

Penerapan Algoritma *K-Means* Untuk Mengelompokkan Kemiskinan Provinsi Di Indonesia Tahun 2022

Muhammad Rizkiyanto
Teknik Informatika, Program Pascasarjana, Universitas Pamulang
e-mail: aarizkiyans@gmail.com

Abstrak—Kemiskinan adalah permasalahan khusus yang perlu ditanggulangi. Jumlah penduduk miskin di Indonesia pada maret 2023 sebanyak 25,90 juta orang. Dengan adanya hal tersebut, Pemerintah Indonesia memerlukan data sosial untuk mengetahui program seperti apa yang harus diberikan terhadap masyarakatnya. Penelitian ini bertujuan untuk mengelompokkan kemiskinan yang ada di Indonesia berdasarkan kesamaan karakteristiknya. Variabel yang digunakan pada penelitian ini adalah Persentase Penduduk Miskin (X_1), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Menurut dan Pendidikan yang Ditamatkan dibawah SD (X_2), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Menurut dan Pendidikan yang Ditamatkan Tamat SD/SMP (X_3), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Menurut dan Pendidikan yang Ditamatkan tamat SMA (X_4), Angka Melek Huruf Penduduk Miskin 15-55 (X_5), Angka Melek Huruf dan Angka Partisipasi Sekolah Penduduk Miskin 13-15 (X_6), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Tidak Bekerja (X_7), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Bekerja di Sektor Informal (X_8), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Bekerja di Sektor Formal (X_9), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Bekerja di Sektor Pertanian (X_{10}), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Bekerja Bukan di Sektor Pertanian (X_{11}). Hasil analisis yang diperoleh dari hasil perhitungan *Silhouette Index* didapat hasil bahwa pembentukan 3 kluster menjadi yang terbaik. Kemudian pada kluster 1 terdiri dari 8 Provinsi, kluster 2 terdiri dari 25 Provinsi, dan kluster 3 terdiri dari 1 Provinsi.

Kata Kunci—*Cluster; K-means; Silhouette Index; Kemiskinan; Indonesia.*

I. PENDAHULUAN

Kemiskinan adalah masalah yang sering dihadapi manusia. Kemiskinan termasuk kedalam sosial yang ada di seluruh dunia[1]. Keadaan dan situasi yang terjadi di suatu negara merupakan faktor yang berpengaruh terhadap kemiskinan di negara tersebut[2]. Upaya pengentasan kemiskinan terdapat dua strategi yang harus di tempuh. Pertama, melindungi keluarga dan kelompok masyarakat miskin melalui pemenuhan kebutuhan mereka dari berbagai bidang. Kedua, melakukan pelatihan kepada mereka agar mempunyai kemampuan untuk melakukan usaha pencegahan terjadinya kemiskinan baru[3].

Di Indonesia, persentase kemiskinan pada maret 2023 sebesar 9.36 persen, di mana persentase penduduk miskin di perkotaan sebesar 7,29 persen dan persentase penduduk miskin di perdesaan sebesar 12,22 persen pada Maret 2023. Lalu, jumlah penduduk miskin di indonesia pada maret 2023 sebesar 25,90 juta orang, di mana jumlah penduduk miskin di perkotaan sebanyak 11,74 orang dan jumlah penduduk miskin di perdesaan sebanyak 14,16 juta pada Maret 2023[4]. Pengalaman pengentasan kemiskinan di masa lalu menunjukkan banyak kelemahan, seperti: (1) orientasi pertumbuhan makro tanpa mempertimbangkan aspek pemerataan, (2) kebijakan yang terpusat, (3) lebih lanjut karikatur dibandingkan transformatif, (4) memposisikan masyarakat sebagai objek bukan subjek, (5) orientasi pengentasan kemiskinan cenderung bersifat karikatur dan instan dibandingkan produktivitas berkelanjutan, dan (6) perspektif dan solusi yang bersifat generik terhadap permasalahan kemiskinan yang ada tanpa memandangnya keanekaragaman yang ada. Karena begitu beragamnya sifat tantangan yang ada, maka penanganannya permasalahan kemiskinan harus menyentuh akar permasalahan yang sebenarnya, baik secara langsung maupun tidak langsung secara tidak langsung[5].

Berdasarkan data diatas, banyak sekali ilmu komputer yang dapat menyelesaikan permasalahan tersebut, salah satunya algoritma *k-means*. *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* yaitu sangat sensitif terhadap pembangkitan titik pusat awal secara random dan hasil pengelompokkan bersifat tidak unik (selalu berubah-ubah)[2]. Dari permasalahan diatas, tujuan Penelitian ini adalah untuk mengelompokkan kemiskinan di Indonesia yang mengacu pada kesamaan karakteristiknya berdasarkan Data dan Informasi Kemiskinan 2022.

II. TINJAUAN PUSTAKA

A. Data Mining

Data mining adalah istilah yang digunakan untuk menggambarkan hasil pengetahuan dalam database. Penambangan data adalah proses penggunaan teknik statistik, matematika, kecerdasan buatan, dan data pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi dan pengetahuan berguna terkait database[6].

Data mining adalah proses pencarian pola atau informasi menarik pada data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma yang digunakan dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat bergantung pada tujuan dan proses penemuan pengetahuan dalam database (KDD) secara keseluruhan[7].

B. K-Means

K-Means merupakan metode clustering data non-hierarki yang dapat membagi data menjadi dua kelompok atau lebih. Metode ini akan membagi data menjadi beberapa kelompok dimana data yang mempunyai karakteristik yang sama akan ditempatkan pada kelompok yang sama sedangkan data yang mempunyai karakteristik yang berbeda akan dikelompokkan pada kelompok yang lain. Tujuan dari clustering adalah untuk meminimalkan fungsi tujuan yang terbentuk dalam proses clustering, dan secara umum akan berusaha untuk meminimalkan varians dalam kelompok dan memaksimalkan varians antar kelompok [8].

Pada algoritma *K-Means*, objek atau data yang ada dikelompokkan menjadi k kelompok atau *cluster*. Untuk melakukan agregasi ini, nilai k harus ditentukan terlebih dahulu. Pengguna biasanya sudah mempunyai informasi awal mengenai objek yang diteliti, termasuk jumlah *cluster* yang paling tepat. Dalam algoritma *K-Means*, ukuran ketidaksamaan digunakan untuk mengelompokkan objek. Perbedaan ini diterjemahkan ke dalam konsep jarak. Jika jarak dua benda cukup dekat maka kedua benda tersebut sebangun. Semakin dekat, semakin besar kemiripannya. Semakin besar nilai jarak maka semakin besar perbedaannya[9]. Berikut ini adalah langkah-langkah dalam algoritma K-Means:

1. Tentukan jumlah c klaster (k) pada data set.
2. 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:

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

Keterangan:

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

N_i = Jumlah anggota klaster ke-i

i, k = indeks dari klaster

j = indeks dari variabel

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

3. Pada masing-masing record, hitung jarak terdekat dengan centroid. Jarak centroid yang digunakan adalah Euclidean Distance, dengan rumus seperti pada persamaan:

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

Keterangan:

De = Eulidean Distance

I = Banyaknya objek

(x, y) = Koordinat objek

(s, t) = Koordinat centroid

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

C. Silhouette Index

Indeks Silhouette mengukur derajat kepercayaan dalam proses clustering pada pengamatan tertentu dengan cluster yang dikatakan terbentuk baik bila nilai indeks mendekati 1 dan kondisi sebaliknya jika nilai indeks mendekati angka -1[10].

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

$a(i)$: Jarak rata-rata i dan seluruh pengamatan lainnya pada kluster yang sama

$b(i)$: Jarak rata-rata antara i dengan pengamatan kluster

D. Kemiskinan

Konsep kemiskinan mencakup berbagai dimensi, berbagai definisi, dan alternatif pengukuran. Kemiskinan merupakan masalah yang sulit untuk didefinisikan dan dijelaskan. Secara umum, kemiskinan dapat diukur dalam dua dimensi: dimensi pendapatan atau kekayaan dan dimensi faktor non finansial. Kemiskinan dalam dimensi pendapatan atau kekayaan tidak hanya diukur dari rendahnya pendapatan yang diterima karena rendahnya pendapatan biasanya bersifat sementara, namun juga dari kepemilikan aset seperti tanah bagi petani kecil dan akses terhadap pelayanan publik. Sedangkan dimensi faktor non finansial ditandai dengan keputusan atau ketidakberdayaan yang juga dapat mempengaruhi berbagai rumah tangga berpendapatan rendah[11].

III. METODE PENELITIAN

A. Jenis Penelitian

Pada penelitian ini menggunakan jenis penelitian kuantitatif. Penelitian kuantitatif merupakan pendekatan penelitian yang mewakili paham positivisme, sementara itu penelitian kualitatif merupakan pendekatan penelitian yang mewakili paham naturalistik (fenomenologis)[12].

B. Jenis Data

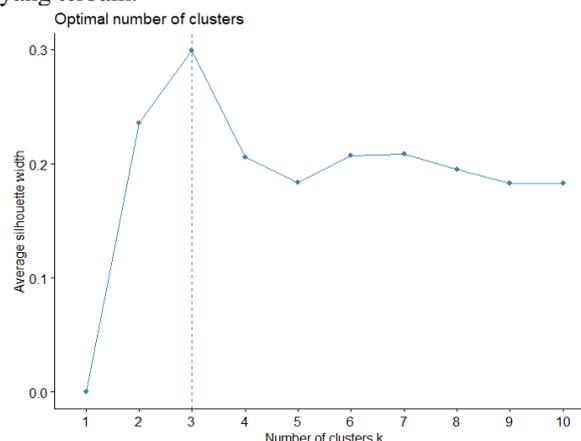
Data yang digunakan pada penelitian ini adalah Data dan Informasi Kemiskinan 2022 yang didapat dari Badan Pusat Statistik. Penelitian ini mengamati 34 provinsi yang ada di Indonesia yang terdiri dari Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua.

C. Variabel Penelitian

Indikator kemiskinan yang digunakan pada penelitian ini yaitu Persentase Penduduk Miskin (X_1), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Menurut dan Pendidikan yang Ditamatkan dibawah SD (X_2), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Menurut dan Pendidikan yang Ditamatkan Tamat SD/SMP (X_3), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Menurut dan Pendidikan yang Ditamatkan tamat SMA (X_4), Angka Melek Huruf Penduduk Miskin 15-55 (X_5), Angka Melek Huruf dan Angka Partisipasi Sekolah Penduduk Miskin 13-15 (X_6), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Tidak Bekerja (X_7). Persentase Penduduk Miskin Usia 15 Tahun ke Atas Bekerja di Sektor Informal (X_8), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Bekerja di Sektor Formal (X_9), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Bekerja di Sektor Pertanian (X_{10}), Persentase Penduduk Miskin Usia 15 Tahun ke Atas Bekerja Bukan di Sektor Pertanian (X_{11}). Metode yang digunakan adalah algoritma *K-means*.

IV. HASIL DAN PEMBAHASAN

Penerapan algoritma *K-means* yang dilakukan terhadap variabel penelitian akan dilakukan pengujian *Silhouette Index* untuk mencari pembentukan kluster mana yang terbaik.



Gambar 1. *Silhouette Index*

Pada gambar 4.1 menunjukkan hasil *Silhouette Index* yang menandakan bahwa dari pembentukan kluster 1 sampai 10 adalah kluster 3 mendapat nilai tertinggi. Artinya pembentukan 3 kluster memiliki nilai keakuratan yan terbaik diantara kluster lainnya. Kemudian dilakukan pembentukan 3 kluster untuk mengetahui pembagian provinsi terhadap klasternya masing-masing.

Tabel 1. Nilai Centroid Kluster

K	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	5.970	14.0450	54.58125	31.37625	99.12375	91.60375	51.1350	26.3725	22.49375	14.2875	34.5775
2	10.958	20.7816	53.17880	26.03880	97.98320	92.27240	44.4336	39.2956	16.27080	32.0040	23.5636
3	26.560	47.4800	33.58000	18.94000	73.64000	73.30000	25.3700	71.9500	2.68000	70.7000	3.9300

Tabel 2. Nilai Kategori Centroid Kluster

K	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	P	P	P	P	P	Netral	N	N	P	N	P
2	Netral	Netral	Netral	Netral	Netral	P	Netral	Netral	Netral	Netral	Netral
3	N	N	N	N	N	N	P	P	N	P	N

P: Nilai positif

N: Negatif

Netral: Nilai sedang/diantara positif dan negatif

Tabel 3. Pembagian Kluster

No	Kab.Kota	Kluster
1	ACEH	2
2	SUMATERA UTARA	2
3	SUMATERA BARAT	2
4	RIAU	2
5	JAMBI	2
6	SUMATERA SELATAN	2
7	BENGKULU	2
8	LAMPUNG	2
9	BANGKA BELITUNG	1
10	KEPULAUAN RIAU	1
11	DKI JAKARTA	1
12	JAWA BARAT	1
13	JAWA TENGAH	2
14	DI YOGYAKARTA	2
15	JAWA TIMUR	2
16	BANTEN	1
17	BALI	1
18	NUSA TENGGARA BARAT	2
19	NUSA TENGGARA TIMUR	2
20	KALIMANTAN BARAT	2
21	KALIMANTAN TENGAH	2
22	KALIMANTAN SELATAN	2
23	KALIMANTAN TIMUR	1
24	KALIMANTAN UTARA	2
25	SULAWESI UTARA	1
26	SULAWESI TENGAH	2
27	SULAWESI SELATAN	2
28	SULAWESI TENGGARA	2
29	GORONTALO	2
30	SULAWESI BARAT	2
31	MALUKU	2
32	MALUKU UTARA	2
33	PAPUA BARAT	2
34	PAPUA	3

Dengan melihat hasil tabel 1 dapat ditentukan bahwa kluster kemiskinan dibagi menjadi 3 kategori. Dengan kategori kemiskinan rendah, kategori kemiskinan sedang, dan kategori kemiskinan tinggi. Di mana dapat dilihat berdasarkan pengaruh nilai positif, nilai negatif, dan nilai netral yang dihasilkan dari variabel-variabel yang ada. Dari tabel 4.3 diperoleh pola pembagian kluster sebagai berikut.

- a. Pada kluster 1 terdiri dari Bali, Bangka Belitung, Banten, DKI Jakarta, Jawa Barat, Kalimantan Timur, Kepulauan Riau, dan Sulawesi Utara. Kluster 1 termasuk ke dalam kategori kemiskinan rendah karena dilihat dari tabel 4.2 nilai dari variabel yang ada di kluster 1 mayoritas bernilai positif.
- b. Pada kluster 2 terdiri dari Aceh, Bengkulu, DI Yogyakarta, Gorontalo, Jambi, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Utara, Lampung, Maluku, Maluku Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua Barat, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Sumatera Barat, Sumatera Selatan, dan Sumatera Utara. Kluster 2 termasuk ke dalam kategori kemiskinan sedang karena dilihat dari tabel 4.2 nilai variabel yang ada di kluster 2 mayoritas bernilai netral.

- c. Pada klaster 3 terdiri dari Papua. Klaster 3 termasuk ke dalam kategori kemiskinan tinggi karena dilihat dari tabel 4.2 nilai dari variabel yang ada di klaster 3 mayoritas bernilai negatif.

V. KESIMPULAN

Berdasarkan hasil analisis pengelompokan kemiskinan 34 Provinsi di Indonesia tahun 2022 menggunakan algoritma *k-means* yang telah dilakukan. Pengelompokan kemiskinan dilakukan berdasarkan nilai dari perhitungan *Silhouette Index* yang membandingkan pembentukan 1 sampai 10 klaster, di mana pembentukan 3 klaster mendapatkan posisi grafik tertinggi sekaligus menjadi pembentukan klaster yang terbaik. Pembentukan 3 klaster memiliki kategori masing-masing yaitu kemiskinan dengan kategori rendah, kemiskinan dengan kategori sedang, dan kemiskinan dengan kategori tinggi. Pada klaster 1 merupakan kelompok kemiskinan dengan kategori rendah terdiri dari Bali, Bangka Belitung, Banten, DKI Jakarta, Jawa Barat, Kalimantan Timur, Kepulauan Riau, dan Sulawesi Utara. Pada klaster 2 merupakan kelompok kemiskinan dengan kategori sedang terdiri dari Aceh, Bengkulu, DI Yogyakarta, Gorontalo, Jambi, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Utara, Lampung, Maluku, Maluku Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua Barat, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Sumatera Barat, Sumatera Selatan, dan Sumatera Utara. Dan pada klaster 3 merupakan kelompok kemiskinan dengan kategori tinggi terdiri dari Papua.

DAFTAR PUSTAKA

- [1] N. Nurwati, "Kemiskinan : Model Pengukuran , Permasalahan dan Alternatif Kebijakan," *J. Kependud. Padjadjaran*, vol. 10, no. 1, pp. 1–11, 2008.
- [2] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 76–83, 2020, doi: 10.47065/bits.v2i2.492.
- [3] D. V. Ferezagia, "Analisis Tingkat Kemiskinan di Indonesia," *J. Sos. Hum. Terap.*, vol. 1, no. 1, pp. 1–6, 2018, doi: 10.7454/jst.v1i1.6.
- [4] Badan Pusat Statistik Indonesia, "BERITA RESMI STATISTIK INDONESIA 2023," *Badan Pus. Stat.*, no. 57, pp. 1–8, 2018, [Online]. Available: <https://www.bps.go.id/pressrelease/2018/07/16/1483/persentase-penduduk-miskin-maret-2018-turun-menjadi-9-82-persen.html>
- [5] A. V. D. Sano and H. Nindito, "Application of K-Means Algorithm for Cluster Analysis on Poverty of Provinces in Indonesia," *ComTech Comput. Math. Eng. Appl.*, vol. 7, no. 2, p. 141, 2016, doi: 10.21512/comtech.v7i2.2254.
- [6] E. Turban and J. E. Aronson, *Deciclon Support Systems and Intelligent System*. Yogyakarta: Andi Offset, 2005.
- [7] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- [8] E. Prasetyo, *Data Mining : Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi, 2014.
- [9] B. & Santosa and Ardian Umam, *Data Mining dan Big Data Analytics*. Yogyakarta: Penebar Media Pustaka, 2018.
- [10] N. Afira and A. W. Wijayanto, "Analisis Cluster Kemiskinan Provinsi di Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarki Cluster Analysis with Partitioning and Hierarchical Methods on Provincial Poverty Information Data in Indonesia in 2019," vol. 10, no. 28, 2021, doi: 10.34010/komputika.v10i2.4317.
- [11] W. K. Bellingier, *The Economics Analysis of Public Policy*. Routledge: Oxon, 2007.
- [12] M. Mulyadi, "PENELITIAN KUANTITATIF DAN KUALITATIF SERTA PEMIKIRAN DASAR MENGGABUNGKANNYA," vol. 15, no. 1, pp. 127–138, 2011.