

Algoritma K-Means dalam mengelompokkan kota dan kabupaten di Jawa Timur berdasarkan indikator kesejahteraan rakyat

Isa Durrofiq^{a,1,*}, Rachmadania Akbarita^{a,2}

^a Universitas Nahdlatul Ulama Blitar, Indonesia;

¹ isarofiq77@gmail.com

*Correspondent Author

KATAKUNCI

Kemiskinan
Clustering
K-Means Cluster

KEYWORDS

Poverty
Clustering
K-Means Cluster

ABSTRAK

Kemiskinan merupakan permasalahan utama di suatu negara. Pemerintah Indonesia telah melakukan berbagai upaya untuk menanggulangi masalah kemiskinan. Berdasarkan data yang dirilis oleh BPS Provinsi Jawa Timur persentase penduduk miskin (penduduk yang berada dibawah Garis Kemiskinan) pada Desember 2021 sebesar 11,40%, angka ini melebihi persentase kemiskinan nasional dengan selisih 1,69%. Tingginya angka kemiskinan di Jawa Timur tersebut perlu dilakukan upaya-upaya pengentasan kemiskinan yang berbasis pada data dengan memahami karakteristik dari masing-masing wilayah, sehingga upaya pengentasan yang dilakukan bisa efektif dan tepat. Maka perlu dilakukan pengelompokan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan kemiripan karakteristik yang ditinjau dari indikator kesejahteraan rakyat yang meliputi kepadatan penduduk, pengangguran terbuka, pengeluaran perkapita disesuaikan, angka harapan hidup, harapan lama sekolah, jumlah presentase penduduk miskin dan PDRB. Berdasarkan hasil analisis diperoleh tiga klaster yang terdiri dari Kabupaten/Kota dengan persamaan karakteristik pada masing-masing klaster. Klaster 1 dengan tingkat kesejahteraan rakyat rendah berisi 3 kota dan kabupaten. Klaster 2 dengan tingkat kesejahteraan rakyat tinggi berisi 10 kota dan kabupaten. Klaster 3 dengan tingkat kesejahteraan rakyat sedang berisi 25 kota dan kabupaten

K-Means Algorithm in Grouping Cities and Regencies in East Java Based on People's Welfare Indicators

Poverty is a major problem in a country. The Indonesian government has made various efforts to tackle the problem of poverty. Based on data released by the BPS of East Java Province, the percentage of poor people (population below the Poverty Line) in December 2021 was 11.40%, this figure exceeded the national poverty percentage by a difference of 1.69%. The high poverty rate in East Java requires data-based poverty alleviation efforts by understanding the characteristics of each region, so that the alleviation efforts made can be effective and appropriate. So it is necessary to group districts/cities in East Java Province based on similar characteristics in terms of people's welfare indicators which include population density, open unemployment, adjusted per capita expenditure, life expectancy, school year expectations, percentage of the poor and GRDP. Based on the results of the analysis obtained three clusters consisting of districts/cities with similar characteristics in each cluster. Cluster 1 with a low level of people's welfare contains 3 cities and regencies. Cluster 2 with a high

level of people's welfare contains 10 cities and regencies. Cluster 3 with a moderate level of people's welfare contains 25 cities and regencies.

This is an open-access article under the [CC-BY-SA](#) license.



Pendahuluan

Setiap negara yang menjadi tujuan utama dalam pembangunan adalah peningkatan kesejahteraan rakyat. Seperti halnya di Indonesia, kesejahteraan rakyat adalah salah satu tujuan negara yang tertuang dalam Pembukaan UUD 1945 alinea IV. Kesejahteraan rakyat pada suatu wilayah merupakan suatu manifestasi yang diraih masyarakat dan diperoleh dari berbagai upaya, termasuk upaya dan kegiatan aktifitas ekonomi masyarakat tersebut. Dengan demikian pengentasan kemiskinan merupakan prioritas utama pembangunan kesejahteraan rakyat. Berdasarkan data yang dirilis oleh BPS Provinsi Jawa Timur persentase penduduk miskin (penduduk yang berada dibawah Garis Kemiskinan) pada Desember 2021 sebesar 11,40%, angka ini melebihi persentase kemiskinan nasional dengan selisih 1,69%. Tingginya angka kemiskinan di Jawa Timur tersebut perlu dilakukan upaya-upaya pengentasan kemiskinan yang berbasis pada data dengan memahami karakteristik dari masing-masing wilayah, sehingga upaya pengentasan yang dilakukan bisa efektif dan tepat. Maka perlu dilakukan identifikasi karakteristik dari masing-masing kabupaten/kota di Jawa Timur sebagai salah satu penunjang keberhasilan program pengentasan kemiskinan.

Aspek yang bisa diukur dalam indikator kesejahteraan rakyat diantaranya seperti kepadatan penduduk, pengangguran terbuka, pengeluaran perkapita disesuaikan, angka harapan hidup, harapan lama sekolah, jumlah presentase penduduk miskin dan PDRB. Menurut Soemartini (2017), dalam melaksanakan program pembangunan perlu adanya identifikasi berdasarkan karakteristik tingkat kesejahteraan rakyat tiap daerah agar dalam mengambil kebijakan dan strategi pembangunan bisa tepat sasaran dan tepat guna. Sebab salah satu prasyarat keberhasilan program- program pembangunan sangat tergantung pada ketepatan pengidentifikasian target group dan target area. Sebagai upaya untuk mencapai kondisi ideal diperlukan perencanaan dan aktivitas, selain oleh rakyat, peran pemerintah juga sangat dominan, baik menyangkut perencanaan, kegiatan dan bantuan biaya berupa permodalan. Secara umum kesejahteraan dapat diartikan sebagai suatu keadaan dimana segenap warga negara selalu berada dalam kondisi serba kecukupan segala kebutuhannya, baik material maupun spiritual. Indikator kesejahteraan rakyat menyajikan data tentang tingkat perkembangan kesejahteraan rakyat dari waktu ke waktu. Istilah kesejahteraan mencakup berbagai aspek kehidupan yang sangat luas dan tidak semuanya dapat diukur. Oleh karenanya, pengentasan kasus kemiskinan di Jawa Timur perlu adanya analisis untuk mengidentifikasi karakteristik berdasarkan indikator-indikator kesejahteraan rakyat dengan pengelompokan daerah. Salah satu metode yang dapat digunakan untuk mengelompokkan variabel atau objek adalah analisis kluster (cluster analysis) (soemartini, 2017).

Pada analisis kluster objek-objek dikelompokkan berdasarkan kemiripan/kesamaan objek. Prinsip dasar dalam analisis kluster adalah mengelompokkan objek pada suatu kluster (kelompok) yang memiliki kemiripan sangat besar dengan objek lain dalam kluster yang sama (similarity), tetapi sangat tidak mirip dengan objek lain pada kluster yang berbeda (dissimilarity). Hal ini bahwa kluster yang baik akan mempunyai homogenitas yang tinggi antar anggota dalam satu kelompok dan heterogenitas yang tinggi antar kluster yang satu dengan yang lainnya (Rachmatin, 2014). Secara umum metode dalam analisis kluster dibagi dua yakni Hierarchical Clustering (metode hirarki) dan Non-Hierarchical Clustering (metode tak hirarki) atau partitioning. Metode hirarki jumlah kelompok yang akan dibentuk belum

ditentukan. Prosedur kelompok hirarki terdiri atas dua bagian, yaitu metode (penggabungan) agglomerative dan metode (pemisah) divisive. Metode tak hirarki merupakan metode pengelompokan dimana kluster yang ingin dibentuk ditentukan terlebih dahulu, sehingga objek-objek akan dikelompokkan kedalam k kelompok yang telah ditentukan. Metode yang sering digunakan adalah K-Means. K-Means Cluster adalah salah satu metode dan clustering non-hierarki yang berusaha mengelompokkan data kedalam suatu kluster sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu kluster yang sama. Pemilihan metode K-Means karena metode ini efektif dan efisien. Ini dikarenakan K-means sangat mudah dipelajari dan dari segi waktu proses komputasinya relatif singkat.

Penelitian terdahulu yang menggunakan analisis kluster diantaranya (Wardono, 2019) menerapkan pengelompokan kabupaten/kota di provinsi Jawa Tengah berdasarkan indikator kesejahteraan dengan metode kmeans cluster. Pada penelitian tersebut menggunakan 6 variabel, yaitu Kepadatan Penduduk, Pengangguran Terbuka, Pengeluaran Perkapita disesuaikan, Angka Harapan Hidup, Rata-Rata Lama Sekolah dan Jumlah Presentase Penduduk Miskin. Sedangkan pada penelitian yang kali ini di Badan Pusat Statistik Kabupaten Blitar menggunakan 7 variabel, yaitu kepadatan penduduk, pengangguran terbuka, pengeluaran perkapita disesuaikan, angka harapan hidup, harapan lama sekolah, jumlah presentase penduduk miskin dan PDRB. Kemudian (Bobby, 2019) algoritma K-means dalam mengelompokkan kecamatan di tana luwu berdasarkan produktifitas hasil pertanian. Serta (Akramunnisa, 2020) menerapkan metode K-Means Clustering Analysis Pada Persebaran Tingkat Pengangguran Kabupaten/Kota Di Sulawesi Selatan. Berdasarkan uraian di atas, penelitian ini bertujuan untuk mengelompokkan kota dan kabupaten di Jawa Timur berdasarkan indikator kesejahteraan rakyat menggunakan metode K-means. Mengacu pada penelitian yang telah dilakukan sebelumnya, terdapat keterbaruan dalam penelitian ini yaitu ditambahkan beberapa indikator yang berpengaruh terhadap kesejahteraan rakyat.

Metode

Metode yang digunakan peneliti dalam melaksanakan penelitian di Kantor Badan Pusat Statistik Kabupaten Blitar adalah sebagai berikut:

1. Metode Studi Kepustakaan
Studi kepustakaan yaitu pengumpulan data dan informasi dari kepustakaan dengan membaca dan mempelajari buku-buku, literatur, artikel, serta bahan-bahan yang bersifat teoritis, pembelajaran yang didapat di perkuliahan ataupun umum, serta sumber informasi lainnya yang berhubungan dengan penelitian.
2. Metode Pengumpulan Data
Pengumpulan data dalam penelitian ini menggunakan data sekunder. Data sekunder diperoleh dari Website Badan Pusat Statistik Jawa Timur. Data yang dikumpulkan tersebut disusun untuk mendapatkan hasil data yang jelas dalam bentuk angka.
3. Metode Pengolahan Data
Metode pengolahan data menggunakan metode K-means untuk mengelompokkan kota dan kabupaten di Jawa Timur berdasarkan indikator kesejahteraan rakyat.

Hasil dan Pembahasan

1. Standarisasi Data

Variabel dalam data harus distandarasi agar terhindar dari masalah yang disebabkan oleh pemakaian nilai skala yang tidak seragam antar variabel pengelompokan objek. Standarisasi yang sering digunakan yaitu mengkonversi tiap variabel dengan nilai standar atau disebut dengan istilah z-score yakni menghitung nilai tengah serta membagi hasilnya dengan standar deviasi dari masing-masing variabel.

Dari data yang diperoleh pada Tabel 4.1 dilakukan standarisasi data dengan menggunakan

z-score. Hal ini perlu dilakukan karena satuan antar variabel yang memiliki jangkauan yang terlampaui jauh sehingga perlu dilakukan standarisasi pada setiap variabel.

2. Uji Multikolinearitas

Multikolinieritas merupakan suatu kondisi dimana ada korelasi antara variabel prediktor saat tipe regresi yang digunakan melebihi satu prediktor. Uji multikolinieritas bertujuan supaya mengerti apakah ada korelasi antara variabel independen dalam data yang digunakan. Dalam uji multikolinieritas ini salah satu yang peneliti gunakan untuk mengetahui terjadinya multikolinieritas yaitu koefisien korelasi pearson. Dalam penelitian ini uji multikolinieritas dilakukan dengan mencari nilai VIF menggunakan SPSS. Apabila nilai VIF lebih besar dari 10 menandakan terjadinya multikolinieritas antar variabel-variabel prediktor.

3. Interpretasi Analisis Kluster

Penelitian ini dilakukan untuk mengelompokkan kota dan kabupaten di Jawa Timur berdasarkan indikator kesejahteraan rakyat. Untuk itu agar dapat terbentuk kluster yang diinginkan akan diujikan memakai aplikasi SPSS dengan interpretasi analisis kluster menggunakan metode K-means Kluster sebagai berikut:

Table 1. *Initial Cluster Centers*

	<i>Initial Cluster Centers</i>		
	<i>1</i>	<i>2</i>	<i>3</i>
Zscore: Kepadatan Penduduk	.34523	2.68703	-.33094
Zscore: Pengangguran Terbuka	-.34810	2.06962	-1.20376
Zscore: Pengeluaran Perkapita	-.60803	2.83974	-1.24741
Zscore: Angka Harapan Hidup	.00107	1.24684	-2.04990
Zscore: Harapan Lama Sekolah	-.79734	1.56491	.27843
Zscore: Jumlah Persentase Penduduk Miskin	.41273	-1.29060	.84280
Zscore: PDRB	-4.44390	.90618	.29354

Pada tabel intial cluster centers ini merupakan proses pertama untuk pengelompokan data aplikasi SPSS. Karena pengelompokan data ini merupakan proses untuk pembentukan tiga kluster secara kasarnya.

Dapat diketahui nilai pusat cluster akhir dari proses clustering dimana data masih terdapat keterkaitan dengan standarisasi yang ditunjukkan dengan nilai z-score dengan makna:

Nilai Zscore positif (+) = data diatas rata-rata keseluruhan

Nilai Zscore negatif (-) = data dibawah rata-rata keseluruhan

1. Kluster 1 berisi nilai kepadatan penduduk (0.06356), pengangguran terbuka (-.03635), pengeluaran per kapita (-1.05609), angka harapan hidup (-.86827), harapan lama sekolah (-1.25944), jumlah presentase penduduk miskin(1.73967) dan PDRB (-3.00911). pada kluster 1 dapat dilihat pengeluaran perkapita dan PDRB memiliki nilia rata-rata yang lebih tinggi dibandingkan kluster 2 dan kluster 3. Namun kepadatan penduduk, angka harapan hidup, harapan lama sekolah dan jumlah presentase penduduk miskin memiliki nilia rata-rata yang lebih rendah dibandingkan kluster 2 dan kluster 3. Maka kluster 1 termasuk dalam kelompok kota dan kabupaten dengan kesejahteraan rakyat rendah.

2. Klaster 2 berisi nilai kepadatan penduduk (-.36142), pengangguran terbuka (1.18412), pengeluaran per kapita (1.30036), angka harapan hidup (.78297), harapan lama sekolah (1.22998), jumlah presentase penduduk miskin (-.99421) dan PDRB (.45092). Pada klaster 2 kepadatan penduduk, angka harapan hidup, harapan lama sekolah dan jumlah persentase penduduk miskin memiliki nilai rata-rata diatas klaster 1 dan klaster 3. Pengeluaran perkapita dan PDB memiliki nilai rata-rata diatas klaster 3 dan dibawah klaster 1. Namun hanya pengangguran terbuka yang memiliki nilai rata-rata dibawah klaster 1 dan klaster 3. Maka klaster 2 termasuk dalam kelompok kota dan kabupaten dengan kesejahteraan rakyat tinggi.
3. Klaster 3 berisi nilai kepadatan penduduk (.13694), pengangguran terbuka (-.46929), pengeluaran per kapita (-.46929), angka harapan hidup (-.20899), harapan lama sekolah (-.34086), jumlah presentase penduduk miskin (.18893) dan PDRB (.18072). Pada klaster 3 kepadatan penduduk, angka harapan hidup, harapan lama sekolah dan jumlah persentase penduduk miskin memiliki nilai rata-rata diatas klaster 1 dan dibawah klaster 2. Hanya pengangguran terbuka yang memiliki nilai rata-rata diatas klaster 1 dan klaster 2. Namun pengeluaran perkapita dan PDRB memiliki nilai rata-rata dibawah rata-rata. Maka klaster 3 termasuk dalam kelompok kota dan kabupaten dengan kesejahteraan rakyat sedang.

Simpulan

Berdasarkan hasil penelitian yang dilakukan dengan K-Means Klaster pada data indikator kesejahteraan rakyat kota dan kabupaten di Jawa Timur tahun 2021, maka diperoleh klaster 1 berisikan 3 kota dan kabupaten dengan dengan tingkat indikator kesejahteraan rakyat rendah, yaitu Kabupaten Bojonegoro, Kabupaten Bangkalan dan Kabupaten Sampang. Klaster 2 berisikan 10 kota dan kabupaten dengan dengan tingkat indikator kesejahteraan rakyat tinggi, yaitu Kabupaten Sidoarjo, Kabupaten Gresik, Kota Kediri, Kota Blita, Kota Malang, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya dan Kota Batu. Klaster 3 berisikan 25 kota dan kabupaten dengan dengan tingkat indikator kesejahteraan rakyat sedang, yaitu Kabupaten Pacitan, Kabupaten Ponorogo, Kabupaten Trenggalek, Kabupaten Tulungagung, Kabupaten Blita, Kabupaten Kediri, Kabupaten Malang, Kabupaten Lumajang, Kabupaten Jember, Kabupaten Banyuwangi, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan, Kabupaten Mojokerto, Kabupaten Jombang, Kabupaten Nganjuk, Kabupaten Madiun, Kabupaten Magetan, Kabupaten Ngawi, Kabupaten Tuban, Kabupaten Lamongan, Kabupaten Pamekasan, Kabupaten Sumenep dan Kota Probolinggo.

Daftar Pustaka

- [1] Alwi, W., & Hasrul, M. (2018). Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, 6(1), 35-35.
- [1] Badan Pusat Statistika Provinsi Jawa Timur, (2021): <https://jatim.bps.go.id/>
- [2] Fajriani, F. (2020). K-Means Clustering Analysis pada Persebaran Tingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan. *Jurnal Varian*, 3(2), 103-112.
- [3] Hasrul, M. (2018). Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat (Doctoral dissertation, Universitas Islam Negeri Alauddin Makassar).
- [4] Hidayat, R., Wasono, R., & Darsyah, M. Y. (2017). Pengelompokan Kabupaten/Kota di Jawa Tengah Menggunakan Metode K-Means dan Fuzzy C-Means. In *PROSIDING SEMINAR NASIONAL & INTERNASIONAL*.
- [5] Johnson, R.A dan Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. USA: Pearson Education, Inc.

-
- [6] Kristin Daya Rohani Sianipar, Indra Gunawan, (2021). Algoritma K-Means Dalam Pengelompokan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia Di Sumatera Utara, vol 6, No. 2.
- [7] M. Ivan Putra Eriansya, Muhammad Syafrullah, (2018). "Implementasi Algoritma ST-DBSCAN dan K-MEANS Untuk Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota Pulau Jawa Tahun 2014-2016 Berbasis Web Di Badan Pusat Statistik" vol 1, No. 3.
- [8] Prasetyo, Eko. (2012). Data Mining; Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta: Andi Offset.
- [9] Rachmatin, D dan Sawitri, K. (2014). Perbandingan Antara Metode Agglomeratif, Metode Divisif, dan Metode K-Means dalam Analisis Kluster. Seminar Nasional Matematika UNPAR. Soemartini dan Supartini, E. 2015. Prosiding Nasional Penelitian Matematika dan Pembelajarannya II (KNPMP II) UMS. Surakarta, 18 Maret 2018
- [10] Rindang Ndaru Puspita, (2021). "Analisis K-Means Cluster Pada Kabupaten/Kota Di Provinsi Banten Berdasarkan Indikator Indeks Pembangunan Manusia" vol 2, No. 3, Hal. 267-281.
- [11] Santoso, Singgih. (2014). Statistik Multivariat. Jakarta: PT Elex Media Komputtindo.
- [12] Sari, D. N. P., & Sukestiyarno, Y. L. (2021). Analisis cluster dengan metode K-Means pada persebaran kasus COVID-19 berdasarkan Provinsi di Indonesia. In PRISMA, Prosiding Seminar Nasional Matematika (Vol. 4, pp. 602-610).
- [13] Supranto, J. (2004). Analisis Multivariat: Arti dan Interpretasi. Jakarta: PT Rineka Cipta.
- [14] Talakua, M. W., Leleury, Z. A., & Taluta, A. W. (2017). Analisis cluster dengan menggunakan metode k-means untuk mengelompokkan Kabupaten/Kota di provinsi maluku berdasarkan indikator indeks pembangunan manusia tahun 2014. BAREKENG: Jurnal Ilmu Matematika dan Terapan, 11(2), 119-128.
- [15] Wardono, W., Sunarmi, S., & Wirawan, M. R. (2019). Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Dengan Metode Kmeans Cluster. EDUSAINTEK, 3.