

Penerapan Metode K-Means Clustering Untuk Daging Ayam Buras

Viya Miralda¹, Muhammad Zarlis², Eka Irawan¹

¹ Prodi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

² Universitas Sumatera Utara, Medan, Indonesia

Email: ¹Viyamiralda12@gmail.com, ²m.zarlis@yahoo.com, ³eka.irawan@amiktunasbangsa.ac.id

Abstrak

Ayam buras merupakan unggas lokal negara Indonesia yang populasinya banyak di temukan di setiap wilayah Indonesia. Namun tingkat konsumsi daging ayam buras di Provinsi Sumatera Utara saat ini masih tergolong rendah. Data penelitian ini meliputi data produksi daging ayam buras Provinsi Sumatera Utara. Salah satu teknik data mining yaitu clustering. Clustering sendiri merupakan suatu metode dengan melakukan pengelompokan data. Data penelitian ini meliputi data produksi daging ayam buras Provinsi Sumatera Utara yang dimulai sejak tahun 2010, 2011, 2012, 2013, 2014, 2016, 2017. (7 tahun), dan data tersebut diperoleh dari Direktorat Jenderal Peternakan dan Kesehatan Hewan dan di akses melalui situs BPS Sumatera Utara (Badan Pusat Statistik). Tujuan penelitian ini adalah mengelompokkan Kabupaten/Kota di Provinsi Sumatera Utara menjadi 2 bagian. Yaitu kelompok Kabupaten/Kota dengan Daging ayam buras tinggi dan Kabupaten/Kota dengan Daging ayam buras rendah. Hal ini dapat menjadi rekomendasi kepada pemerintah Provinsi Sumatera Utara, Kabupaten/Kota yang menjadi perhatian lebih pada populasi daging ayam buras berdasarkan cluster yang telah dilakukan. Hasil dari penelitian ini menyimpulkan dari 33 data daging ayam buras di kabupaten/kota di provinsi Sumatera Utara dapat di ketahui, 3 kabupaten cluster tingkat tinggi yakni : Mandailing Natal, Langkat, Serdang Bedagai dan 30 kabupaten/kota lainnya termasuk kedalam cluster tingkat rendah.

Kata Kunci: Data Mining; Clustering; K-Means; Populasi Ayam Buras; Provinsi Sumatera Utara; Indonesia.

Abstract

Free-range chicken is a local Indonesian poultry whose population is widely found in every region of Indonesia. However, the level of consumption of free-range chicken meat in North Sumatra Province is still relatively low. This research data contains data on the production of free-range chicken in North Sumatra Province. One data mining technique is clustering. Grouping itself is a method by grouping data. The data of this study contains data on the production of free-range chicken in North Sumatra Province which began in 2010, 2011, 2012, 2013, 2014, 2017. (7 years), and data obtained from the Directorate General of Animal Husbandry and Animal Health as well as on animal access to BPS Sumatra site North (Central Statistics Agency). The purpose of this study is to group Regencies / Cities in North Sumatra Province into 2 parts. These are the Regency / City group with high free-range chicken meat and Regency / City with low-free chicken meat. This can be a contribution of the provincial government of North Sumatra, Regency / City which is of more concern than free-range chicken based on the cluster that has been done. The results of this research are taken from 33 data of free-range chicken meat in regencies / cities in North Sumatra province, 3 high-level districts, namely: Mandailing Natal, Langkat, Serdang Bedagai and 30 other regencies / cities including

Keywords: Data Mining; Clustering, K-Means; Population of Free-range Chicken; North Sumatra Province; Indonesia.

1. PENDAHULUAN

Algoritma K-Means merupakan algoritma klusterisasi yang mengelompokkan data berdasarkan titik pusat kluster (*centroid*) terdekat dengan data. Tujuan dari K-Means adalah mengelompokkan data dengan memaksimalkan kemiripan data dalam satu kluster dan meminimalkan kemiripan data antar kluster. Ukuran kemiripan yang digunakan dalam kluster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap titik centroid [1]. *Data mining* atau sering disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran *data mining* ini bisa dipakai untuk membantu pengambilan keputusan di masa depan (Mujiasih, 2015). Analisis kluster merupakan suatu metode statistik yang digunakan untuk mengelompokkan sekumpulan objek ke dalam kelompok-kelompok berdasarkan karakteristik yang dimilikinya. Objek diklasifikasikan ke dalam satu atau lebih kluster sehingga objek-objek yang berada di dalam kluster mempunyai kemiripan atau kesamaan karakter [2].

Kebutuhan akan protein hewani sangat di perlukan manusia dalam menjalani aktivitas sehari hari, protein dapat diperoleh dari protein hewani seperti ayam pedaging (ayam putih) dan juga ayam buras. Di Indonesia terdapat dua jenis ayam, yaitu ayam ras dan ayam buras (bukan ras). Ayam ras terdiri dari dua jenis yaitu ayam ras pedaging dan ayam ras petelur, sedangkan ayam buras adalah sebutan bagi semua jenis ayam yang bukan golongan ayam ras. Ayam buras atau sering di sebut ayam lokal yang banyak di temukan di seluruh daerah di Indonesia, ayam buras populer di sebut dengan ayam kampung, padahal ayam buras bukan hanya ayam kampung saja, mencakup juga jenis ayam seperti ayam arab, ayam kedu, ayam pelung, ayam kate, dan jenis ayam lainnya. Kualitas ayam buras lebih unggul dibandingkan dengan ayam ras, ayam buras lebih unggul dari segi kualitas daging, protein, harga jual dan ayam buras lebih memiliki sedikit lemak di bandingkan dengan ayam ras, Dalam pemeliharaannya, ayam ras pedaging di beri antibiotik dan makanannya menggunakan pelet, yang jelas tidak

baik jika dagingnya sering di konsumsi secara terus menerus, berbeda dengan ayam buras yang tidak menerima antibiotik dan ayam buras tidak makan pelet makanan ayam buras juga lebih sehat karena di pelihara dengan sistem umbar sehingga ayam buras lebih baik untuk kesehatan. Seiring dengan meningkatnya pengetahuan masyarakat dalam memilih makanan sumber protein yang baik dan sehat untuk tubuh, maka semakin tinggi lah minat masyarakat terhadap daging ayam buras, tetapi tingginya permintaan masyarakat akan ayam buras tidak di barengi dengan produksinya, produksi ayam buras masih lebih rendah jika dibandingkan dengan ayam ras, sehingga masyarakat kesulitan menemukan daging ayam ras di pasaran.

Dalam hal itu, penulis membuat suatu penelitian dengan data mining metode *clustering* untuk mengelompokkan Kabupaten/Kota di Provinsi Sumatera Utara menjadi dua bagian, yang nantinya akan menghasilkan 2 kategori daerah dengan Daging ayam buras tinggi dan kategori daerah dengan Daging ayam buras rendah. Maka daerah yang termasuk dalam Daging ayam buras rendah akan lebih di tingkatkan lagi produksinya sehingga di harapkan akan bisa memenuhi permintaan akan daging ayam buras di pasar pasar yang ada di daerah kabupaten/kota di provinsi sumatera utara dan dapat memperkecil jumlah impor daging dan menambah hasil produksi ayam buras di Provinsi Sumatera Utara, agar produksi ayam buras bisa memenuhi kebutuhan protein nasional khususnya di Provinsi Sumatera Utara.

Kajian Peneliti (Syam, 2017) yang berjudul "Implementasi Metode Klastering K-Means Untuk Mengelompokkan Hasil Evaluasi Mahasiswa" menyimpulkan bahwa, metode *Clustering* Algoritma *K-Means* dapat diterapkan pada pengelompokan hasil evaluasi mahasiswa FKIP Universitas Riau, sehingga metode ini sangat membantu pihak akademik dalam menentukan mahasiswa berprestasi, berpotensi berprestasi, berpotensi bermasalah dan bermasalah. Begitu juga dengan penelitian [3] yang berjudul "Penerapan Metode K-Means dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018" menyimpulkan bahwa, dengan menggunakan metode *K-Means*, data jumlah kejadian bencana alam periode waktu 5 tahun terakhir (2013-2018) di Indonesia dapat dikelompokkan menjadi 2 klaster. Dimana klaster pertama yaitu daerah yang rawan terjadinya bencana alam terdiri dari 3 provinsi, di antaranya Provinsi Jawa Tengah, Jawa Timur, dan Jawa Barat. Sedangkan klaster 2 atau daerah yang tidak rawan terjadinya bencana alam terdiri dari 31 provinsi lainnya.

Diharapkan penelitian ini dapat memberikan manfaat dan rekomendasi bagi pemerintah yg khususnya Kementerian Pertanian dan Direktorat Jendral Perternakan yang menangani persoalan Daging Ayam Buras di Provinsi Sumatera Utara agar dapat meningkatkan hasil produksi Ayam Buras di Provinsi Sumatera Utara sehingga impor daging juga dapat di kurangi.

2. METODE PENELITIAN

2.1 Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* secara keseluruhan [4].

2.2 K-Means

K-means merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok [5]. Langkah-langkah melakukan clustering dengan metode K-means adalah sebagai berikut : 1) Pilih jumlah cluster k; 2) Inisialisasi ke pusat cluster ini bisa dilakukan dengan berbagai cara. Cara yang paling sering dilakukan adalah dengan random atau acak. Pusat-pusat cluster diberiduberi nilai awal dengan angka-angka random; 3) Alokasikan semua data/objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut (Bastian et al., 2018).

Dalam penelitian ini terdapat prosedur pembentukan *K-Means Clustering* terdapat langkah-langkah yang dapat dilakukan, antara lain:

- Tentukan *k* sebagai jumlah *cluster* yang ingin dibentuk.
- Bangkitkan *K* centroid (titik pusat klaster) awal secara random kemudian untuk menghitung centroid klaster ke-*i* berikutnya, digunakan rumus sebagai berikut:

$$y_i = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Keterangan:

y = centroid pada cluster

x_i = obyek pengamatan ke *i*



- n = banyaknya obyek yang menjadi anggota klaster.
 c. Hitung jarak setiap data ke masing-masing *centroid*.

$$d = \sum_{i=1}^n (x_i - y_i)^2 \tag{2}$$

Keterangan:

x_i = obyek pengamatan ke i

y_i = centroid ke i

n = banyaknya obyek yang menjadi anggota klaster

d. Setiap data memilih *centroid* yang terdekat.

Tentukan posisi *centroid* yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada *centroid* yang sama.

- e. Kembali ke langkah-3 jika posisi *centroid* baru dengan *centroid* yang lama tidak sama.

2.3 Clustering

Clustering merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data clustering dan *non-hierarchical* (non hirarki) data clustering [6]. *Clustering* dapat juga dikatakan sebagai identifikasi kelas objek yang memiliki kemiripan. Dengan menggunakan teknik *clustering* kita bisa lebih lanjut mengidentifikasi kepadatan dan jarak daerah dalam objek ruang dan dapat menemukan secara menyeluruh pola distribusi dan korelasi antara atribut.

3. ANALISA DAN PEMBAHASAN

Sumber data penelitian diperoleh dari Badan Pusat Statistik (BPS) dengan situs <https://www.bps.go.id>. Data yang digunakan dalam penelitian ini adalah data presentase kemiskinan menurut provinsi dari tahun 2012-2018 yang terdiri dari 34 provinsi. Data akan diolah dengan melakukan *clustering* presentase kemiskinan menurut provinsi dalam 2 *cluster* yaitu *cluster* tingkat tinggi dan *cluster* tingkat rendah.

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



Provinsi	2012	2013	2014	2015	2016	2017	2018
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*. Proses *clustering* dilakukan mulai dari penentuan data yang ingin di *cluster*. Dalam hal ini variabel data yang ingin di *cluster* adalah data dari tahun 2012-2018 dan disini penulis mengambil nilai rata-rata dari tahun 2012-2018 sebagai data yang akan di *cluster*. 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.

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

$$R_{sumut} = 15,875 + 15,255 + 14,305 + 15,925 + 15,485 + 14,86 + 13,69/7 = 15,05642857$$

$$R_{papua} = 46,44 + 47,065 + 43,95 + 42,37 + 42,78 + 41,5 + 41,455/7 = 43,65142857$$

Hasil dari keseluruhan dapat dilihat pada tabel 2. sebagai berikut :

Tabel 2. Nilai Rata-Rata

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

$$C1 = \frac{25,74166667/15,05642857/10,92285714/11,91642857/12,38357143/20,40428571/25,62357143/21,37642857/7,847857143/9,467142857/5,61142857/13,53428571/20,28642857/21,11642857/18,235/8,354285714/6,462142857/25,29/31,765/12,09785714/8,524285714/7,172142857/9,362142857/5,711428571/12,31928571/21,44785714/14,40785714/19,35571429/25,06857143/17,98285714/28,89642857/10,59142857/38,53928571/43,65142857}{43,65142857}$$

$$C2 = \frac{25,74166667/15,05642857/10,92285714/11,91642857/12,38357143/20,40428571/25,62357143/21,37642857/7,847857143/9,467142857/5,61142857/13,53428571/20,28642857/21,11642857/18,235/8,354285714/6,462142857/25,29/31,765/12,09785714/8,524285714/7,172142857/9,362142857/5,711428571/12,31928571/21,44785714/14,40785714/19,35571429/25,06857143/17,98285714/28,89642857/10,59142857/38,53928571/43,65142857}{5,61142857}$$

Tabel 4. Data Centroid

<i>Cluster</i>	Nilai
<i>Centroid</i> 1	43,6514286
<i>Centroid</i> 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
...
...
...
32	Maluku Utara	10,59142857	33,06000003	4,98	4,98
33	Papua Barat	38,53928571	5,11214289	32,927857	5,11214289
34	Papua	43,65142857	0	38,04	0

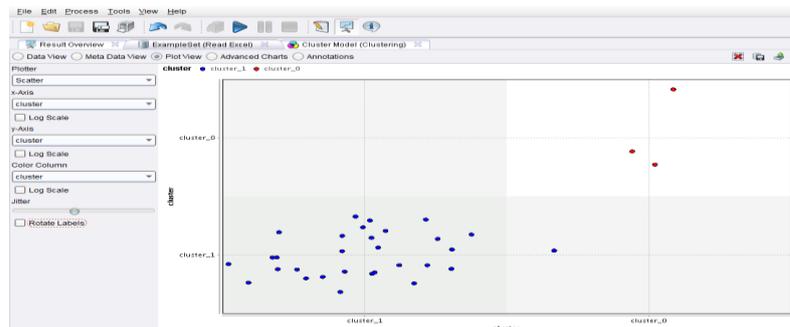
e. Menentukan *Cluster* atau Pengelompokan

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 :

Keterangan :

- a. Jumlah *Cluster* 0 (Rendah) berjumlah 30 *items*
- b. Jumlah *Cluster* 1 (Tinggi) berjumlah 3 *items*
- c. Jumlah keseluruhan *items* adalah 33

Sehingga dapat diketahui hasil pengelompokan dari rapidminer 5.3 berikut ini dapat dilihat pada gambar 2.



Gambar 2. Hasil Pengelompokan

Berdasarkan pada gambar 2 dapat diketahui bahwa pada kelompok tinggi memiliki banyak node yaitu 3, sedangkan pada kelompok rendah memiliki 30 node.

4. KESIMPULAN

Berdasarkan pembahasan pada bab sebelumnya dapat disimpulkan 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] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015, doi: 10.1038/hdy.2009.180.
- [2] M. A. M. Fitra Ramdhani, Abdul Hoyyi, "PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN KARAKTERISTIK KESEJAHTERAAN RAKYAT MENGGUNAKAN METODE K-MEANS CLUSTER," *J. GAUSSIAN*, vol. 4, no. 2006, pp. 875–884, 2015.
- [3] A. R. Mira Suci Yana, Lathifah Setiawan, Elvitra Mutia Ulfa, "Penerapan Metode K-Means dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018," *J. Data Anal.*, vol. Vol.1, No., no. December 2018, 2019, doi: 10.24815/jda.v1i2.12584.
- [4] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform. Penelit. Bid. Komput. Sains dan Pendidik. Inform.*, vol. V2.i2(213-, no. ISSN : 2407-0491, 2017.
- [5] Linda Maulida, "PENERAPAN DATAMINING DALAM MENGELOMPOKKAN KUNJUNGAN WISATAWAN KE OBJEK WISATA UNGGULAN DI PROV. DKI JAKARTA DENGAN K-MEANS," *JISKa (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, pp. 167–174, 2018.
- [6] S. Rosmini, Abdul Fadlil, "Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," *IT J. Res. Dev.*, vol. Vol.3, No., no. 1, pp. 22–31, 2018.