

Implementasi *Data Mining* Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan *Algoritma K-Means*

Yuni Radana Sembiring¹, Saifullah², Riki Winanjaya³

^{1,2,3} STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia
Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara

¹yunioradanasembiring@gmail.com, ²saifullah@amiktunasbangsa.ac.id,

³riki@amiktunasbangsa.ac.id

Abstract

Poverty is a certain condition that is below the standard line of minimum needs, good for food and non-food. Poor households generally have a greater average number of members compared to households that only have members who have fewer members. This situation is followed by the low level of education of household heads and workers who generally only work in the agricultural sector. Factors such as education, labor, health, fertility, housing, and the environment are a picture of the level of people's welfare which is thought to affect the amount of poverty. This study used data sourced from the Central Bureau of statistics the year 2007-2019. The method used is *Datamining the K-Means Clustering*, *Clustering* is a method used in *datamining* the how it works find and classify data that has a semblance and characteristics of data between one another with the data. Using this algorithm the data already obtained can be grouped into Clusters based on this data. This data can be entered to the local Government to recommend to the Government so that the Government can handle the number of poor people in this country.

Keywords: The Number Of Poor People, Data Mining, K-Means

Abstrak

Kemiskinan merupakan suatu kondisi yang berada di bawah garis standar kebutuhan minimal, baik pangan maupun non pangan. Rumah tangga miskin umumnya memiliki jumlah rata-rata anggota yang lebih banyak dibandingkan dengan rumah tangga yang hanya memiliki anggota yang anggota lebih sedikit. Situasi ini diikuti dengan rendahnya tingkat pendidikan kepala rumah tangga dan pekerja yang umumnya hanya bekerja di sektor pertanian. Faktor-faktor seperti pendidikan, tenaga kerja, kesehatan, kesuburan, perumahan, dan lingkungan merupakan gambaran tingkat kesejahteraan masyarakat yang diperkirakan akan mempengaruhi jumlah kemiskinan. Penelitian ini menggunakan data yang bersumber dari Badan Pusat Statistik tahun 2007-2019. Metode yang digunakan adalah *Datamining the K-Means Clustering*, *Clustering* adalah metode yang digunakan dalam mendata cara kerjanya mencari dan mengelompokkan data yang memiliki kemiripan dan karakteristik data antara satu dengan yang lain dengan datanya. Dengan menggunakan algoritma ini, data yang sudah diperoleh dapat dikelompokkan ke dalam Cluster berdasarkan data ini. Data ini dapat dimasukkan ke Pemerintah Daerah untuk direkomendasikan kepada Pemerintah agar Pemerintah dapat menangani jumlah penduduk miskin di negara ini.

Keywords: Jumlah Penduduk Miskin, Data Mining, K-Means

1. Pendahuluan

Kemiskinan merupakan permasalahan sosial yang dihadapi hampir di setiap negara. Kemiskinan tidak hanya dihadapi oleh pemerintah pusat akan tetapi juga menjadi

permasalahan yang serius bagi pemerintah daerah. Persoalan kemiskinan yang dihadapi pemerintah berkaitan erat dengan rendahnya pendapatan sehingga tidak dapat memenuhi kebutuhan pokoknya. Rendahnya pendapatan yang diperoleh berimbas pada kurangnya kesempatan dalam mengakses pendidikan dan fasilitas pemerintah lainnya. Begitu pula dengan pendapatan daerah yang rendah menyebabkan kurang terdistribusinya pemerataan pendapatan bagi masyarakat.

Melihat kondisi penduduk miskin di Indonesia, sudah seharusnya pemerintah melakukan upaya untuk mengurangi jumlah penduduk miskin di setiap Provinsi melalui berbagai program bantuan. Untuk program ini diperlukan sebuah langkah pemetaan bagi tiap provinsi sesuai dengan tingkat garis kemiskinan sehingga dapat diambil sebuah solusi yang tepat sesuai dengan kelompoknya. Salah satu teknik yang dikenal dalam *Data Mining* yaitu teknik *clustering*. Teknik *clustering* mengelompokkan data secara otomatis tanpa perlu diberitahu label kelasnya. Banyak metode *clustering* yang telah diusulkan oleh para ahli, salah satunya adalah *K-Means*. Metode *K-Means* merupakan algoritma klasterisasi yang paling banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan mengaplikasikannya.

(Brilian Rahmat C.T.I) menerapkan metode *K-Means clustering* pada Rapidminer untuk analisis daerah rawan kecelakaan. Dari hasil penelitian ini, dapat disimpulkan bahwa analisis data kecelakaan menggunakan aplikasi RapidMiner dapat mengekstraksi beberapa informasi untuk data kecelakaan menjadi 3 buah kelompok/cluster. Peneliti lain juga menerapkan metode *k-means* dalam pengelompokan daerah terjangkit demam berdarah Dengue (DBD) berdasarkan Provinsi [1]. Data ini diolah menggunakan *Rapidminer* untuk ditentukan nilai *centroid* dalam 3 cluster yaitu cluster tingkat tinggi, cluster tingkat sedang, dan cluster tingkat rendah.

Menyikapi hal tersebut, pada penelitian ini penulis mencoba untuk menerapkan teknik *clustering* dengan menggunakan algoritma *K-Means* sebagai metode penyelesaian dalam mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi. Dari hasil yang diperoleh, penelitian ini diharapkan mampu mengidentifikasi faktor-faktor jumlah kemiskinan di setiap provinsi dan dapat memberikan informasi atau masukan bagi pemerintah dalam mengambil langkah yang tepat sebagai upaya untuk menurunkan jumlah kemiskinan pada setiap provinsi yang ada di Indonesia.

2. Metodologi Penelitian

2.1. Data mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (induction-based learning) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. Knowledge Discovery in Databases (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD[2] [3].

2.2. Clustering

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 serta tidak memerlukan target *output*. Dalam data mining ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering* [4][5].

2.3. Algoritma K-Means

K-Means merupakan Algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah K *cluster* yang sudah ditetapkan diawal. Algoritma *K-Means* sederhana untuk diimplementasikan dan dijalankan, relative cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, *K-Means* menjadi salah satu algoritma yang paling penting dalam bidang data mining. [6].

Berikut ini adalah langkah-langkah dalam algoritma *K-Means* [7]:

- a) Tentukan jumlah *cluster* (k) pada data set.
- b) Tentukan nilai pusat (*centroid*).

Penentuan nilai *centroid* pada tahap awal dilakukan secara *random*, sedangkan pada tahap iterasi digunakan rumus seperti pada persamaan (1) berikut ini :

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}^2 \quad (1)$$

Keterangan :

V_{ij} = *centroid* rata-rata *cluster* ke-I untuk variabel ke-j

N_i = Jumlah anggota *cluster* ke-i

i, k = indeks dari *cluster*

j = indeks dari variabel

X_{kj} = nilai data ke-k variabel ke-j untuk *cluster* tersebut

- c) Pada masing-masing *record*, hitung jarak terdekat dengan *centroid*. Jarak *centroid* yang digunakan adalah *Euclidean Distance*, dengan]
- d) rumus seperti pada persamaan (2) :

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2)$$

Keterangan :

De = *Euclidean Distance*

i = Banyaknya objek ²

(x, y) = Koordinat objek

(s, t) = Koordinat *centroid*

- e) Kelompokkan objek berdasarkan jarak ke *centroid* terdekat.
Ulangi langkah ke-2, lakukan *iterasi* hingga *centroid* bernilai optimal.

3. Hasil dan Pembahasan

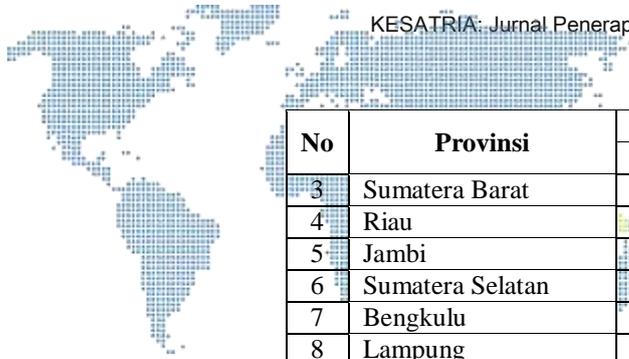
Dalam menentukan hasil dari penelitian yang dilakukan, berikut uraian perhitungan manual proses *clustering* jumlah penduduk miskin menggunakan *algoritma K-Means*. Proses *clustering* dilakukan mulai dari penentuan data yang ingin di *cluster*. Dalam hal ini variabel data yang ingin di *cluster* adalah data dari tahun 2007 sampai 2019 yang terdiri dari perkotaan dan perdesaan, disini peneliti mengambil nilai rata rata dari tahun 2007 sampai dengan 2019 sebagai data yang akan di *cluster*. Berikut adalah langkah-langkah penyelesaian yang dilakukan penulis dalam mengelompokkan data jumlah penduduk miskin menggunakan algoritma *K-Means* :

- a) Menentukan jumlah data yang akan di *cluster*, dimana sampel data jumlah penduduk miskin yang akan digunakan dalam proses *clustering* adalah Data jumlah penduduk miskin yang terdapat pada Badan Pusat Statistik pada tahun 2007-2019 yang terdiri dari Perdesaan (X) dan Perkotaan (Y) jumlah data sebanyak 34 Provinsi. Berikut adalah tabel 1 data Jumlah penduduk miskin:

Tabel 1. Data Jumlah Penduduk Miskin

(Sumber : <https://www.bps.go.id>)

No	Provinsi	Tahun	
		X	Y
1	Aceh	169,8	789,5
2	Sumatera Utara	694,2	1082



No	Provinsi	Tahun	
		X	Y
3	Sumatera Barat	120,8	334,2
4	Riau	177,6	421,8
5	Jambi	114,2	219,7
6	Sumatera Selatan	403,7	913,2
7	Bengkulu	103,5	269,3
8	Lampung	251,8	1109
9	Kep. Bangka Belitung	24,03	61,39
10	Kep. Riau	90,87	82,28
11	Dki Jakarta	374,1	299,3
12	Jawa Barat	2525	3095
13	Jawa Tengah	1978	3804
14	Di Yogyakarta	311,6	374,1
15	Jawa Timur	1686	4186
16	Banten	376	508
17	Bali	99,93	132,9
18	Nusa Tenggara Barat	423,4	642
19	Nusa Tenggara Timur	112,6	1025
20	Kalimantan Barat	86,41	360,5
21	Kalimantan Tengah	41,32	130,2
22	Kalimantan Selatan	65,43	160
23	Kalimantan Timur	96,28	191,3
24	Kalimantan Utara	18,41	36,96
25	Sulawesi Utara	66,21	169,3
26	Sulawesi Tengah	70,24	396,4
27	Sulawesi Selatan	150,1	775,8
28	Sulawesi Tenggara	45,62	319,6
29	Gorontalo	22,72	191,3
30	Sulawesi Barat	31,02	141,1
31	Maluku	49,13	314,6
32	Maluku Utara	10,86	80,8
33	Papua Barat	15,74	223,3
34	Papua	37,4	878,8

- b) Menetapkan nilai k jumlah *cluster* jumlah Jumlah Penduduk Miskin sebanyak 3 *cluster* (k-3). *Cluster* yang dibentuk yaitu *cluster* tertinggi, *cluster* sedang, dan *cluster* terendah.
- c) Menentukan nilai *centroid* (pusat *cluster*) awal yang telah ditentukan secara random berdasarkan nilai variabel data yang di *cluster* sebanyak k yang ditentukan sebelumnya.

Tabel 2. Penentuan Pusat Awal *cluster*

Cluster	X	Y
C1	2525	4186,4
C2	318,9	697,61
C3	10,86	36,964

- d) Menghitung jarak setiap data jumlah penduduk miskin terhadap pusat *cluster*
Setelah data nilai pusat *cluster* awal ditentukan maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat *cluster* dengan menggunakan rumus yang perhitungannya dapat kita lihat sebagai berikut:
Dilakukan perhitungan jarak terhadap jumlah penduduk miskin dengan titik pusat (*centroid*) pada *cluster* pertama.

$$D(1.1) = \sqrt{(169,8 - 2525)^2 + (789,5 - 4186,4)^2} = 4133,6$$

$$D(1.2) = \sqrt{(694,2 - 2525)^2 + (1082 - 4186,4)^2} = 3604,3$$

$$D(1.3) = \sqrt{(120,8 - 2525)^2 + (334,2 - 4186,4)^2} = 4540,9$$

$$D(1.4) = \sqrt{(177,6 - 2525)^2 + (421,8 - 4186,4)^2} = 4436,5$$

Dan seterusnya sampai dengan D(1.34).

Perhitungan jarak Jumlah Penduduk Miskin pertama dengan *centroid cluster* kedua, seperti berikut:

$$D(2.1) = \sqrt{(169,8 - 318,9)^2 + (789,5 - 697,61)^2} = 175,2$$

$$D(2.2) = \sqrt{(694,2 - 318,9)^2 + (1082 - 697,61)^2} = 537,02$$

$$D(2.3) = \sqrt{(120,8 - 318,9)^2 + (334,2 - 697,61)^2} = 413,88$$

Dan seterusnya sampai dengan D(2.34)

Perhitungan jarak Jumlah Penduduk Miskinpertama dengan *centroid cluster* ketiga, seperti berikut:

$$D(3.1) = \sqrt{(169,8 - 10,86)^2 + (789,5 - 36,964)^2} = 769,11$$

$$D(3.2) = \sqrt{(694,2 - 10,86)^2 + (1082 - 36,964)^2} = 1248,4$$

$$D(3.3) = \sqrt{(120,8 - 10,86)^2 + (334,2 - 36,964)^2} = 316,95$$

Berikut tabel 3 hasil perhitungan jarak data dengan titik pusat pada iterasi 1 menggunakan *Eulidean Distance*.

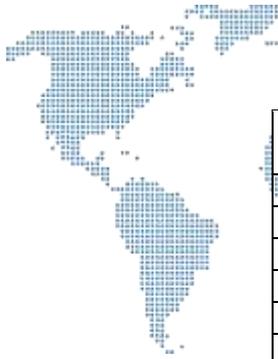
Tabel 3. Hasil Perhitungan Jarak Data Ke Titik Pusat Pada *Cluster* 1

No	Provinsi	Cluster			Jarak Terpendek
		C1	C2	C3	
1	Aceh	4133,6	175,2	769,11	175,2
2	Sumatera Utara	3604,3	537,02	1248,4	537,02
3	Sumatera Barat	4540,9	413,88	316,95	316,95
4	Riau	4436,5	309,86	419,45	309,86
5	Jambi	4641,8	519,88	209,97	209,97
6	Sumatera Selatan	3900,5	231,7	960,31	231,7
7	Bengkulu	4605,2	479,44	250,12	250,12
8	Lampung	3826,4	416,39	1098,3	416,39
9	Kep, Bangka Belitung	4824	701,24	27,757	27,757
10	Kep, Riau	4771,7	656,23	91,959	91,959
...
...
32	Maluku Utara	4814,3	689,47	43,832	43,832
33	Papua Barat	4690,7	562,91	186,43	186,43
34	Papua	4138,7	334,83	842,3	334,83

- e) Menentukan posisi *cluster* masing-masing data jumlah penduduk miskin berdasarkan jarak minimum data terhadap pusat *cluster*. Data yang memiliki jarak terkecil dengan *centroid* akan menjadi anggota pada kelompok tersebut. Berikut ini merupakan table 4 posisi data dengan tiap *cluster* pada iterasi ke-1 dengan menggunakan bantuan tanda (1) yang mengatakan data menjadi anggota pada *cluster*.

Tabel 4. Posisi Data Dengan Tiap *Cluster* Pada Iterasi Ke-1

No	Provinsi	Cluster		
		C 1	C 2	C 3
1	Aceh		1	
2	Sumatera Utara		1	
3	Sumatera Barat			1
4	Riau		1	
5	Jambi			1



No	Provinsi	Cluster		
		C 1	C 2	C 3
6	Sumatera Selatan		1	
7	Bengkulu			1
8	Lampung		1	
9	Kep. Bangka Belitung			1
10	Kep. Riau			1
...
...
31	Maluku			1
32	Maluku Utara			1
33	Papua Barat			1
34	Papua		1	

- f) Menghitung titik pusat baru menggunakan hasil dari setiap anggota pada masing-masing *cluster*. Contoh perhitungan titik pusat baru pada *cluster* X adalah:

$$C 1 = \frac{2525 + 1978 + 1686}{3} = 2063$$

$$169,8 + 694,2 + 177,6 + 403,7 + 251,8 + 374,1 + 311,6 + 376 + 423,4 + 112,6 + 150,1 + 37,4$$

$$C 2 = \frac{120,8 + 114,2 + 103,5 + 24,03 + 90,87 + 99,93 + 86,41 + 41,32 + 65,43 + 96,28 + 18,41 + 66,21 + 70,24 + 45,62 + 22,72 + 31,02 + 49,13 + 10,86 + 15,74}{12} = 290,2$$

$$C 3 = \frac{19}{19} = 61,7241$$

Contoh perhitungan titik pusat baru pada *cluster* Y adalah:

$$C 1 = \frac{3095 + 3804 + 4186}{3} = 3695$$

$$789,5 + 1082 + 421,8 + 913,2 + 1109 + 299,3 + 374,1 + 508 + 642 + 1025 + 775,8 + 878,8$$

$$C 2 = \frac{334,2 + 219,7 + 269,3 + 61,39 + 82,28 + 132,9 + 360,5 + 130,2 + 160 + 191,3 + 36,96 + 169,3 + 396,4 + 319,6 + 191,3 + 141,1 + 314,6 + 80,8 + 223,3}{12} = 734,85$$

$$C 3 = \frac{19}{19} = 200,801$$

- g) Selanjutnya dilakukan kembali langkah 4 sampai 6. Jika nilai *centroid* hasil iterasi dengan nilai *centroid* sebelumnya bernilai sama atau nilai *centroid* sudah optimal serta posisi *cluster* data jumlah penduduk miskin tidak mengalami perubahan lagi maka proses iterasi berhenti. Namun jika nilai *centroid* tidak sama atau belum optimal serta posisi data jumlah penduduk miskin masih berubah maka proses iterasi berlanjut pada iterasi berikutnya.

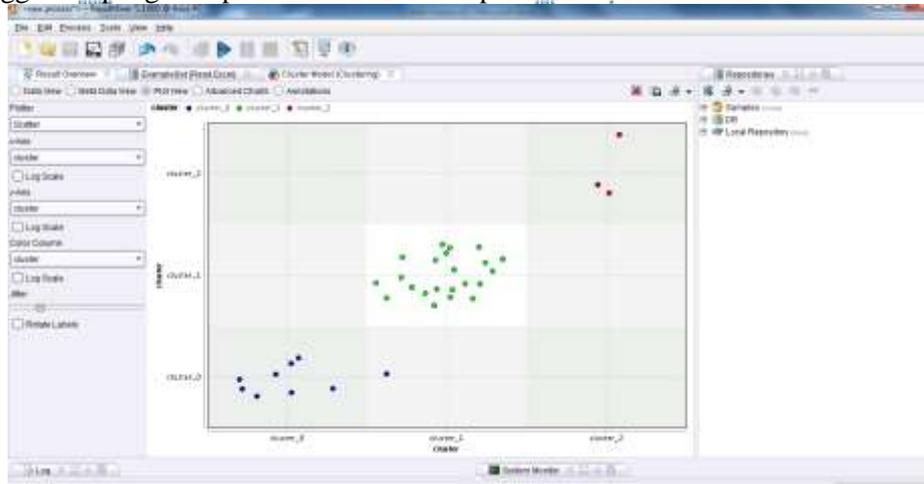
Dari perhitungan manual data Penduduk miskin menggunakan algoritma *K-Means* dapat dilihat bahwa hasil iterasi 2 dan iterasi 3 bernilai sama, hasil dari kedua iterasi tersebut bernilai C 1 = 3, C 2 = 9, dan C 3 = 22, posisi *cluster* data Penduduk miskin tidak mengalami perubahan lagi maka proses iterasi berhenti.

Berdasarkan posisi *cluster* masing-masing data penduduk miskin dan nilai *cluster* hasil iterasi ketiga maka dapat disimpulkan bahwa:

- Cluster* Pertama dengan jumlah penduduk miskin sebanyak 3 provinsi yaitu Jawa Barat, Jawa Tengah dan Jawa Timur
- Cluster* Kedua dengan jumlah penduduk miskin sebanyak 9 provinsi yaitu Aceh, Sumatra Utara, Sumatra Selatan, Lampung, Banten, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Selatan, Papua.
- Cluster* Ketiga dengan jumlah penduduk miskin sebanyak 22 provinsi yaitu Sumatra Barat, Riau, Jambi, Bengkulu, Kep Bangka Belitung, Kep Riau, DKI Jakarta, Di Yogyakarta, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah,

Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat.

Data hasil clustering menggunakan tools rapidminer menjelaskan bahwa terdapat 3 cluster yang dimulai dari cluster 0, cluster 1, dan cluster 2. Cluster 0 merupakan cluster tertinggi, cluster 1 merupakan cluster sedang, dan cluster 2 merupakan cluster terendah, sehingga didapat grafik plot view dari hasil rapidminer versi 5.0 berikut ini.



Gambar 1. Grafik Clustering

Berdasarkan pada Gambar 1. dapat diketahui bahwa titik berwarna merah adalah *cluster 0* (Tertinggi), titik berwarna hijau merupakan *cluster 1* (sedang). dan titik berwarna biru merupakan *cluster 2* (Terendah).



Gambar 2. Folder Fiew

Pada gambar 2 ini dijelaskan provinsi mana mana saja yang termasuk *cluster* tertinggi, *cluster* sedang dan *cluster* terendah. Cluster 0 Yaitu Aceh, Sumatra Utara, Sumatra Selatan, lampung, Banten, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Selatan, Papua. Cluster 1 Yaitu, Sumatra Barat, Riau, Jambi, Bengkulu, Kep Bangka Belitung, Kep Riau, DKI Jakarta, Di Yogyakarta, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, maluku Utara, Papua Barat. Cluster 2 Yaitu. Jawa Barat, Jawa Tengah Jawa Timur.

4. Kesimpulan

Dari hasil penelitian yang telah dilakukan, maka penulis dapat menarik beberapa kesimpulan, diantaranya sebagai berikut:

- a) Mengelompokkan jumlah penduduk miskin yang ada diprovinsi indonesia berhasil diterapkan menggunakan *Algoritma K-Means Clustering*.
- b) Pengujian yang dilakukan dalam penelitian ini, iterasi *clustering* pada data jumlah penduduk miskin terjadi sebanyak 3 kali *iterasi*,
- c) Dari Hasil Perhitungan k-means dengan Aplikasi Rapidminer hasil yang didapat bernilai sama yaitu *Cluster Pertama* dengan jumlah data penduduk miskin sebanyak 10 provinsi yaitu Sumatera Utara, Sumatera Selatan, Lampung, Banten, NTB, NTT, Suawesi Selatan dan Papua. *Cluster Kedua* jumlah data penduduk miskin sebanyak 22 Provinsi yaitu, Sumatera Barat, Riau, Jambi, Bengkulu, Kep Bangka Belitung, Kep Riau, DKI Jakarta, Di Yogyakarta, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, dan Papua Barat. *Cluster Ketiga* jumlah data penduduk miskin sebanyak 3 Provinsi yaitu, Jawa Barat, Jawa Tengah dan Jawa Timur.

Daftar Pustaka

- [1] K. Fatmawati And A. P. Windarto, "Data Mining : Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," Vol. 3, No. 2, Pp. 173–178, 2018.
- [2] T. Imandasari, E. Irawan, A. P. Windarto, And A. Wanto, "Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," *Semin. Nas. Ris. Inf. Sci.*, 2019, Doi: 10.30645/Senaris.V1i0.81.
- [3] J. Eska, "Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5," *Jurteks (Jurnal Teknol. Dan Sist. Informasi)*, Vol. 2, 2018, Doi: 10.31227/Osf.Io/X6svc.
- [4] Dan G. F. Ade Bastian, Harun Sujadi, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," No. 1, Pp. 26–32, 2018.
- [5] 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.
- [6] H. Sulastrri And A. I. Gufroni, "Jurnal Teknologi Dan Sistem Informasi Penerapan Data Mining Dalam Pengelompokan Penderita," Vol. 02, Pp. 299–305, 2017.
- [7] T. Khotimah, "Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means," *Simetris*, Vol. 5, No. 1, Pp. 83–88, 2014.