

ANALISIS KLASTER DATA PENDIDIKAN KOTA SURABAYA TAHUN 2022-2023

Nanda Reza Handitia

Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: nandareza.21034@mhs.unesa.ac.id

A'yunin Sofro

Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya
Penulis Korespondensi: ayuninsofro@unesa.ac.id

Abstrak

Pendidikan merupakan kebutuhan esensial setiap individu bagi pengembangan sumber daya manusia suatu negara. Undang-undang pendidikan nasional menekankan pentingnya pemerataan akses, peningkatan kualitas, dan tata kelola pendidikan yang efisien demi menghadapi tantangan global. Namun, pemerataan pendidikan di Indonesia, khususnya melalui sistem zonasi masih menghadapi tantangan signifikan di kota-kota besar seperti Surabaya karena adanya ketidakmerataan akses dan fasilitas pendidikan. Studi ini membandingkan efektivitas tiga metode *clustering* non hierarki dalam analisis data pendidikan Kota Surabaya tahun 2022-2023. Data pada penelitian diperoleh dari publikasi terbaru Badan Pusat Statistik (BPS) Kota Surabaya berjudul "Kota Surabaya dalam Angka 2024". Data yang diambil mencakup informasi data pendidikan pada tahun 2022-2023 dari 31 kecamatan di Kota Surabaya terkait jumlah sekolah, jumlah murid, dan jumlah tenaga pendidik pada setiap jenjang pendidikan, yaitu SD/MI, SMP/MTS, dan SMA/SMK/MA. Hasil penelitian mendapatkan hasil bahwa metode *K-Means* memperoleh nilai rata-rata koefisien paling tinggi, dengan nilai rata-rata koefisien *silhouette* sebesar 0.592. Dengan demikian, metode *K-Means* memiliki ketepatan klaster paling optimal di antara metode lainnya. Temuan ini menyoroti perlunya perhatian lebih pada kecamatan dengan kondisi pendidikan rendah untuk memastikan pemerataan akses pendidikan dengan jumlah sekolah, jumlah murid, dan jumlah tenaga pendidik yang merata di seluruh wilayah Kota Surabaya.

Kata Kunci: pendidikan, *K-Means*, *K-Medoids*, *Fuzzy C-Means*

Abstract

Education is an essential need for every individual for the development of a country's human resources. The national education law emphasizes the importance of equal access, improving quality, and efficient education management to face global challenges. However, the equitable distribution of education in Indonesia, especially through the zoning system, still faces significant challenges in big cities such as Surabaya due to unequal distribution of educational access and facilities. This study compared the effectiveness of three methods of non-hierarchical clustering analysis of Surabaya City education data for 2022-2023. The data used was obtained from the latest publication of the Statistics of Surabaya Municipality (BPS) entitled "Surabaya Municipality in Figures 2024". The data includes information on education data for 2022-2023 from 31 sub-districts in Surabaya, regarding the quantity of schools, quantity of students, and quantity of teachers at each level of education, namely elementary school, junior high school, and senior high school. The results of this research indicate that the *K-Means* method has the highest average coefficient value, with the average *silhouette* coefficient amounting to 0.592. Thus, *K-Means* method has the most optimal cluster accuracy among other methods. These findings highlight the need for more attention to sub-districts with low educational conditions to ensure equal access to education, with an even distribution of schools, students, and the teacher throughout the city of Surabaya.)

Keywords: education, *K-Means*, *K-Medoids*, *Fuzzy C-Means*

PENDAHULUAN

Bagi suatu negara/bangsa, pendidikan merupakan kebutuhan bagi setiap manusia untuk melalui kehidupan serta menyiapkan kualitas sumber daya manusia (Kurniawan et al., 2021). Berdasarkan Undang-undang Nomor 20 Tahun 2003 tentang sistem pendidikan nasional menegaskan

bahwa pemerintah wajib memastikan pemerataan harapan memperoleh pendidikan, peningkatan kualitas, relevansi, dan efektivitas manajemen pendidikan untuk mengatasi tantangan terkait desakan peralihan kehidupan dalam tingkatan regional, nasional, dan internasional. Oleh karena itu, diperlukan inovasi pendidikan yang terkonsep, terpadu, dan berkelanjutan. Disamping itu, dalam

penyelenggaraan pendidikan yang efisien dan efektif dibutuhkan faktor pendukung dalam kegiatan belajar mengajar, yakni berupa penyediaan fasilitas belajar dan tenaga pendidik yang menguasai media pembelajaran (Sari, 2019). Namun, pemerataan pendidikan di Indonesia masih menghadapi berbagai tantangan. Salah satu isu yang terus dibahas adalah penerapan sistem zonasi yang diharapkan dapat memberikan keadilan dalam mendapatkan akses pendidikan kepada peserta didik, tetapi tidak diimbangi dengan pemerataan fasilitas pendidikan seperti jumlah sekolah, jumlah murid, dan jumlah tenaga pendidik di setiap kecamatan.

Pemerataan wilayah dalam memperoleh pendidikan belum terlaksana secara maksimal, oleh karena pemerintah harus memberikan perhatian khusus kepada wilayah yang dianggap memiliki akses pendidikan yang kurang. Dengan adanya klasterisasi tingkat pendidikan maka akan terlihat wilayah mana saja yang memiliki jumlah sekolah, jumlah murid, dan jumlah tenaga pendidik yang rendah atau tinggi. Klasterisasi tingkat pendidikan suatu wilayah dapat memudahkan pemerintah mempertimbangkan wilayah yang memerlukan perhatian lebih besar terkait pelaksanaan pendidikan sehingga tidak tertinggal oleh wilayah lainnya.

Salah satu metode dalam data *mining* dengan tujuan menentukan pola dari suatu data adalah analisis klaster, yakni dengan mengelompokkan objek ke dalam klaster yang mempunyai karakteristik sama. Ada beberapa metode dalam *clustering*, tetapi untuk analisis klaster dengan pendekatan non hierarki terdapat 3 metode, yakni *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*. Beberapa penelitian sebelumnya terkait perbandingan metode tersebut telah dilakukan, salah satunya terkait perbandingan algoritma *K-Means* dan *K-Medoids* untuk pengelompokan program BPJS Ketenagakerjaan (Meiriza et al., 2023). Ketiga metode tersebut juga dapat diterapkan pada bidang pendidikan seperti penelitian yang bertujuan untuk pengelompokan data guru di Indonesia (Idris et al., 2019) dan pemetaan penyebaran guru di Provinsi Banten (Priambodo & Prasetyo, 2018). Namun, penelitian-penelitian tersebut cenderung berfokus pada kondisi persebaran di tingkat provinsi dan kota/kabupaten, sementara belum banyak penelitian yang fokus pada satu kota. Dengan demikian, hasil pengklasteran dapat fokus pada kondisi dari setiap

kecamatan karena di sana dampak dari sistem zonasi paling terasa.

Pada penelitian ini akan dibandingkan hasil *cluster* menggunakan metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means* dalam menentukan metode mana yang paling efektif untuk pengelompokan data pendidikan di Surabaya. Dengan pengelompokan yang akurat mengenai tingkat pendidikan, diharapkan pemerintah dapat mengambil kebijakan yang lebih tepat untuk mengatasi ketimpangan pendidikan yang ada.

KAJIAN TEORI

PENDIDIKAN

Kondisi pendidikan saat ini belum mencapai standar yang dimandatkan oleh undang-undang. Pemerataan masih jauh dari kata sempurna dan isu ini tidak semata-mata terjadi di Indonesia, namun terjadi juga di berbagai negara lain (Lynch & Baker, 2005; Reay, 2020; Tchamyu, 2020). Ketidakmerataan pendidikan terdiri dari dua komponen, yakni mutu pendidikan (layanan pendidikan) dan kuantitas pendidikan (persebaran sekolah, pengaksesan, dan proporsi total sekolah dengan total penduduk) dan lain sebagainya (Hakim, 2016). Persebaran sekolah menjadi aspek penting dalam menjamin akses pendidikan yang adil dan merata di seluruh wilayah, peningkatan jumlah sekolah dan distribusi yang merata diharapkan dapat mengurangi kesenjangan akses pendidikan antara wilayah yang satu dengan yang lain. Sementara itu, distribusi siswa juga menjadi faktor penting dalam dinamika pendidikan, mempengaruhi kapasitas dan keefektifan pembelajaran di setiap sekolah.

Guru memegang kunci penting dalam proses menuntut ilmu di sekolah. Guru memiliki peranan penting dalam penyelenggaraan pendidikan yang adil dan merata, meski masalah kesenjangan dalam distribusi tersebut masih jauh untuk terselesaikan (Adamson & Darling-hammond, 2011). Pendistribusian guru yang tidak merata di Indonesia merupakan salah satu hambatan dalam upaya peningkatan kualitas pendidikan (Nurfatimah et al., 2022). Melalui Menteri Pendidikan dan Kebudayaan Nadiem Makarim, pemerintah telah berupaya membuat kebijakan baru untuk menyelesaikan persoalan distribusi guru, misalnya dengan program zonasi pendidikan yang juga turut menyasar

pemerataan guru profesional. Namun, nyatanya kajian konseptual menemukan tiga faktor utama yang menjadi tantangan dalam pendistribusian guru, yakni diantaranya manajemen pendidikan, faktor geografi, dan kendala pribadi (Haekal, 2022).

KAISER-MEYER-OLKIN (KMO)

Kaiser-Meyer-Olkin sering dimanfaatkan untuk menguji syarat kelayakan dari data. Jika didapatkan nilai *KMO* berada pada rentang 0,5 hingga 1 maka data dapat dianggap mewakili populasi atau data representatif. Untuk menghitung nilai *KMO* dapat digunakan rumus sebagai berikut (Widarjono, 2010):

$$KMO = \frac{\sum_{j=1}^p \sum_{k=1, k \neq j}^p r^2 X_j X_k}{\sum_{j=1}^p \sum_{k \neq j}^p r^2 X_j X_k + \sum_{j=1}^p \sum_{k \neq j}^p \rho^2 X_j X_k X_1} \quad (1)$$

dimana:

- p : banyaknya atribut
 $rX_j X_k$: korelasi antar atribut X_j dan X_k
 \bar{X}_j : rata-rata atribut X_j
 \bar{X}_k : rata-rata atribut X_k
 N : banyaknya observasi (objek)
 $\rho X_j X_k, X_1$: korelasi parsial antara atribut X_j dan X_k dengan X_1

BARTLETT TEST OF SPHERICITY

Bartlett Test of Sphericity bertujuan untuk menganalisis ada atau tidaknya korelasi antar atribut dalam kasus multivariat agar analisis dapat dilanjutkan (Putri & Fithriasari, 2015). Matriks korelasi dikatakan matriks identitas jika atribut bersifat independen atau saling bebas yang menandakan bahwa di antara peubah tidak terdapat korelasi. Hipotesis percobaan yang dipakai pada *Bartlett Test* adalah sebagai berikut:

H_0 : Matriks korelasi sama dengan matriks identitas (tidak ada multikolinearitas)

H_1 : Matriks korelasi tidak sama dengan matriks identitas

Nilai statistik *Bartlett Test* dapat ditulis dengan persamaan berikut:

$$X_{obs}^2 = - \left[(N - 1) \frac{(2p+5)}{6} \ln |R| \right] \quad (2)$$

dimana:

- N : jumlah pengamatan
 $|R|$: determinan matriks pengamatan
 p : banyaknya atribut

Dalam pengambilan keputusan *Bartlett Test* jika $X_{obs}^2 > X_{\alpha, \frac{p(p-1)}{2}}$ atau p -value $< \alpha$ yang berarti bahwa terjadi multikolinearitas antar atribut.

NORMALISASI DATA

Normalisasi merupakan metode penskalaan nilai atribut dari data untuk dapat berada pada rentang tertentu (Maria Kristilia, 2014). Transformasi data dengan normalisasi dapat diselesaikan dengan berbagai metode, yaitu *z-score normalization*, *decimal scaling*, *min-max normalization*, *softmax*, dan *sigmoid*. Normalisasi dengan *Min-Max* adalah dengan menggunakan transformasi linier pada data asli untuk memperoleh nilai perbandingan antara data sebelum dan sesudah proses yang seimbangan. Metode normalisasi *Min-Max* adalah salah satu teknik untuk mengubah data yang kompleks tanpa menghilangkan isi, sehingga data lebih mudah diolah. Metode ini melalui cara standarisasi data, yaitu meletakkan data dalam rentang 0 hingga 1, dimana nilai terkecil adalah 0, dan nilai terbesar adalah 1. Metode ini menghasilkan proporsi data yang seimbang antara satu dengan yang lain (Tasmalaila Hanifa et al., 2017). Metode normalisasi *min-max* dapat menggunakan rumus sebagai berikut:

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{range.max} - \text{range.min}) + \text{range.min} \quad (3)$$

v' merupakan nilai dari data sesudah dilakukan normalisasi, v adalah data sesudah normalisasi, \min_A merupakan nilai minimal pada atribut sebelum normalisasi, \max_A adalah nilai maksimal pada atribut sebelum normalisasi, *range min* adalah 0, dan *range max* adalah 1.

ELBOW METHOD

Elbow method adalah salah satu metode untuk memutuskan total kluster optimal melalui cara melihat persentase hasil membandingkan total kluster yang membentuk siku pada suatu titik (Madhulata, 2012). Metode *elbow* memberikan gagasan melalui cara memilih nilai kluster lalu menambahnya untuk membuat model data dalam menentukan kluster optimal. Hasil persentase setiap nilai kluster ditampilkan dalam bentuk grafik sebagai sumber informasi (Merliana et al., 2019). Dikatakan nilai kluster tersebut paling baik jika pada grafik membentuk siku atau mengalami penurunan yang paling besar (Bholowalia & Kumar, 2014). Dalam menentukan nilai perbandingannya maka dilakukan perhitungan *Sum of Square Error (SSE)* dari setiap nilai kluster, nilai *SSE* merupakan jumlah rata-rata jarak *euclidean* dari setiap titik ke *centroid*. Untuk

mendapatkan nilai SSE digunakan persamaan sebagai berikut (Merliana et al., 2019):

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - c_k\|^2 \quad (4)$$

dimana:

- K : banyak kluster
- x_i : data ke- i
- S_k : himpunan dari elemen kluster k
- c_k : rata-rata (pusat) dalam kluster k

K-MEANS

K-Means adalah salah satu algoritma dalam analisis kluster dengan pendekatan non hierarki. Algoritma ini mempartisi data ke dalam kluster sehingga data dengan karakteristik sejenis dikelompokkan dalam satu kluster, sebaliknya data dengan karakteristik berbeda ditempatkan pada kluster lain. Secara sederhana, proses algoritma *K-Means* dibagi menjadi 5 langkah berikut (Maryani et al., 2018):

1. Menentukan nilai kluster (k) sesuai keinginan
2. Tentukan juga nilai pusat (*centroid*) awal dari setiap kluster dengan nilai sembarang
3. Mengalokasikan seluruh objek pada kluster dengan jarak terdekat terhadap tiap *centroid* menggunakan teori jarak *euclidean* menggunakan rumus sebagai berikut:

$$D(a, b) = \sqrt{\sum_{k=1}^n (a_k - b_k)^2} \quad (5)$$

dimana:

- n : banyak atribut
 - $a_k - b_k$: data dengan atribut k
4. Perbarui nilai *centroid* melalui perhitungan rata-rata seluruh data pada setiap kluster
 5. Dilakukan perulangan langkah (3) dan (4) sampai tidak terjadi pergantian pada *centroid* dan keseluruhan objek tidak dapat diklasifikasikan lagi.

K-MEDOIDS

K-Medoids merupakan proses *clustering* untuk mengelompokkan kumpulan objek pada kluster dan memiliki keserupaan dengan algoritma *K-Means*, yaitu termasuk ke dalam pengelompokan data tanpa adanya struktur hierarki. Algoritma *K-Medoids* memakai *medoid* untuk pusat kluster, yaitu menggunakan salah satu data yang ada dengan inisialisasi secara acak untuk mewakili kluster (Meiriza et al., 2023). Langkah-langkah algoritma *K-Medoids* dibagi menjadi 6 langkah berikut:

1. Menentukan nilai kluster (k) sesuai keinginan

2. Mengalokasikan semua objek pada kluster dengan jarak terdekat memakai teori jarak *euclidean* dan menggunakan rumus sebagai berikut:

$$D(a, b) = \sqrt{\sum_{k=1}^n (a_k - b_k)^2} \quad (6)$$

dimana:

- n : banyak atribut
 - $a_k - b_k$: data dengan atribut k
3. Memilih sembarang objek pada masing-masing kluster sebagai bakal *medoid* baru
 4. Menghitung jarak setiap objek pada masing-masing kluster dengan kandidat *medoid* baru.
 5. Menentukan total simpangan (S) dengan cara menghitung nilai total jarak baru dikurangi total jarak sebelumnya. Apabila $S < 0$, maka harus menukar objek dengan n data kluster sehingga terbentuk membentuk sekumpulan k objek.
 6. Dilakukan perulangan untuk langkah (3) hingga (5) sampai *medoid* tidak berubah lagi, sehingga dihasilkan kluster dan anggotanya baru sebagai *medoid*.

FUZZY C-MEANS

Metode *Fuzzy C-Means* merupakan proses mengelompokkan data berdasarkan derajat keanggotaan. Selama proses perhitungan, nilai derajat keanggotaan setiap objek akan selalu dihitung dan digunakan untuk menetapkan objek ke dalam satu atau lebih kluster. (Nithila & Kumar, 2016). Data dikelompokkan berdasarkan derajat keanggotaannya, yakni dalam rentang 0 hingga 1, dan ada beberapa jenis data hanya menampilkan keanggotaan sebagian. Algoritma *Fuzzy C-Means* bisa menghasilkan pengelompokan untuk objek tidak teratur dan tersebar. Secara sederhana, proses algoritma *Fuzzy C-Means* dibagi menjadi 7 langkah (Andriyani et al., 2013) sebagai berikut:

1. Memasukkan data yang akan dikelompokkan berupa matriks X berukuran $n \times m$ (n = banyaknya sampel data dan m = banyaknya variabel setiap data). X_{ij} = data sampel ke- i ($i = 1, 2, \dots, n$), atribut ke- j ($j = 1, 2, \dots, m$)
2. Menentukan nilai-nilai yang diperlukan sebelum melakukan perhitungan, yakni sebagai berikut:
 - Banyaknya kluster yang akan dibentuk (c)
 - Pangkat pembobot (w)
 - Maksimum iterasi (*MaxIter*).
 - Error terkecil (ξ)

- Fungsi obyektif awal ($P_0 = 0$)
 - Iterasi awal ($t = 1$)
3. Membangkitkan matriks partisi awal $U_{n \times c} = [\mu_{ik}]$, μ_{ik} merupakan bilangan *random* yang menunjukkan derajat keanggotaan
 4. Menghitung pusat kluster ke- k (V_{kj}) menggunakan $k = 1, 2, \dots, c$ dan $j=1, 2, \dots, m$ dengan rumus sebagai berikut:

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w \cdot X_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \quad (7)$$

5. Mencari nilai fungsi objektif untuk iterasi ke- t , dimana P_t menyatakan jumlah jarak data ke pusat kluster yang dapat dihitung dengan rumus sebagai berikut:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (8)$$

dimana:

- P_t : fungsi objektif
- X_{ij} : elemen X baris i , kolom j
- V_{kj} : pusat kluster

6. Mencari nilai perubahan matriks partisi melalui rumus sebagai berikut:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}} \quad (9)$$

dimana:

- i : 1, 2, ..., n
- k : 1, 2, ..., c
- X_{ij} : sampel data ke- i , atribut ke- j
- V_{kj} : pusat kluster ke- k untuk atribut ke- j
- w : pangkat pembobotan

7. Melakukan pemeriksaan kondisi untuk tetap melakukan iterasi atau stop, dengan pemutusan keputusan sebagai berikut:
 - Jika ($|P_t - P_{t-1}| < \xi$) atau ($t > \text{MaxIter}$) maka stop
 - Jika tidak maka $t = t + 1$, lalu dilakukan perulangan untuk langkah (4) hingga (7).

SILHOUETTE COEFFICIENT

Silhouette coefficient berfungsi dalam menentukan seberapa bagus objek diletakkan dalam suatu kluster setelah proses *clustering*, dengan kata lain berupa nilai kualitas dan kekuatan dari kluster. Metode *Silhouette coefficient* adalah perpaduan dari 2 metode, yakni *cohesion* dan *separation* (Dewi & Pramita, 2019). Langkah-langkah dalam perhitungan *silhouette coefficient* dibagi menjadi 3 langkah sebagai berikut (Mario et al., 2016):

1. Menghitung rata-rata jarak dari dokumen, misalkan i dengan semua dokumen berada dalam satu kluster, digunakan rumus sebagai berikut:

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (10)$$

dimana j merupakan dokumen lain dalam satu kluster A dan $d(i, j)$ adalah jarak antara dokumen i dengan j

2. Menghitung rata-rata jarak dari dokumen i tersebut dengan semua dokumen di kluster lain, kemudian diambil nilai terkecilnya melalui rumus sebagai berikut:

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (11)$$

dimana $d(i, C)$ adalah jarak rata-rata dokumen i dengan seluruh objek pada kluster lain yang mana $A \neq C$, lalu $b(i) = \min C \neq A d(i, j)$.

3. Menentukan nilai koefisien *silhouette* dengan rumus sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (12)$$

Berikut ini adalah kriteria pengukuran nilai *silhouette coefficient* (Rousseeuw, 1987). Kriteria *Silhouette Coefficient* disajikan pada Tabel sebagai berikut:

Tabel 1. Kriteria *Silhouette Coefficient*

Koefisien <i>Silhouette</i>	Kriteria Kluster
0,71 - 1,00	Strong
0,51 - 0,70	Good
0,26 - 0,50	Weak
0,00 - 0,25	Bad

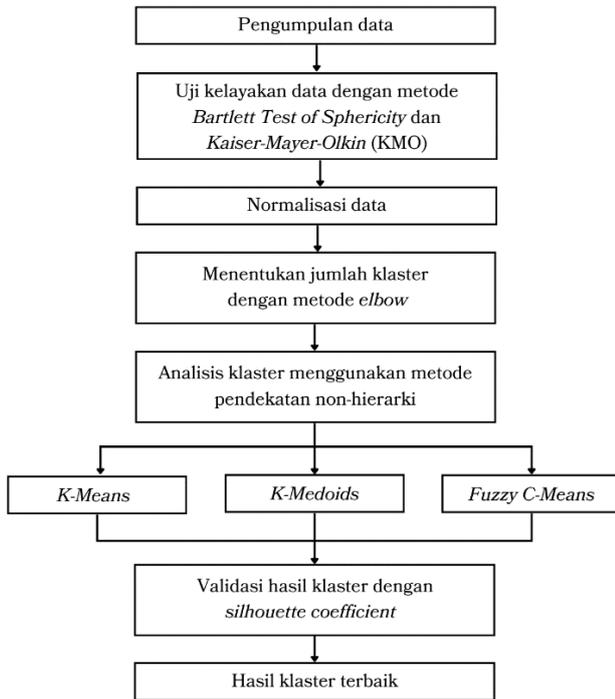
METODE

DATA PENELITIAN

Pada penelitian ini jenis data yang digunakan adalah data sekunder yang diperoleh dari publikasi terbaru Badan Pusat Statistik (BPS) Kota Surabaya berjudul “Kota Surabaya dalam Angka 2024”. Data yang diambil mencakup informasi data pendidikan pada tahun 2022-2023 dari 31 kecamatan di Kota Surabaya terkait jumlah sekolah, jumlah murid, dan jumlah guru pada masing-masing jenjang pendidikan, yaitu SD/MI, SMP/MTS, dan SMA/SMK/MA.

DIAGRAM ALIR PENELITIAN

Diagram alir berikut ini menunjukkan langkah-langkah yang akan dilakukan dalam penelitian:



Gambar 1. Diagram Alir Penelitian

Berdasarkan Gambar 1, tahapan awal pada penelitian ini yaitu pengumpulan data pendidikan pada tahun 2022-2023 dari 31 kecamatan di Kota Surabaya terkait jumlah sekolah, jumlah murid, dan jumlah tenaga pendidik pada masing-masing jenjang pendidikan, yaitu SD/MI, SMP/MTS, dan SMA/SMK/MA. Tahapan selanjutnya yaitu melakukan uji kelayakan data dengan menggunakan dua metode, yaitu *Bartlett Test of Sphericity* dan *Kaiser-Meyer-Olkin (KMO)*. Sesudah itu, dilakukan normalisasi data untuk memastikan bahwa setiap atribut memiliki skala yang sama dan tidak ada atribut yang mendominasi hasil akhir. Langkah selanjutnya menentukan jumlah kluster optimal menggunakan metode *elbow*. Kemudian dilakukan analisis kluster menggunakan metode pendekatan non hirarki, yaitu *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*. Setelah memperoleh hasil analisis kluster, dilakukanlah validasi kluster menggunakan *silhouette coefficient* untuk mengetahui metode terbaik dengan hasil kluster paling optimal.

HASIL DAN PEMBAHASAN

Pada bagian ini merupakan pembahasan mengenai analisis kluster 31 kecamatan di Surabaya berdasarkan jumlah sekolah, jumlah murid, dan

jumlah tenaga pendidik tahun 2022-2023 pada masing-masing jenjang pendidikan. Berdasarkan data yang akan dianalisis didapatkan bahwa nilai minimum pada jenjang SD/MI untuk total sekolah adalah 12 (Kecamatan Jambangan), selanjutnya untuk minimum pada total murid dan guru berturut-turut adalah 3768 dan 200 (Kecamatan Asemrowo). Sementara itu, pada jenjang SMP/MTS nilai minimum dari total sekolah, murid, dan guru berturut-turut adalah 5, 1389, dan 71 (Kecamatan Pabean Cantian). Nilai minimum pada jenjang SMA/SMK/MA untuk total sekolah adalah 1 (Kecamatan Tegalsari), sedangkan untuk nilai minimum pada total murid dan guru berturut-turut adalah 512 dan 28 (Kecamatan Gununganyar).

UJI KELAYAKAN DATA

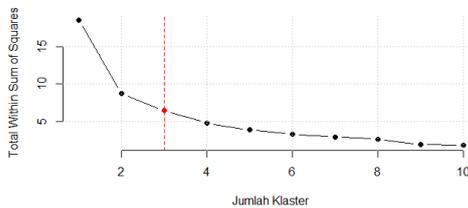
Data pendidikan yang digunakan perlu dilakukan uji kelayakan untuk memastikan apakah terdapat pengaruh signifikan secara keseluruhan terhadap model. Uji kelayakan atribut ini menggunakan metode *Bartlett Test of Sphericity* dan *Kaiser-Meyer-Olkin (KMO)*. Uji *Kaiser-Meyer-Olkin (KMO)* berguna dalam penentuan *appropriateness* dari suatu analisis faktor. Rentang dalam nilai uji KMO adalah antara 0 hingga 1. Jika nilai KMO tidak lebih dari 0.5, maka analisis faktor tidak layak untuk dilanjutkan. Sebaliknya, apabila nilai KMO lebih dari 0.5, analisis faktor layak untuk dilanjutkan. Berdasarkan perhitungan dihasilkan bahwa nilai KMO sebesar 0.68 dan terletak di antara 0 dan 1, sehingga proses analisis faktor dapat dilakukan. Selain itu, hasil analisis menggunakan *Bartlett's test of sphericity* menghasilkan nilai signifikansinya mencapai 0.000 atau tidak lebih dari 0.05. Dengan demikian, atribut-atribut yang dipakai berkorelasi dan proses perhitungan dapat dilanjutkan.

NORMALISASI DATA

Data pendidikan yang layak dipakai tidak dapat langsung diolah karena adanya perbedaan besaran angka yang cukup jauh. Normalisasi data merupakan langkah penting dalam proses pengelompokan untuk memastikan bahwa setiap atribut memiliki skala yang sama dan tidak ada atribut yang mendominasi hasil akhir hanya karena skala yang berbeda.

PENENTUAN JUMLAH KLASTER

Untuk menentukan jumlah k (klaster) optimal pada data akan digunakan metode *elbow*. Dengan menerapkan metode *elbow* pada data yang telah dilakukan normalisasi didapatkan grafik seperti berikut:



Gambar 2. Grafik Metode *Elbow*

Berdasarkan Gambar 2 terjadi pergerakan grafik yang landai setelah pergerakan grafik yang curam atau berbentuk siku setelah titik 3. Dengan demikian, pada metode *elbow* didapatkan nilai k optimal adalah $k=3$.

CLUSTERING DENGAN K-MEANS

Pengujian pertama dilakukan dengan metode *K-Means* menggunakan k optimal yang sudah didapatkan sebelumnya, yakni sejumlah tiga klaster yang kemudian dilakukan *clustering* berdasarkan posisi *centroid* terdekat. Proses iterasi berhenti ketika tidak terjadi perubahan pada *centroid* dan keseluruhan objek tidak dapat diklasifikasikan lagi, yaitu ketika iterasi ketujuh dan didapatkan hasil sebagai berikut:

Tabel 2. Hasil *Clustering K-Means*

Klaster	Nama Kecamatan
Klaster 1	Karangpilang, Gayungan, Tenggiling Mejoyo, Gunung Anyar, Wiyung Tegalsari, Bulak, Pabean Cantian, Bubutan, Asemrowo, Benowo, dan Pakal.
Klaster 2	Wonocolo, Sukolilo, Mulyorejo, Gubeng, Wonokromo, Sukomanunggal, Sawahan, Genteng, Tambaksari, Kenjeran, Semampir, dan Krembangan.
Klaster 3	Jambangan, Rungkut, Dukuh Pakis, Lakarsantri, Sambikerep, Tandes, dan Simokerto.

Dengan demikian, pada klaster 1 didapati sebanyak 12 kecamatan, klaster 2 terdiri dari 12 kecamatan, dan klaster 3 memiliki anggota sebanyak 7 kecamatan.

CLUSTERING DENGAN K-MEDOIDS

Pengujian selanjutnya dilakukan dengan metode *K-Medoids* menggunakan k optimal yang sudah didapatkan sebelumnya, yakni sejumlah tiga klaster dan dilakukan *clustering* berdasarkan *medoid* terdekat yang dipilih ulang untuk setiap iterasi. Proses iterasi berhenti ketika setiap *medoid* baru yang dipilih tetap sebagai *medoid* setelah pertukaran titik, yaitu ketika iterasi keempat dan didapatkan hasil sebagai berikut:

Tabel 3. Hasil *Clustering K-Medoids*

Klaster	Nama Kecamatan
Klaster 1	Karangpilang, Jambangan, Gayungan, Tenggiling Mejoyo, Gunung Anyar, Dukuh Pakis, Wiyung, Bulak, Simokerto, Pabean Cantian, Bubutan, Asemrowo, Benowo, dan Pakal.
Klaster 2	Wonocolo, Sukolilo, Mulyorejo, Gubeng, Wonokromo, Sambikerep, Sukomanunggal, Genteng, Tambaksari, Semampir, dan Krembangan.
Klaster 3	Rungkut, Lakarsantri, Tandes, Sawahan, Tegalsari, dan Kenjeran

Dengan demikian, pada klaster 1 didapati sebanyak 14 kecamatan, klaster 2 terdiri dari 11 kecamatan, dan klaster 3 memiliki anggota sebanyak 6 kecamatan.

CLUSTERING DENGAN FUZZY C-MEANS

Pengujian dengan metode *Fuzzy C-Means* menggunakan $k=3$, dilakukan *clustering* dengan memperbarui derajat keanggotaan fuzzy dan posisi *centroid* secara iteratif. Proses iterasi berhenti ketika perubahan nilai objektif lebih kecil dari nilai *error* (ξ) yang ditentukan di awal atau iterasi telah mencapai batas *MaxIter* yang ditentukan di awal juga. Pada perhitungan ini, proses iterasi berhenti saat iterasi ke-15 ketika nilai *error* (ξ) < 0.0001 dan didapatkan hasil sebagai berikut:

Tabel 4. Hasil *Clustering Fuzzy C-Means*

Klaster	Nama Kecamatan
Klaster 1	Wonocolo, Rungkut, Mulyorejo, Gubeng, Lakarsantri, Sambikerep, Tandes, Sukomanunggal, Sawahan, Semampir, dan Krembangan
Klaster 2	Karangpilang, Jambangan, Gayungan, Tenggiling Mejoyo, Gunung Anyar, Dukuh Pakis, Wiyung, Tegalsari,

	Genteng, Bulak, Simokerto, Pabean Cantian, Bubutan, Asemrowo, Benowo, dan Pakal
Klaster 3	Sukolilo, Wonokromo, Tambaksari, dan Kenjeran

Dengan demikian, pada klaster 1 didapati sebanyak 11 kecamatan, klaster 2 terdiri dari 16 kecamatan, dan klaster 3 memiliki anggota sebanyak 4 kecamatan.

UJI VALIDITAS KLASTER

Sesudah melakukan pengklasteran, pada tahapan ini dilakukan validasi klaster menggunakan koefisien *silhouette*. Penggunaan koefisien *silhouette* berguna dalam melakukan validasi hasil klaster dari masing masing metode untuk menemukan klaster paling optimal sebagai langkah pemecahan masalah persebaran pendidikan di Kota Surabaya. Hasil nilai koefisien *silhouette* dengan bantuan R studio dihasilkan sebagai berikut:

Tabel 5. Hasil Nilai Koefisien *Silhouette*

Metode	Nilai Rata-rata Koefisien <i>Silhouette</i>	Interpretasi
K-Means	0.592	<i>Good</i>
K-Medoids	0.257	<i>Weak</i>
Fuzzy C-Means	0.338	<i>Weak</i>

Dari nilai rata-rata koefisien *silhouette* tiga metode yang digunakan, diperoleh nilai rata-rata 0,592 untuk metode *K-Means* diartikan ke dalam kriteria *Good Classification* dan nilai rata-rata 0,257 untuk metode *K-Medoids* diartikan ke dalam kriteria *Weak Classification*. Untuk *Fuzzy C-Means* memiliki nilai rata-rata 0,338 diartikan ke dalam kriteria *Weak Classification*. Dengan demikian, hasil klaster paling optimal dimiliki oleh metode *K-Means*.

PENUTUP

SIMPULAN

Dalam penelitian ini, metode terbaik yang diperoleh dari ketiga metode yang telah digunakan adalah *K-Means* dengan nilai rata-rata koefisien *silhouette* sebesar 0.592. Berdasarkan hasil analisis persebaran pendidikan di kota Surabaya menggunakan metode *K-Means* diperoleh sebanyak 12 kecamatan berada pada klaster 1, yakni Karangpilang, Gayungan, Tenggilis Mejoyo, Gunung

Anyar, Wiyung, Tegalsari, Bulak, Pabean Cantian, Bubutan, Asemrowo, Benowo, dan Pakal. Kecamatan tersebut menunjukkan kondisi pendidikan yang rendah. Oleh karena itu, kecamatan-kecamatan tersebut memerlukan perhatian khusus pemerintah kota Surabaya untuk mengatasi kurangnya jumlah sekolah, jumlah murid, dan jumlah tenaga pendidik. Setelah menemukan hasil pengelompokan untuk persebaran pendidikan di Surabaya maka diharapkan nantinya dapat menjadi langkah yang tepat untuk mengatasi ketidakmerataan pendidikan di kota Surabaya.

SARAN

Untuk penelitian selanjutnya disarankan menggunakan data yang lebih luas, termasuk atribut sosial-ekonomi dan kualitas pendidikan. Hal ini dapat memberikan gambaran yang lebih komprehensif dan mendalam mengenai faktor-faktor yang mempengaruhi ketidakmerataan pendidikan. Disarankan pula untuk menerapkan metode *clustering* lain seperti *DBSCAN* atau *Hierarchical Clustering* untuk melakukan perbandingan dan bisa menemukan metode yang lebih baik lagi.

DAFTAR PUSTAKA

- Adamson, F., & Darling-hammond, L. (2011). Addressing the Inequitable Distribution of Teachers: What It Will Take to Get Qualified, Effective Teachers in All Communities. *Stanford Center for Opportunity Policy in Education*, 2(December), 171-192.
- Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications*, 105(9), 975-8887.
- Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi Dan Informatika*, 9(3), 102-109. <https://doi.org/10.31940/matrix.v9i3.1662>
- Haekal, M. (2022). Tantangan Distribusi Guru di Daerah Terpencil Indonesia: Antara Manajemen, Isu Personal, dan Faktor Geografi. *TA'DIB: Jurnal Pemikiran Pendidikan*, 12(1), 18-23. <https://doi.org/10.54604/tdb.v12i1.100>
- Hakim, L. (2016). Pemerataan Akses Pendidikan bagi Rakyat Sesuai dengan Amanat Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional. *EduTech: Jurnal Ilmu Pendidikan Dan Ilmu Sosial*, 2(1), 53-64.

- Handoyono, R., M. R. R., & Nasution, S. M. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means pada Pengelompokan Dokumen. *JSM STMik Mikroskil*, 15(2), 73–82.
- Idris, F., Azmi, F., & Kusuma, D. P. (2019). Pengelompokan Data Guru Di Indonesia Menggunakan K-Means Clustering. *E-Proceeding of Engineering*, 6(2), 5643–5658.
- Kurniawan, R., Mukarobin, M., & Mahradianur. (2021). Klasterisasi Tingkat Pendidikan di DKI Jakarta pada Tingkat Kecamatan Menggunakan Algoritma K-Means. *Technologia: Jurnal Ilmiah*, 12(4), 234. <https://doi.org/10.31602/tji.v12i4.5633>
- Lynch, K., & Baker, J. (2005). Equality in education: An equality of condition perspective. In *Theory and Research in Education* (Vol. 3, Issue 2). <https://doi.org/10.1177/1477878505053298>
- Madhulata, T. S. (2012). An Overview of Clustering Methods. *IOSR Journal of Engineering*, 2(4), 719–725. <https://doi.org/10.3233/ida-2007-11602>
- Maria Kristilia. (2014). Deteksi Outlier pada Data Campuran Numerik Dan Kategorikal Menggunakan Algoritma Enhanced Class Outlier Distance Based (Ecodb) (Studi Kasus : Data Kredit BPS XYZ}. Fakultas Sains Dan Teknologi, Universitas Sanata Dharma, 1.
- Mario, A., Herry, S., & Nasution, H. (2016). Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness. *Jurnal Sistem Dan Teknologi Informasi*, 1(1), 1–6.
- Maryani, I., Riana, D., Astuti, R. D., Ishaq, A., Sutrisno, & Pratama, E. A. (2018). Customer segmentation based on RFM model and clustering techniques with K-means algorithm. *Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018*, 1–6. <https://doi.org/10.1109/IAC.2018.8780570>
- Meiriza, A., Ali, E., Rahmiati, & Agustin. (2023). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan. *Indonesian Journal of Computer Science*, 12(2), 714–728. <https://doi.org/10.33022/ijcs.v12i2.3184>
- Merliana, N. P. E., Ernawati, & Santoso, A. J. (2019). ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-MEANS CLUSTERING. *Matrix: Jurnal Manajemen Teknologi Dan Informatika*, 102–109. <https://www.unisbank.ac.id/ojs/index.php/sendu/article/view/3333>
- Nurfatimah, S. A., Hasna, S., & Rostika, D. (2022). Membangun Kualitas Pendidikan di Indonesia dalam Mewujudkan Program Sustainable Development Goals (SDGs). *Jurnal Basicedu*, 6(4), 6145–6154. <https://doi.org/10.31004/basicedu.v6i4.3183>
- Priambodo, Y. A., & Prasetyo, S. Y. J. (2018). Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Metode Spatial Clustering K-Means (Studi kasus : Wilayah Provinsi Banten). *Indonesian Journal of Computing and Modeling*, 1(1), 18–27. <https://doi.org/10.24246/j.icm.2018.v1.i1.p18-27>
- Putri, M. M., & Fithriasari, K. (2015). Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kesehatan Masyarakat Menggunakan Metode Kohonen SOM dan K-Means. *Jurnal Sains Dan Seni ITS*, 4(1), 13–18. [10.12962/j23373520.v4i1.8815](https://doi.org/10.12962/j23373520.v4i1.8815)
- Rachmatin, D. (2014). Aplikasi Metode-Metode Agglomerative dalam Analisis Kluster pada Data Tingkat Polusi Udara. *Infinity Journal*, 3(2), 133. <https://doi.org/10.22460/infinity.v3i2.59>
- Reay, D. (2020). Miseducation: Inequality, Education, and the Working Classes. *Journal of Teaching and Learning*, 14(2), 64–66. <https://doi.org/10.26522/ssj.v13i2.2227>
- Sari, P. I. (2019). Peran Pendidik dalam Implementasi Media Pembelajaran terhadap Peserta Didik Generasi 4.0. *Prosiding Seminar Nasional Pendidikan FKIP*, 2(1), 508–517.
- Tasmalaila Hanifa, T., Al-Faraby, S., & Adiwijaya. (2017). Analisis Churn Prediction pada Data Pelanggan PT. Telekomunikasi dengan Logistic Regression dan Underbagging. *E-Proceeding of Engineering*, 4(2), 3210–3225.
- Tchamyoyou, V. S. (2020). Education, lifelong learning, inequality and financial access: evidence from African countries. *Contemporary Social Science*, 15(1), 7–25. <https://doi.org/10.1080/21582041.2018.1433314>
- Widarjono, A. 2010. Analisis Statistika Multivariat Terapan. Edisi pertama. Yogyakarta: UPP STIM YKPN.