

Artikel Ilmiah Hasil Riset

K-MEANS CLUSTERING DENGAN METODE ELBOW UNTUK PENGELOMPOKAN KABUPATEN DAN KOTA DI JAWA TIMUR BERDASARKAN INDIKATOR KEMISKINAN

Anita Fitria Febrianti¹, Antonito Hornay Cabral², Gangga Anuraga³
Program Studi Statistika, Universitas PGRI Adi Buana Surabaya^{1,2,3}
anitiftriafeb98@gmail.com

ABSTRAK

Provinsi Jawa Timur merupakan salah satu provinsi dengan kabupaten dan kota terbanyak di Indonesia yang terdiri dari 29 kabupaten dan 9 kota. Besarnya wilayah di Jawa Timur menyebabkan pemerintah daerah sulit menangani kasus kemiskinan. Oleh karena itu perlu adanya pengelompokan kabupaten dan kota supaya mudah dalam menangani permasalahan dan mengetahui kabupaten atau kota yang membutuhkan bantuan Pemerintah. Tujuan dari penelitian ini adalah untuk mengelompokkan kabupaten dan kota di Jawa Timur berdasarkan indikator kemiskinan. Pada penelitian ini menggunakan Metode *K-Means Cluster*, merupakan salah satu metode yang bisa digunakan untuk mengelompokkan kabupaten dan kota di Jawa Timur. Algoritma pada metode *K-Means Cluster* menetapkan suatu obyek ke dalam *cluster* yang mempunyai *centroid* (mean) terdekat. Hasil penelitian menunjukkan bahwa dengan mengelompokkan kabupaten dan kota kedalam 4 *cluster* maka dapat diperoleh *cluster* pertama terdiri dari 10 kabupaten, *cluster* kedua terdiri dari 5 kabupaten, *cluster* 3 terdiri dari 19 kabupaten dan *cluster* 4 terdiri dari 4 kabupaten. Hasil karakteristik *cluster* pertama memiliki presentase penduduk 15 tahun keatas pendidikan SLTA yang tinggi, *cluster* kedua memiliki rumah tangga menerima raskin yang cukup tinggi, *cluster* ketiga memiliki angka partisipasi sekolah rendah dan karakteristik pada *cluster* keempat memiliki penduduk 15 tahun keatas yang bekerja di sektor pertanian yang sangat tinggi.

Kata kunci: kemiskinan, *K-Means Cluster*, *Elbow Method*

ABSTRACT

East Java Province is one of the provinces with the most regencies and cities in Indonesia, consisting of 29 regencies and 9 cities. The size of the area in eastern Java has made it difficult for regional governments to handle poverty cases. therefore it is necessary to group districts and cities that need government assistance. The purpose of this study is to classify districts and cities in East Java. Based on poverty indicators. In this study using the K-Means Cluster method, is one method that can be used to group districts and cities in East Java. Algorithms in the K-Means Cluster method assign an object to a cluster that has the closest centroid (mean). The results showed that by grouping districts and cities into 4 duster, the first cluster consisted of 10 districts, the second cluster consisted of 5 districts, cluster 3 consisted of 19 districts and cluster 4 consisted of 4 districts. The results of the first cluster characteristics have a percentage of the population of 15 years and above for high school education which is high, the second cluster has households that receive quite high Raskin. The third cluster has low school achievement rates and the characteristics of the fourth cluster have a population of 15 years and over who work in a very high agricultural sector.

Keywords : Poverty, *K-Means Cluster*, *Elbow Method*

PENDAHULUAN

Analisis kelompok atau *cluster analysis* adalah analisis statistika yang bertujuan untuk mengelompokkan obyek-obyek berdasarkan kesamaan karakteristik. Obyek tersebut akan diklasifikasikan ke dalam satu atau beberapa kelompok (*cluster*) sehingga obyek-obyek yang berada dalam satu kelompok akan mempunyai kemiripan satu dengan yang lain. Dengan demikian terdapat homogenitas (kesamaan) yang tinggi antar anggota dalam satu kelompok (*within-cluster*) dan heterogenitas (perbedaan) yang tinggi antar kelompok yang satu dengan kelompok lainnya (*between-cluster*)[1].

Beberapa penelitian terkait analisis kelompok diantaranya pernah dilakukan oleh Anuraga (2015), dalam penelitiannya yang berjudul *Hierarchical Clustering Multiscale Bootstrap* untuk pengelompokan kemiskinan di Jawa Timur menyimpulkan bahwa terdapat lima kelompok (*cluster*) yang terbentuk dengan kemiripan satu dengan yang lain dan signifikan dengan nilai AU (*Approximately Unbiased*) $p_value \geq 0,95$ [2]. Febriyana, (2011) melakukan analisis klaster *k-means* dan *k-median* pada data indikator kemiskinan (studi kasus data indikator Kemiskinan Kabupaten di Indonesia tahun 2009). Hasil yang diperoleh adalah metode pengklasteran *K-Means* lebih baik daripada *K-Median* untuk kasus tersebut, karena ketepatan klasifikasi *K-Means* sebesar 98,51%, sedangkan *K-Median* sebesar 97,57%[3]. Alfina, dkk. (2012) meneliti tentang perbandingan metode *Hierarchical Clustering*, *K-Means* dan

Gabungan Keduanya dalam Cluster Data. Hasil yang diperoleh gabungan metode *single linkage* dan *K-means* memberikan hasil cluster lebih baik dengan parameter uji cluster variance dan metode *silhouette*[4].

Kemiskinan adalah keadaan dimana terjadi ketidakmampuan untuk memenuhi sandang, pangan, papan, serta pendidikan dan kesehatan. Kemiskinan dapat disebabkan oleh kelangkaan kebutuhan dasar ataupun sulitnya akses terhadap pendidikan dan pekerjaan. Angka kemiskinan Kabupaten /Kota di Jawa Timur memiliki posisi cukup tinggi. Angka kemiskinan ini terjadi diduga karena para petani sudah tidak memiliki lahan sendiri. Sebagian besar di Jawa Timur adalah buruhtani. Berdasarkan data dari Badan Pusat Statistik kemiskinan di Provinsi Jawa Timur sejak September 2014 hingga Maret 2015 tercatat garis kemiskinan naik sebesar 6,49%, sedangkan di wilayah perkotaan hanya 3,93%[5]. Tingkat kemiskinan ini juga dipengaruhi karena tingkat pendidikan, lapangan kerja yang kurang memadai, pengeluaran perkapita untuk makanan, dan juga penerima raskin di Kabupaten/Kota.

Tingkat kemiskinan di Jawa Timur dari tahun ke tahun mengalami kenaikan. Untuk itu dalam menangani angka kemiskinan, pemerintah berperan untuk mengelompokkan daerah yang membutuhkan pendidikan yang layak, lapangan pekerjaan sesuai dengan potensi yang dimiliki khususnya pada wilayah yang memiliki angka kemiskinan yang tinggi. Hasil pengelompokan ter-

sebut dapat digunakan oleh pemerintah dalam merumuskan kebijakan terkait persebaran kemiskinan

Dalam penelitian ini digunakan data indikator kemiskinan menurut kabupaten dan kota di Jawa Timur tahun 2016. Indikator kemiskinan diantaranya adalah rumah tangga miskin yang menggunakan air layak dan jamban sendiri, rumah tangga menerima raskin, pengeluaran untuk makanan, angka melek huruf, presentase penduduk miskin yang bekerja di sektor informal, dan sebagainya. Tujuan dari penelitian ini adalah mengelompokkan kabupaten dan kota di Jawa Timur berdasarkan indikator dengan menggunakan analisis *K-Means Cluster*. Pengelompokan yang dilakukan ini diharapkan dapat memberikan informasi yang bermanfaat serta masukan bagi pemerintah untuk menangani daerah yang mempunyai kemiskinan yang tinggi dan dapat melakukan pemetaan sesuai dengan kelompoknya.

METODE

Data yang digunakan dalam penelitian merupakan data sekunder yang diambil dari Badan Pusat Statistik (BPS). Data yang digunakan berupa data indikator kemiskinan seluruh kabupaten dan kota di Jawa Timur tahun 2016, yang terdiri dari 29 kabupaten dan 9 kota. Variabel-variabel tersebut dapat dilihat pada tabel berikut:

Tabel 3.1 Variabel Penelitian

No.	Keterangan	Skala
1	Presentase rumah tangga miskin menggunakan air layak dan jamban sendiri (X_1)	Rasio

2	presentase rumah tangga miskin menerima raskin (X_2)	Rasio
3	Presentase pengeluaran per kapita untuk makanan (X_3)	Rasio
4	Presentase perempuan usia 15-49 tahun menggunakan KB (X_4)	Rasio
5	Presentase penduduk 15 tahun keatas bekerja di sektor pertanian (X_5)	Rasio
6	Presentase penduduk miskin 15 tahun keatas dengan pendidikan SLTA (X_6)	Rasio
7	Presentase angka melek huruf penduduk 15-55 tahun (X_7)	Rasio
8	Presentase angka partisipasi sekolah penduduk usia 13-15 tahun (X_8)	Rasio

Analisis Pengelompokan

Analisis kelompok atau *cluster analysis* merupakan salah satu teknik multivariat yang masuk dalam klasifikasi metode interdependen, yaitu metode yang mempelajari struktur hubungan antara variabel, antara kasus/responden dan antara obyek. *Analisis cluster* atau kelompok bertujuan untuk mengelompokkan obyek atau kasus ke dalam kelompok yang mempunyai sifat yang relatif homogen. Obyek-obyek yang mempunyai kemiripan dikelompokkan dalam *cluster* yang sama, sedangkan obyek-obyek lain yang tidak mirip dikelompokkan dalam *cluster* yang lain[1].

Menurut Santoso (2004) analisis kelompok merupakan pengelompokan data yang dilakukan dengan dua macam metode yaitu metode hierarki dan meto-

de non hierarki. Metode hirarki merupakan pengelompokan yang dilakukan secara hirarki atau berjenjang dari n , ($n-1$) sampai 1 kelompok. Metode non hirarki dilakukan dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan. Salah satu prosedur non hirarki yang terkenal adalah metode *K-Means*[6].

Ukuran kemiripan yang biasa digunakan dalam analisis *cluster* adalah jarak *Euclidean* dan jarak *Mahalanobis*. Jarak *Euclidean* digunakan jika variabel amatan saling bebas atau berkorelasi satu sama lain (tidak terjadi multikolinieritas). Jika terjadi multikolinieritas maka dapat diatasi dengan mentransformasi data menggunakan *Principal Component Analysis* (PCA)[1].

Johnson dan Wichern (1982) menyatakan bahwa Jarak *Euclidean* antara $\mathbf{x} = [x_1, x_2, \dots, x_p]^T$, $\mathbf{y} = [y_1, y_2, \dots, y_p]^T$ adalah sebagai berikut[1]:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + \dots + (x_p - y_p)^2} \quad (1)$$

K-Means Cluster

Johnson dan Wichern (1982) menyarankan bahwa algoritma pada metode *K-Means Cluster* menetapkan suatu obyek ke dalam *cluster* yang mempunyai *centroid* (mean) terdekat[1]. Proses ini dibagi dalam tiga tahap:

1. Mempartisi obyek-obyek ke dalam k *cluster* awal.
2. Menentukan obyek yang mempunyai jarak terdekat dengan *centroid* (mean). (Ukuran jarak yang biasa digunakan adalah jarak *Euclidean*, yang digunakan baik untuk pengamatan yang distandarkan maupun yang tidak distandarkan).

Menghitung ulang *centroid* yang digunakan *cluster* untuk mendapatkan obyek yang baru dan untuk *cluster* yang kehilangan obyek.

3. Mengulangi langkah ke-2 sampai tidak ada lagi obyek yang dipindahkan.

Metode Elbow

Metode dan analisis ini digunakan untuk pemilihan jumlah *cluster* atau kelompok yang optimal. Berikut disajikan algoritma *elbow* dalam menentukan jumlah kelompok yang terbentuk [7].

Yaitu berdasarkan pada *sum of square error* (SSE).

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|X_i - C_k\|^2 \quad (2)$$

dimana K adalah banyaknya kelompok yang digunakan pada algoritma *K-Means* X_i adalah jumlah data dan C_k adalah banyaknya *cluster* pada *cluster* ke k .

Langka-langkah analisis data

Langkah-langkah penelitian dalam penelitian ini adalah sebagai berikut:

1. Analisis deskriptif untuk mengetahui gambaran indikator kemiskinan Jawa Timur
2. Menentukan jumlah *cluster* yang akan dibentuk. Dalam penelitian ini jumlah *cluster* yang dibentuk adalah sebanyak 4 *cluster*
3. Lakukan analisis *cluster* dengan metode *K-Means Cluster* sesuai dengan langkah 1.
4. Membagi anggota-anggota ke dalam setiap *cluster*
5. Deskripsikan karakteristik *cluster* yang didapatkan pada langkah 2

- Melakukan kesimpulan terhadap analisis yang dilakukan

HASIL

Deskripsi indikator kemiskinan Jawa Timur

Tabel 2 Analisis deskriptif dilakukan untuk mengetahui gambaran secara umum karakteristik untuk masing-masing variabel yang membentuk indikator kemiskinan pada setiap kabupaten/kota di Provinsi Jawa Timur.

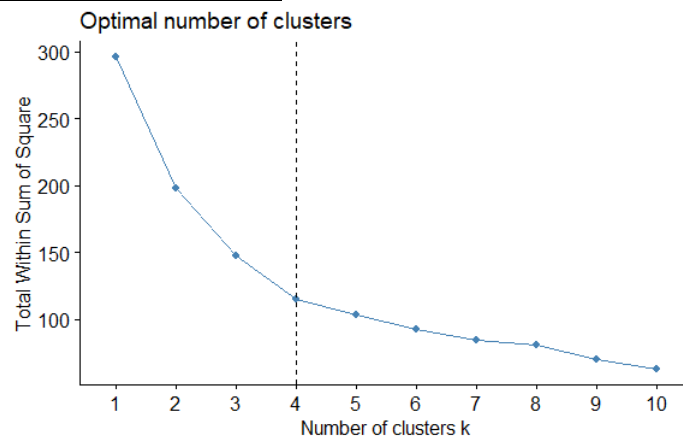
Tabel 2. Statistik deskriptif variabel penelitian

Var	Min	Rata-rata	Standar deviasi	Maks
X1	32.1	76.10	16.17	98.93
X2	20.09	76.48	19.28	99.07
X3	54.78	62.31	3.57	69.5
X4	32.44	91.61	12.92	98
X5	0	29.74	19.06	64.69
X6	4.22	14.93	8.37	35.23
X7	82.18	95.04	4.54	100
X8	73.7	94.25	6.75	100

Berdasarkan tabel 2 diatas terlihat bahwa secara rata-rata sebagian besar rumah tangga menerima raskin masih terkolong tinggi yaitu sebesar 1718.53 dengan standar deviasi sebesar 419.24. selain itu angka melek huruf di Jawa Timur tergolong cukup tinggi yaitu sebesar 95.04%.

Analisis K-Means cluster indikator kemiskinan

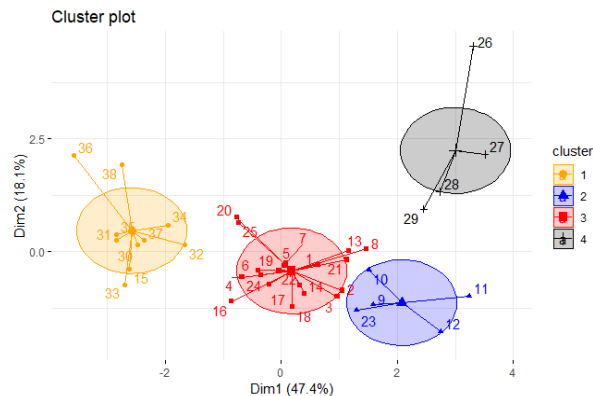
Sebelum melakukan analisis analisis *K-Means Cluster*, perlu ditentukan terlebih dahulu jumlah *cluster*. Dalam penelitian ini jumlah cluster yang ditentukan sebanyak 4 *cluster*. Dengan menggunakan algoritma k-means Cluster dengan kasus indikator kemiskinan sebanyak 4 cluster merupakan cluster optimal. Untuk lebih jelas dapat dilihat pada plot berikut:



Gambar 1. Jumlah cluster optimal

Berdasarkan gambar 1 dapat dilihat bahwa jumlah cluster yang optimal sebanyak 4 cluster. Untuk anggota

cluster dapat dilihat pada gambar 2 dibawah ini.



Gambar 2. Jumlah anggota cluster

Berdasarkan gambar 2 dapat dilihat bahwa cluster 1 terdapat 10 kabupaten/kota, cluster 2 terdapat 5 kabupaten dan kota, cluster 3 terdapat 19 kabupaten/kota dan cluster 4 terdiri dari 4 kabupaten. Berikut adalah hasil keanggotaan kabupaten/kota pada setiap cluster yang disajikan dalam tabel 3:

Tabel 3. Anggota-anggota Cluster

Cluster	Anggota cluster
1	Sidoarjo, kota Kediri, kota Blitar, kotaMalang, kota Probolinggo, kota Pasuruan, kota Mojokerto, kota Madiun, kota Surabaya dan kota Batu.
2	Jember, Banyuwangi, Bondowoso, Situbondo dan Tuban.
3	Pacitan, Ponorogo, Treng-galek, Tulungagung, Blita, Kediri, Malang, Lumajang, Probolinggo, Pasuruan, Mojo-kerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Lamongan, Gresik
4	Bangkalan, Sampang, Pamekasan, sumenep

Masing-masing cluster yang terbentuk memiliki karakteristik yang berbeda-beda. Berikut adalah karakteris-

tik setiap cluster berdasarkan indikator kemiskinan. Adapun hasil rata-rata setiap cluster yang disajikan dalam tabel 4 dibawah ini:

Tabel 4. rata-rata presentase indikator kemiskinan di setiap cluster

Var	Cluster			
	1	2	3	4
X1	0.532	-1.723	0.115	0.275
X2	-1.060	0.713	0.467	-0.460
X3	-1.242	0.717	0.143	1.529
X4	0.463	-0.144	0.272	-2.272
X5	-1.259	0.277	0.276	1.492
X6	1.327	-0.706	-0.341	-0.824
X7	0.726	-0.940	0.253	-1.841
X8	0.324	-1.589	0.303	-0.264

Berdasarkan tabel 4 diatas dapat dilihat bahwa karakteristik pada cluster 1 mempunyai presentase penduduk 15 tahun keatas dengan pendidikan SLTA dan angka melek huruf penduduk 15-55 tahun yang tinggi.Senada dengan penelitian yang dilakukan oleh Febri Angelina M.(2015) menyatakan bahwa ciri-ciri kawasan miskin dipengaruhi dengan tingkat pendidikan dan angka

melek huruf. Ketika tingkat pendidikan dan angka melek huruf tinggi maka akan menurunkan tingkat kemiskinan di suatu kabupaten/kota [8]. Karakteristik pada cluster 2 memiliki persentase rumah tangga menerima raskin dan pengeluaran perkapita untuk makanan cukup tinggi dengan rata-rata sebesar 0.713 dan 0.717 dari sini dapat ditentukan di kluster 2 didominasi oleh kabupaten /kota pesisir di Jawa Timur. Pada karakteristik cluster 3 mempunyai rumah tangga menerima raskin dan angka partisipasi sekolah penduduk usia 13-15 tahun yang rendah. Kemudian cluster 4 mempunyai persentase pengeluaran perkapita untuk makanan dan persentase penduduk miskin 15 tahun keatas yang bekerja di sektor pertanian yang sangat tinggi.

SIMPULAN

Berdasarkan hasil pembahasan dapat disimpulkan bahwa kondisi indikator kemiskinan pada setiap kabupaten dan kota di Jawa Timur masih tidak seimbang, sehingga ada beberapa kabupaten dan kota di provinsi Jawa Timur masih memiliki beberapa indikator kemiskinan yang tinggi. Pengelompokan kabupaten dan kota di Jawa Timur berdasarkan indikator kemiskinan dengan 4 cluster dengan metode *K-Means Cluster* diperoleh cluster 1 terdapat 10 kabupaten dan kota, cluster 2 diperoleh 20 kabupaten dan kota, cluster 3 diperoleh 4 kabupaten dan kota sedangkan cluster 4 diperoleh 4 kabupaten. Hasil kesimpulan ini dapat diharapkan sebagai bahan pertimbangan dalam menangani daerah yang status kemiskinan masih tinggi.

UCAPAN TERIMA KASIH

Ucapan terimakasih kami haturkan kepada program studi statistika UNIPA Surabaya atas terlaksananya penelitian kolaboratif dosen dan mahasiswa.

DAFTAR RUJUKAN

- [1] D. W. W. Richard A. Johnson, *Applied Multivariate Statistical Analysis*. 1982.
- [2] G. Anuraga, "Hierarchical Clustering Multiscale Bootstrap untuk Pengelompokan Kemiskinan di Jawa Timur," *Statistika*, vol. 1, pp. 27–33, 2015.
- [3] Febriyana, *Analisis Kluster K-Means dan K-Median Pada Data Indikator Kemiskinan*. 2011.
- [4] T. Alfina, B. Santosa, and R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering , K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)," *J. Tek. ITS*, vol. 1, no. Data Mining, pp. 1–5, 2012.
- [5] BPS, "Data dan Informasi Kemiskinan Kabupaten Kota Tahun 2015," p. vi + 279, 2016.
- [6] Santoso Singgih. 2004 *Statistik Multivariat dengan SPSS*. (116)
- [7] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018.

- [8] F. A. Manurung and D. B. Santoso,
“Pemetaan Kemiskinan
Kabupaten/Kota di Provinsi Jawa
Timur,” J. Ilm. Mhs.FEB Univ.
Brawijaya, vol. 4,no. 1, pp. 1-
10,2015.