan terendah. Dari hasil pengelompokan diperoleh sebanyak 8 cluster tinggi dan 26 cluster rendah. Diharapkan penelitian ini dapat memberikan masukan kepada pemerintah agar dapat memberikan perhatian lebih kepada provinsi yang termasuk kategori tinggi kemiskinan.

Kata Kunci: Kemiskinan; Data mining; Pengelompokan; K-means; RapidMiner

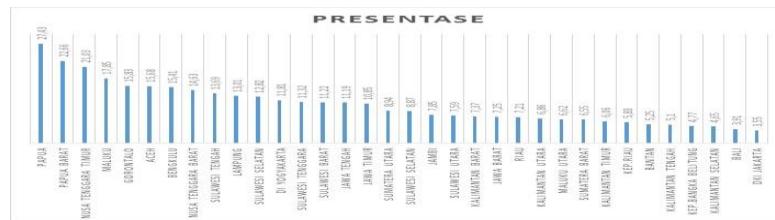
Abstract

Property is one of the problems that inhibits national and regional growth. This research uses data mining techniques. In this study the data used were sourced from the 2012-2018 statistical center. The research uses data mining techniques. In the data processing using k-means method. K-means method is a method of grouping existing data into several groups where the data in one group has the same characteristics with each other and has different characteristics from the data in other groups. The number of records used is 34 provinces which are divided into 2 clusters namely high and low clusters. The purpose of this study is divided into 2 parts, namely the provincial group with a high poverty rate and the provincial group with the lowest poverty level. From the result of grouping there were 8 provinces of high cluster and 26 low clusters. It is hoped that this research can provide input to the government so that it can give more attention to provinces that are categorized as high in poverty.

Keywords: Property; Data Mining; Clustering; K-Means; RapidMiner

1. PENDAHULUAN

Kemiskinan merupakan salah satu masalah yang menghambat pertumbuhan nasional maupun regional. kemiskinan dikatakan sebagai masalah multidimensi apabila berkaitan dengan ketidak-mampuan akses secara ekonomi, sosial, budaya, politik dan partisipasi dalam masyarakat [1]. Kondisi suatu negara dan situasi global merupakan permasalahan yang mengakibatkan kemiskinan. Indonesia merupakan salah satu negara berkembang yang mempunyai kepadatan penduduk nomor 4 di dunia setelah China, India, dan Amerika Serikat. Salah satu tolak ukur dalam keberhasilan pembangunan suatu negara dapat dilihat dari menurunnya jumlah penduduk miskin. Badan Pusat Statistik (BPS) mencatat bahwa pada September 2018 jumlah penduduk miskin di negara Indonesia sebanyak 25,26 juta orang atau sekitar 9,66% dan termasuk ke dalam kategori negara dengan jumlah penduduk miskin terbesar di dunia. Berikut ini adalah grafik tingkat kemiskinan di seluruh provinsi Indonesia dapat dilihat pada gambar berikut.



Gambar 1. Presentase Tingkat Kemiskinan di Indonesia

(Sumber : Badan Pusat Statistik (BPS), 2018)

Dari gambar 1. menjelaskan bahwa hampir seluruh provinsi di Indonesia memiliki tingkat kemiskinan tinggi. Provinsi Papua berada di posisi tertinggi dengan tingkat kemiskinan sebesar 27,43% dan DKI Jakarta berada di posisi terendah dengan tingkat kemiskinan sebesar 3,55%. Hal ini di sebabkan dari berbagai faktor yaitu laju pertumbuhan penduduk besar, angka pengangguran tinggi, pertumbuhan ekonomi yang lambat, tingkat pendidikan yang rendah dan distribusi yang tidak merata. Kemiskinan dapat menyebabkan banyak dampak terhadap perkembangan negara karena mengakibatkan negara tersebut tidak dapat bersaing di masa yang akan datang.



Banyak indikator pendukung yang menyebabkan data tingkat kemiskinan tiap provinsi berbeda. Untuk mempermudah mengenal tingkat kemiskinan suatu wilayah perlu adanya penentuan dan pengukuran indikator-indikator kemiskinan. Misalnya dengan melihat jumlah penduduk dan pendapatan ekonomi. Oleh karena itu, perlu adanya gambaran wilayah mana yang perlu penanganan cepat dan serius. Sulitnya dalam menentukan wilayah mana yang mengalami tingkat kemiskinan yang tinggi menyebabkan pemerintah kurang dalam penanganannya. Untuk mengatasi masalah tersebut perlu adanya pengclusteran untuk mengetahui provinsi mana yang termasuk *cluster* provinsi tingkat kemiskinan tinggi dan *cluster* provinsi tingkat kemiskinan rendah. Sehingga pemerintah dapat dengan cepat dan tepat dalam mengambil kebijakan untuk menanggulangi kemiskinan disuatu provinsi.

Berdasarkan uraian di atas banyak cabang kecerdasan buatan dalam ilmu komputer yang dapat menyelesaikan permasalahan tersebut secara kompleks, salah satunya *data mining*. *Data mining* merupakan proses analisa data untuk menemukan suatu pola dari kumpulan data. Beberapa penelitian terdahulu tentang kemiskinan oleh [2] diperoleh penilaian berdasarkan pemetaan karakteristik dari setiap kelompok yang terbentuk berdasarkan nilai rata-rata tertinggi dan terendah dari setiap provinsi Jawa Barat tahun 2018 didapat 5 cluster dengan cluster ke-1 yang terdiri dari Bogor, Sukabumi, Cianjur, Garut, Ciamis, Subang, Purwakarta dan Pangandaran sebagai wilayah yang diprioritaskan dalam peningkatan kesejahteraan penduduknya. Penelitian selanjutnya oleh [3] menunjukkan bahwa data diolah dengan menggunakan *k-means clustering* untuk menemukan pola yang memberikan gambaran tentang *cluster* setiap wilayah.

Dari penelitian terdahulu dapat dilihat bahwa penelitian tersebut hanya mengacu pada 1 provinsi saja. Disini, penulis menerapkan sebuah metode yaitu *data mining*. Salah satu teknik dari *data mining* adalah *Clustering*. *Clustering* merupakan metode analisis data yang tujuannya mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama. Salah satu algoritma dari *clustering* adalah *K-means*. *Metode K-means* adalah suatu metode penganalisaan data yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Diharapkan dengan adanya penelitian ini menjadi masukan kepada pemerintah agar pemerintah dapat menangani penyebarluasan bantuan untuk daerah tingkat penduduk miskin tinggi yang berdampak pada pertumbuhan ekonomi di Indonesia.

2. METODE PENELITIAN

2.1 Data Mining

Menurut [4] *Data mining* merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran computer secara otomatis untuk menganalisa dan mengekstraksi pengetahuan. *Data mining* sering disebut dengan istilah *Knowledge discovery in database (KDD)*. *Knowledge discovery in database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [5][6].

2.2 K-Means

Menurut [2] "K-Means merupakan salah satu metode *clustering* yang menggunakan konsep *descriptive* dan dapat digunakan untuk menjelaskan algoritma dalam penentuan suatu objek kedalam *cluster* tertentu berdasarkan rataan terdekat". *K-Means* pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976. Pada algoritma *K-Means* jumlah *cluster* telah ditentukan terlebih dahulu. Kelebihan *K-Means* yaitu mampu mengelompokkan objek besar dan penciran objek dengan sangat cepat sehingga mempercepat proses pengelompokan. Adapun kekurangan yang dimiliki oleh *K-Means* yaitu sangat sensitif terhadap pembangkitan titik pusat awal secara random dan hasil pengelompokan bersifat tidak unik (selalu berubah-ubah).

Langkah-langkah metode *k-means clustering* diantaranya [7], [8]:

- Tentukan jumlah *cluster*.
- Inisialisasi ke pusat *cluster* bisa dilakukan dengan berbagai cara. Cara yang sering dilakukan adalah dengan random atau acak.
- Alokasikan semua data/objek ke *cluster* terdekat. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* dengan persamaan :

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Dimana $D(i, j)$ = jarak data ke pusat *cluster* j .

X_{ki} = Data ke i pada tribut data ke k .

X_{kj} = Titik pusat ke j pada tribut ke k .

- Kembali ke step 3, apabila masih ada data yang berpindah *cluster* atau apabila perubahan nilai *centroid*.

2.3. Kemiskinan



Kemiskinan merupakan masalah yang cukup serius, terutama untuk negara berkembang seperti Indonesia. Kemiskinan dapat berdampak pada terhambatnya pembangunan nasional. Kondisi dimana kehidupan yang serba kekurangan, tidak mampu memenuhi kebutuhan yang layak bagi kehidupannya. Beberapa aspek menyangkut dalam dimensi kemiskinan yaitu aspek ekonomi, aspek politik dan aspek sosial-psikologis. Secara ekonomi kemiskinan didefinisikan sebagai kekurangan sumber daya yang digunakan untuk memenuhi kebutuhan hidup dan meningkatkan kesejahteraan [2].

3. ANALISA DAN PEMBAHASAN

Sumber data penelitian diperoleh dari Badan Pusat Statistik (BPS) dengan situs <https://www.bps.go.id>. Data akan diolah dengan melakukan *clustering* presentase kemiskinan menurut provinsi dalam 2 *cluster* yaitu *cluster* tingkat tinggi dan *cluster* tingkat rendah. Data yang diperoleh untuk penelitian ini adalah data kemiskinan di setiap provinsi selama 6 tahun, yakni dari tahun 2012-2018 dapat dilihat pada tabel 1.

Tabel 1. Data Penduduk Miskin Menurut Provinsi (Persen)

Provinsi	2012	2013	2014	2015	2016	2017	2018
Aceh	28,75	26,46	26,54	25,635	24,945	24,85	23,81
Sumatera Utara	15,875	15,255	14,305	15,925	15,485	14,86	13,69
Sumatera Barat	12,19	11,92	10,855	10,665	10,66	10,245	9,925
Riau	12,245	11,93	12,115	12,83	11,815	11,485	10,995
Jambi	12,56	12,28	12,115	13,42	12,325	12,14	11,845
Sumatera Selatan	20,52	21,27	20,72	21,135	20,235	19,74	19,21
Bengkulu	26,45	27,215	26,025	26,46	25,835	24,245	23,135
Lampung	24,005	22,055	21,385	21,115	21,22	20,21	19,645
Kep.Bangka Belitung	8,215	7,835	7,845	7,81	7,74	7,85	7,635
Kep.Riau	10,525	9,635	9,9	9,13	8,9	9,065	9,115
DKI Jakarta	5,54	5,41	5,965	5,735	5,625	5,66	5,345
Jawa Barat	15,035	14,325	14,03	14,315	13,335	12,625	11,075
Jawa Tengah	22,83	21,78	21,25	20,24	19,865	19,125	16,915
DI Yogyakarta	23,98	22,945	22,275	21,49	19,895	19,2	18,035
Jawa Timur	19,94	18,915	18,56	18,48	17,975	17,37	16,405
Banten	8,705	8,685	8,105	8,775	8,1	8,245	7,865
Bali	6,155	6,195	6,91	7,365	6,325	6,32	5,965
NTB	27,64	26,595	25,775	25,37	24,49	23,595	23,565
NTT	31,085	30,15	29,62	33,9	33,195	32,54	31,865
Kalimantan Barat	12,15	12,61	12,575	12,25	11,87	11,81	11,42
Kalimantan Tengah	9,605	9,045	9,065	8,895	8,34	7	7,72
Kalimantan Selatan	7,565	7,15	7,085	7,35	7,11	7,08	6,865
Kalimantan Timur	9,87	9,41	9,575	9,28	9,11	9,23	9,06
Kalimantan Utara	0	0	0	9,04	9,72	10,7	10,52
Sulawesi Utara	12	12,13	12,88	13,14	12,44	12,05	11,595
Sulawesi Tengah	22,87	21,83	20,14	21,695	21,49	21,25	20,855
Sulawesi Selatan	15,02	14,7	15,05	14,45	14,02	14,12	13,495
Sulawesi Tenggara	20,24	19,695	20,435	19,77	19,265	18,795	17,29
Gorontalo	25,94	26,515	26,145	27,4	18,535	26,22	24,725
Sulawesi Barat	19,745	18,415	18,295	18,35	17,335	16,89	16,85
Maluku	32,16	29,125	28,35	29,19	28,81	27,595	27,045
Maluku Utara	12,81	11,32	11,005	9,95	9,535	9,57	9,95
Papua Barat	41,72	40,24	40,26	38,685	37,87	36,66	34,34
Papua	46,44	47,065	43,95	42,37	42,78	41,5	41,455

3.1 Pengolahan Data

Untuk mendapatkan hasil dari penelitian yang dilakukan, berikut uraian perhitungan manual proses *clustering* tingkat kemiskinan menggunakan algoritma *k-means*. Berikut adalah cara-cara perhitungan *k-means*.

a. Menentukan Data Yang Akan Diolah

Data yang digunakan dalam proses *clustering* adalah data presentase penduduk miskin menurut provinsi pada tahun 2012-2018 (7 tahun) dengan jumlah data sebanyak 34 provinsi. Untuk mencari rata-rata adalah dengan menjumlahkan data setiap tahun dan di bagi dengan jumlah banyaknya tahun. Berikut cara untuk mencari nilai rata-rata.

$$\begin{aligned} R_{aceh} &= 28,75 + 26,46 + 25,635 + 24,945 + 24,85 + 23,81/7 \\ &= 25,74166667 \end{aligned}$$



$$\begin{aligned}
 R_{sumut} &= 15,875 + 15,255 + 14,305 + 15,925 + 15,485 + 14,86 + 13,69/7 \\
 &= 15,05642857 \\
 R_{papua} &= 46,44 + 47,065 + 43,95 + 42,37 + 42,78 + 41,5 + 41,455/7 \\
 &= 43,65142857
 \end{aligned}$$

Hasil dari keseluruhan dapat dilihat pada tabel 2 sebagai berikut:

Tabel 2. Nilai Rata-Rata

Provinsi	Rata-Rata	Provinsi	Rata-Rata
Aceh	25,74166667	Nusa Tenggara Barat	25,29
Sumatera Utara	15,05642857	Nusa Tenggara Timur	31,765
Sumatera Barat	10,92285714	Kalimantan Barat	12,09785714
Riau	11,91642857	Kalimantan Tengah	8,524285714
Jambi	12,38357143	Kalimantan Selatan	7,172142857
Sumatera Selatan	20,40428571	Kalimantan Timur	9,362142857
Bengkulu	25,62357143	Kalimantan Utara	5,711428571
Lampung	21,37642857	Sulawesi Utara	12,31928571
Kep.Bangka Belitung	7,847857143	Sulawesi Tengah	21,44785714
Kep.Riau	9,467142857	Sulawesi Selatan	14,40785714
DKI Jakarta	5,611428571	Sulawesi Tenggara	19,35571429
Jawa Barat	13,53428571	Gorontalo	25,06857143
Jawa Tengah	20,28642857	Sulawesi Barat	17,98285714
DI Yogyakarta	21,11642857	Maluku	28,89642857
Jawa Timur	18,235	Maluku Utara	10,59142857
Banten	8,354285714	Papua Barat	38,53928571
Bali	6,462142857	Papua	43,65142857

Tabel 3. Parameter

Parameter	Nilai	Keterangan
Cluster	2	(Max/Min)
Data	34	(Provinsi)
Atribut	7	(Tahun)

b. Menentukan Jumlah *Cluster*

Jumlah *Cluster* yang digunakan pada data penduduk miskin sebanyak 2 *cluster*. *Cluster* tersebut diantaranya (C1) Tinggi dan (C2) Rendah.

c. Menentukan *Centroid*

Penentuan pusat awal *cluster* (*centroid*) ditentukan secara manual atau random yang diambil dari data yang ada dalam *range*. Nilai *cluster* 0 diambil dari data paling tinggi dan nilai *cluster* 1 diambil dari data paling rendah.

Cara mencari nilai *centroid* data awal untuk iterasi ke-1 adalah :

$$\begin{aligned}
 C1 = & 25,74166667/15,05642857/10,92285714/11,91642857/12,38357143/20,40428571/25,62357143/21,3764 \\
 & 2857/7,847857143/9,467142857/5,611428571/13,53428571/20,28642857/21,11642857/18,235/8,354285 \\
 & 714/6,462142857/25,29/31,765/12,09785714/8,524285714/7,172142857/9,362142857/5,711428571/12,3 \\
 & 1928571/21,44785714/14,40785714/19,35571429/25,06857143/17,98285714/28,89642857/10,59142857 \\
 & 38,53928571/43,65142857 = 43,65142857
 \end{aligned}$$

$$\begin{aligned}
 C2 = & 25,74166667/15,05642857/10,92285714/11,91642857/12,38357143/20,40428571/25,62357143/21,3764 \\
 & 2857/7,847857143/9,467142857/5,611428571/13,53428571/20,28642857/21,11642857/18,235/8,354285 \\
 & 714/6,462142857/25,29/31,765/12,09785714/8,524285714/7,172142857/9,362142857/5,711428571/12,3 \\
 & 1928571/21,44785714/14,40785714/19,35571429/25,06857143/17,98285714/28,89642857/10,59142857 \\
 & 38,53928571/43,65142857 = 5,61142857
 \end{aligned}$$

Tabel 4. Data *Centroid*

Cluster	Nilai
Centroid 1	43,6514286
Centroid 2	5,61142857

d. Menghitung Jarak dari *Centroid*

Untuk menghitung jarak antara titik *centroid* dengan titik tiap objek menggunakan *Euclidian Distance*. Maka perhitungan untuk menghitung jarak dari *centroid* adalah sebagai berikut :

$$D(aceh, c1) = \sqrt{(43,65142857 - 25,74166667)^2} = 17,9097619$$

$$D(aceh, c2) = \sqrt{(25,74166667 - 5,61142857)^2} = 20,130238$$

$$D(sumut, c1) = \sqrt{(43,65142857 - 15,05642857)^2} = 28,59500003$$



$$D(sumut, c2) = \sqrt{(15,05642857 - 5,61142857)^2} = 9,445$$

⋮

$$D(papua, c1) = \sqrt{(43,65142857 - 43,65142857)^2} = 0$$

$$D(papua, c2) = \sqrt{(43,65142857 - 5,61142857)^2} = 38,04$$

Hasil dari keseluruhan dapat dilihat pada tabel 5 sebagai berikut :

Tabel 5. Perhitungan Jarak Data Iterasi ke-1

No	Provinsi	Rata-Rata	C1	C2	Terdekat
1	Aceh	25,74166667	17,90976193	20,130238	17,90976193
2	Sumatera Utara	15,05642857	28,59500003	9,445	9,445
3	Sumatera Barat	10,92285714	32,72857146	5,3114286	5,3114286
4	Riau	11,91642857	31,73500003	6,305	6,305
5	Jambi	12,38357143	31,26785717	6,7721429	6,7721429
6	Sumatera Selatan	20,40428571	23,24714289	14,792857	14,792857
7	Bengkulu	25,62357143	18,02785717	20,012143	18,02785717
8	Lampung	21,37642857	22,27500003	15,765	15,765
9	Kep.Bangka Belitung	7,847857143	35,80357146	2,2364286	2,2364286
10	Kep.Riau	9,467142857	34,18428574	3,8557143	3,8557143
11	DKI Jakarta	5,611428571	38,04000003	0	0
12	Jawa Barat	13,53428571	30,11714289	7,9228571	7,9228571
13	Jawa Tengah	20,28642857	23,36500003	14,675	14,675
14	DI Yogyakarta	21,11642857	22,53500003	15,505	15,505
15	Jawa Timur	18,235	25,4164286	12,623571	12,623571
...	Banten	8,354285714	35,29714289	2,7428571	2,7428571
...
31
32	Maluku	28,89642857	14,75500003	23,285	14,75500003
33	Maluku Utara	10,59142857	33,06000003	4,98	4,98
34	Papua Barat	38,53928571	5,11214289	32,927857	5,11214289
	Papua	43,65142857	0	38,04	0

e. Menentukan *Cluster* atau Pengelompokan

Dalam menentukan *cluster* dengan mencari nilai *cluster* berdasarkan nilai minimal dari nilai *cluster* dan diletakkan pada *cluster* yang sesuai dengan nilai minimal pada iterasi 1. Berikut tabel *cluster* pada iterasi 1 sebagai berikut:

Tabel 6. Cluster Iterasi ke-1

C1	C2	C1	C2
1		1	
	1		1
	1		1
	1		1
	1		1
1			1
	1		1
	1		1
	1		1
	1		1
	1	1	
	1		1
	1	1	
	1		1
	1		1
	1		1

Tabel 7. Hasil Iterasi Ke-1

Cluster	Provinsi	Hasil
C1	(1,7,18,19,29,31,33,34)	8

Cluster	Provinsi	Hasil
C2	(2,3,4,5,6,8,9,10,11,12,13,14,15,16,17,20,21,22,23,24,25,26,27,28,30,32)	26

Selanjutnya dalam metode *k-means*, perhitungan berhenti apabila *cluster* pada iterasi yang dihasilkan sama pada iterasi sebelumnya. Maka selanjutnya mencari *cluster* pada iterasi selanjutnya sampai nilai iterasinya sama. Untuk mencari nilai *centroid* selanjutnya dengan menggunakan *centroid* baru pada iterasi ke-1 dengan menjumlahkan nilai sesuai yang tertera pada *cluster* tabel di atas. Adapun *centroid* baru untuk mencari *cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *cluster* tersebut kemudian membagikannya sebanyak jumlah nilai sebagai berikut :

$$C1x = \frac{25,74166667 + 25,62357143 + 25,29 + 31,765 + 25,06857143 + 28,89642857 + 38,53928571 + 43,65142857}{8} = 30,57199405$$

$$C2x = \frac{15,05642857 + 10,92285714 + 11,91642857 + 12,38357143 + 20,40428571 + 21,37642857 + 7,847857143 + 9,467142857 + 5,611428571 + 13,5342857 + 20,286428571 + 20,286428571 + 21,11642857 + 18,235 + 8,35428571 + 7,172142857 + 9,362142857 + 5,711428571 + 12,31928571 + 21,44785714 + 14,40785714 + 19,35571429 + 17,98285714 + 10,59142857}{26} = 13,15184066$$

Maka didapat data *centroid* baru iterasi ke-2 adalah sebagai berikut :

Tabel 8. *Centroid* Baru Iterasi Ke-2

C1	C2
<i>Centroid 1</i>	30,57199405
<i>Centroid 2</i>	13,15184066

Dengan menggunakan langkah-langkah yang sama seperti sebelumnya untuk menentukan jarak dari *centroid* pada iterasi ke-2 dengan menggunakan *centroid* baru. Berikut perhitungannya.

$$D(aceh, c1) = \sqrt{(30,57199405 - 25,74166667)^2} = 4,83032738$$

$$D(aceh, c2) = \sqrt{(25,74166667 - 13,15184066)^2} = 12,58982601$$

$$D(sumut, c1) = \sqrt{(30,57199405 - 15,05642857)^2} = 15,51556548$$

$$D(sumut, c2) = \sqrt{(15,05642857 - 13,15184066)^2} = 1,90458791$$

⋮

$$D(papua, c1) = \sqrt{(30,57199405 - 43,65142857)^2} = 13,07943452$$

$$D(papua, c2) = \sqrt{(43,65142857 - 13,15184066)^2} = 30,49958791$$

Maka didapatkan hasil keseluruhan dari iterasi ke-2 dapat dilihat pada tabel 9. sebagai berikut :

Tabel 9. Perhitungan Jarak Data Iterasi Ke-2

No	Provinsi	Rata-Rata	C1	C2	Terdekat
1	Aceh	25,74166667	4,83032738	12,58982601	4,83032738
2	Sumatera Utara	15,05642857	15,51556548	1,90458791	1,90458791
3	Sumatera Barat	10,92285714	19,64913691	2,22898352	2,22898352
4	Riau	11,91642857	18,65556548	1,23541209	1,23541209
5	Jambi	12,38357143	18,18842262	0,76826923	0,76826923
6	Sumatera Selatan	20,40428571	10,16770834	7,25244505	7,25244505
7	Bengkulu	25,62357143	4,94842262	12,47173077	4,94842262
8	Lampung	21,37642857	9,19556548	8,22458791	8,22458791
9	Kep.Bangka Belitung	7,847857143	22,72413691	5,303983517	5,303983517
10	Kep.Riau	9,467142857	21,10485119	3,684697803	3,684697803
11	DKI Jakarta	5,611428571	24,96056548	7,540412089	7,540412089
12	Jawa Barat	13,53428571	17,03770834	0,38244505	0,38244505
13	Jawa Tengah	20,28642857	10,28556548	7,13458791	7,13458791
14	DI Yogyakarta	21,11642857	9,45556548	7,96458791	7,96458791
15	Jawa Timur	18,235	12,33699405	5,08315934	5,08315934
...
...
31	Maluku	28,89642857	1,67556548	15,74458791	1,67556548
32	Maluku Utara	10,59142857	19,98056548	2,56041209	2,56041209
33	Papua Barat	38,53928571	7,96729166	25,38744505	7,96729166
34	Papua	43,65142857	13,07943452	30,49958791	13,07943452

Dari tabel jarak *centroid* diatas, maka *cluster* atau pengelompokan iterasi ke-2 adalah sebagai berikut :

Tabel 10. Cluster Iterasi Ke-2

C1	C2		C1	C2
1			1	
	1		1	
	1			1
	1			1
	1			1
	1			1
	1			1
1			1	
	1		1	
	1			1
	1			1
	1			1
	1		1	
	1			1
	1			1
	1		1	
	1			1

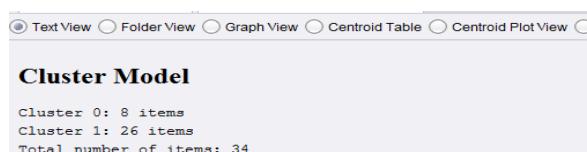
Tabel 11. Hasil Iterasi Ke-2

<i>Cluster</i>	Provinsi	Hasil
C1	(1,7,18,19,29,31,33,34)	8
C2	(2,3,4,5,6,8,9,10,11,12,13,14,15,16,17,20,21,22,23,24,25,26,27,28,30,32)	26

Perhitungan manual pada data penduduk miskin diatas didapatkan hasil akhir yang dimana pada iterasi ke 2 pengelompokan data yang dilakukan didapatkan hasil yang sama. Hasil dari kedua iterasi tersebut bernilai $C1 = 8$ dan $C2 = 26$. Sehingga posisi *cluster* pada data tersebut tidak mengalami perubahan lagi maka proses iterasi berhenti.

3.2 Pengujian Data

Untuk mendapatkan hasil pengelompokan maka pada tahap selanjutnya menampilkan hasil akhir dalam menggunakan *tools rapidminer*. Dapat dilihat pada gambar 2. berikut :

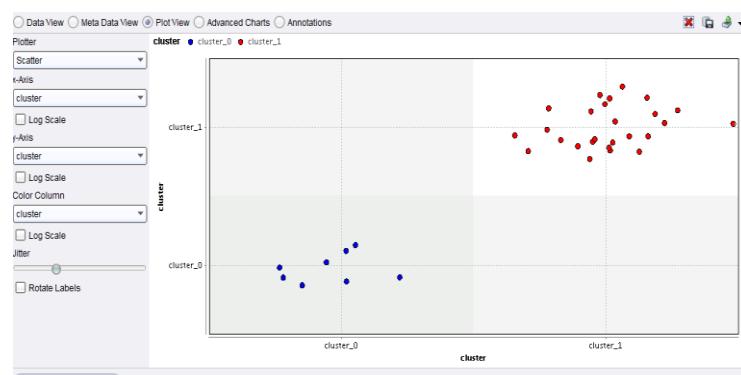


Gambar 2. Nilai Cluster Model

Keterangan :

- a. Jumlah *cluster* 0 (Tinggi) berjumlah 8 *items*
 b. Jumlah *cluster* 1 (Rendah) berjumlah 26 *items*

Sehingga dapat diketahui hasil pengelompokan dari **rapidminer** berikut ini dapat dilihat pada gambar 4.1



Gambar 3. Hasil Pengelompokan



4. KESIMPULAN

Berdasarkan pembahasan pada bab sebelumnya dapat disimpulkan bahwa penerapan *data mining* dengan algoritma *k-means* dapat diterapkan. Sumber data yang digunakan pada penelitian ini adalah www.bps.go.id dengan subjek data penduduk miskin berdasarkan provinsi (2012-2018). Jumlah *record* yang digunakan sebanyak 34 provinsi dengan menghasilkan 2 *cluster* yakni *cluster* tinggi sebanyak 8 provinsi dan *cluster* rendah sebanyak 26 provinsi. Berdasarkan hasil pengujian *k-means* untuk kasus presentase data penduduk miskin menggunakan *tools* *RapidMiner* versi 5.3 diperoleh hasil yang sama dengan analisis perhitungan algoritma dimana diperoleh 8 provinsi dengan *cluster* tinggi yang menjadi pusat perhatian bagi pemerintah dalam melakukan sosialisasi dan pemetaan dalam pemberian bantuan pada provinsi tersebut.

REFERENCES

- [1] N. Nurwati, "Kemiskinan : Model Pengukuran , Permasalahan dan Alternatif Kebijakan," vol. 10, no. 1, pp. 1–11, 2008.
- [2] N. I. Febianto and N. D. Palasara, "Analisis Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018," vol. 08, no. September, pp. 130–140, 2019.
- [3] W. Nengsih, "Descriptive Modelling Menggunakan K-Means Untuk Pengclusteran Tingkat Kemiskinan Di Propinsi Riau," no. January, 2017.
- [4] 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.
- [5] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Eeccis*, vol. 7, no. 1, pp. 59–64, 2013.
- [6] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "MEMANFAATKAN ALGORITMA K-MEANS DALAM MENENTUKAN PEGAWAI YANG LAYAK MENGIKUTI ASESSMENT CENTER UNTUK CLUSTERING PROGRAM SDP," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 3, no. 1, pp. 87–93, 2018.
- [7] R. W. Sari and D. Hartama, "Data Mining : Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi," *Semin. Nas. Sains Teknol. Inf.*, pp. 322–326, 2018.
- [8] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.