

**ANALISIS K-MEAN CLUSTERING DALAM PENGELOMPOKKAN ANGKA
KEMATIAN ANAK DI INDONESIA****ANALYSIS OF K-MEAN CLUSTERING IN GROUPING CHILD MORTALITY
RATE IN INDONESIA****Eka Sabna¹, Mitra²**Universitas Hang Tuah Pekanbaru^{1,2}Email : es3jelita@yahoo.com**Abstract**

The death rate is an indicator of health status in society. One of these indicators is the Child Mortality Rate (CMR), which reflects environmental health conditions which directly influence children's health levels. The K-Means algorithm divides data into 3 groups (Clusters), namely Clusters 0, 1 and 2 to produce information on child mortality rates in all districts/cities in Indonesia. This modeling produces 3 clusters, namely cluster 0 (Districts/Cities with Medium Child Mortality Rates) with 124 Regencies/Cities, cluster 1 (Regency/Cities with High Child Mortality Rates) with 31 Regencies/Cities and cluster 2 (Regency/Cities with High Child Mortality Rates). Low Child Mortality) 391 Districts/Cities. This clustering produces a Davies Bouldin Index value of -0.493, the center point of Cluster 0 is 2.744, the center point of Cluster 1 is 6.185 and the center point of Cluster 2 is 13.756.

Keywords: CMR, K-Means, Cluster, Data Mining, Regencies/Cities

Abstrak

Angka kematian merupakan salah satu indikator status kesehatan di masyarakat. Salah satu Indikator tersebut adalah Angka Kematian Anak (AKA) mencerminkan kondisi kesehatan lingkungan yang langsung mempengaruhi tingkat kesehatan anak. Algoritma K-Means membagi data menjadi 3 kelompok (Cluster) yaitu Cluster 0, 1 dan 2 untuk menghasilkan informasi tingkat Angka Kematian Anak di seluruh Kabupaten/Kota di Indonesia. Pemodelan ini menghasilkan 3 cluster yaitu cluster 0 (Kabupaten/Kota yang Angka Kematian Anak Sedang) sebanyak 124 Kabupaten/Kota, cluster 1 (Kabupaten/Kota yang Angka Kematian Anak Tinggi) sebanyak 31 Kabupaten/Kota dan cluster 2 (Kabupaten/Kota yang Angka Kematian Anak Rendah) 391 Kabupaten/Kota. Pengklusteran ini menghasilkan nilai Davies Bouldin Indeks sebesar -0,493, titik pusat Cluster 0 adalah 2,744, titik pusat Cluster 1 adalah 6,185 dan titik pusat Cluster 2 adalah 13,756.

Kata Kunci: AKA, K-Means, Cluster, Data Mining, Kabupaten/Kota

PENDAHULUAN

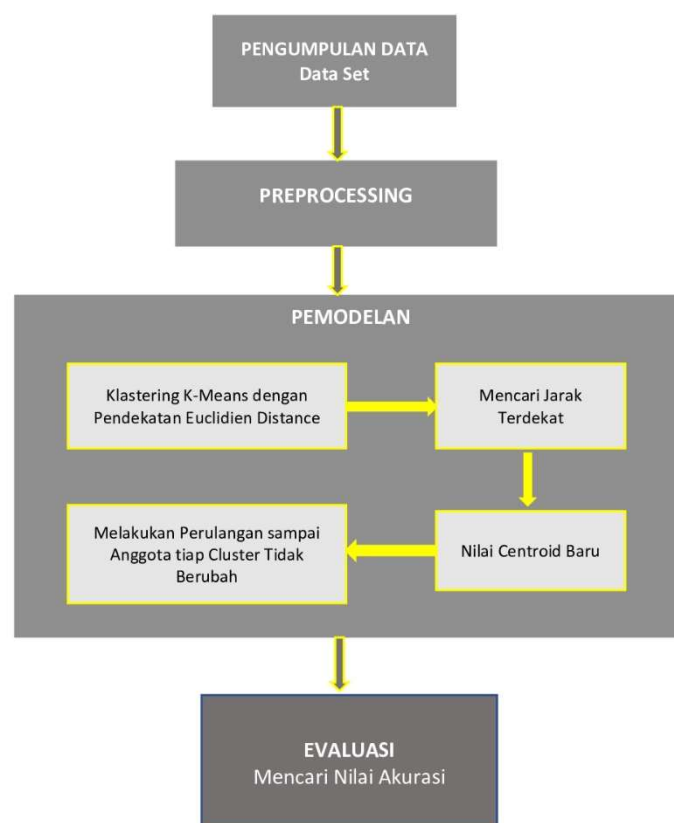
Angka kematian merupakan salah satu indikator status kesehatan di masyarakat. Salah satu Indikator tersebut adalah Angka Kematian Anak (AKA). mencerminkan kondisi kesehatan lingkungan yang langsung mempengaruhi tingkat kesehatan anak. Menurut Budi Utomo (1985) dengan melihat Angka Kematian Anak yang tinggi maka diindikasikan terjadi keadaan salah gizi atau gizi buruk, kebersihan diri dan kebersihan yang buruk, tingginya prevalensi penyakit menular pada anak, atau kecelakaan yang terjadi di dalam atau di sekitar rumah (Magelang, 2021). Berdasarkan data Dana Anak-Anak PBB (UNICEF), tingkat kematian anak di bawah usia lima tahun atau balita di Indonesia mencapai 22,17 kematian per 1.000 kelahiran hidup pada 2021. Angka itu menempatkan Indonesia di papan tengah skala Asia Tenggara. Menurut data UNICEF, tingkat kematian balita di Indonesia terus menurun dalam dekade terakhir namun berdasarkan posisi Indonesia maka negara Indonesia memiliki tantangan yang besar untuk menurunkan angka kematian anak ini (Mutia Annur, 2023) (Kautsar, 2023). Data mining adalah proses pengumpulan informasi dari suatu data yang besar, mengambil data

dari sumber data yang belum dipahami. Sebuah proses data mining biasanya menggunakan metode statistika dan matematika hingga teknologi artificial intelligence.

Clustering adalah metode untuk menganalisis data yang sering digunakan sebagai salah satu metode data mining. Tujuan dari clustering adalah untuk mengelompokkan data dengan karakteristik yang sama ke satu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain. Manfaat clustering sebagai segmentasi data yang berguna untuk memprediksi dan menganalisa masalah bisnis, serta mengidentifikasi obyek dalam berbagai bidang (Binus, 2021). Studi menggunakan Data Mining dengan Algoritma Klastering telah diimplementasikan untuk berbagai bidang, diantaranya adalah untuk klasterisasi kematian bayi berbasis desa- desa di Jawa Barat menggunakan Algoritma K-Means (Pujianti, 2023) dan penelitian oleh Zulfa Nabila (2021) yaitu untuk menganalisis data kasus Covid-19 agar dapat mengetahui pengelompokan pada masalah kasus Covid-19 di Provinsi Lampung (Nabila et al., 2021). Penelitian ini akan dilakukan pengelompokan data Angka Kematian Anak berdasarkan kota/kabupaten yang ada di Indonesia menggunakan pendekatan Algoritma klastering K-Means. Perangkat lunak yang digunakan pada penelitian ini adalah Rapid Miner sebagai tools untuk mempelajari penambangan data. Tujuan penelitian ini adalah untuk mengelompokkan kabupaten/kota di Indonesia berdasarkan Angka Kematian Anak . Penelitian ini memberikan informasi kelompok Angka Kematian Anak yang tinggi , sedang dan rendah dari seluruh kabupaten/kota di Indonesia sehingga melalui penelitian ini akan menjadi dasar bagi pihak yang terkait untuk membuat keputusan diantaranya adalah terkait pelayanan kesehatan Anak. Penelitian ini dilakukan untuk mengaplikasikan data mining dalam pengelompokan data Angka Kematian Anak (AKA) untuk menemukan tingkat AKA yang rendah supaya dapat digunakan dalam pengambilan keputusan oleh pihak terkait

METODE PENELITIAN

Tahapan pada penelitian ini terdiri dari 4 langkah yaitu Pengumpulan Data, Preprocessing, Pemodelan dan Evaluasi .



Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Pada penelitian ini dalam pengumpulan data dilakukan untuk mendapatkan sumber dataset yang akan digunakan dengan pengumpulan data sekunder yang berasal dari web Badan Pusat Statistik (BPS) bagian Kependudukan dan Migrasi yaitu data Angka Kematian Anak (Child Mortality Rate/CMR)

Hasil Long Form SP2020 . Data sekunder merupakan data yang bisa didapatkan oleh peneliti tanpa harus terjun langsung ke lapangan untuk pengambilan datanya . Jumlah data yang dipakai pada penelitian ini ialah sebanyak 546 data yang terdiri dari 2 (dua) atribut yaitu atribut CMR dan atribut Kabupaten/Kota.

2. Preprocessing

Adalah langkah pembersihan data sebelum dikelompokkan. Preprocessing adalah pembersihan data mentah yang diperoleh.

3. Klastering

Algoritma k-means mengatur satu set item n ke dalam cluster k menggunakan parameter input k sedemikian rupa sehingga kesamaan antara anggota cluster tinggi dan kesamaan antara objek individu rendah (Trivusi, 2022).

Langkah Algoritma K-Means :

- a. Klastering K-Means dengan Pendekatan Eucliden Distance . Pilih angka K untuk menentukan jumlah cluster.
- b. Mencari Jarak Terdekat . Pilih titik K atau centroid secara acak.Tetapkan setiap titik data ke centroid terdekat, yang akan membentuk cluster K yang telah ditentukan.
- c. Nilai Centroid Baru. Hitung varians dan tempatkan centroid baru dari setiap cluster.Ulangi menetapkan kembali setiap titik data ke centroid terdekat baru dari setiap cluster.
- d. Melakukan Perulangan sampai Anggota tiap Cluster Tidak Berubah

4. Evaluasi

Evaluasi termasuk untuk memastikan bahwa tes itu benar. Tes dilakukan untuk mengukur keakuratan hasil dan menemukan hasil tes yang terbaik.

HASIL DAN PEMBAHASAN

A. Data Acquisition

Dari variabel pada Gambar 2, dilakukan analisis terhadap nilai-nilai pada setiap variabel, hasilnya dapat dilihat pada gambar berikut :

Row No.	Kabupaten-...	CMR
1	ACEH	3.470
2	SIMEULUE	7.080
3	ACEH SINGKIL	8.320
4	ACEH SELAT...	6.430
5	ACEH TENG...	3.880
6	ACEH TIMUR	4.200
7	ACEH TENG...	5.790
8	ACEH BARAT	6.150
9	ACEH BESAR	3.310
10	PIDIE	8.880
11	BIREUEN	3.340
12	ACEH UTARA	5.330
13	ACEH BARA...	6.400

Gambar 2. Tampilan Data pada Rapidminer

B. Preprocessing

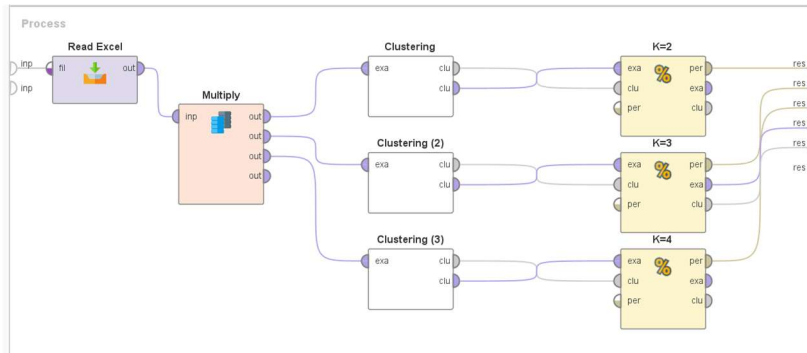
Data yang telah didapatkan dari situs Badan Pusat Statistik perlu dibersihkan terlebih dahulu dengan pengecekan data duplikat, missing value, dan outlier, hasil nya sebagai berikut :

Name	Type	Missing	Statistics	Filter (3 / 3 attributes):	Search for Attributes
✓ Kabupaten-Kota	Polynomial	0	Least YALIMO (1)	Most GORONTALO (2)	Values GORO
✓ cluster	Nominal	0	Least cluster_2 (28)	Most cluster_0 (394)	Values cluster
✓ CMR	Real	0	Min 1.490	Max 21.510	Average 4.094

Gambar 3. Tampilan Data Hasil Preprocessing pada Rapidminer

C. Pemodelan dan Evaluasi

Hasil pemodelan adalah sebagai berikut :



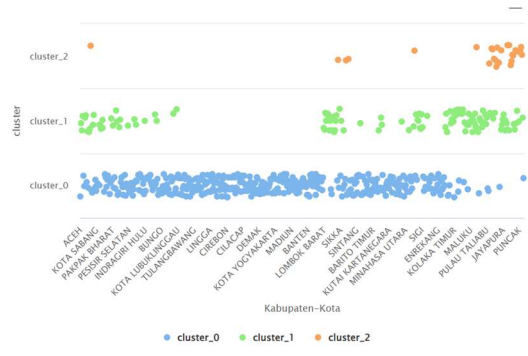
Gambar 4. Tampilan Proses Algoritma Clustering K-Means

Hasil modelan ini menghasilkan 3 cluster yaitu cluster 0 (Kabupaten/Kota yang Angka Kematian Anak Sedang) , cluster 1 (Kabupaten/Kota yang Angka Kematian Anak Tinggi) dan cluster 2 (Kabupaten/Kota yang Angka Kematian Anak Rendah). Cluster 0 sebanyak data 124 item, Cluster 1 sebanyak 31 item dan Cluster 2 sebanyak 391 item.

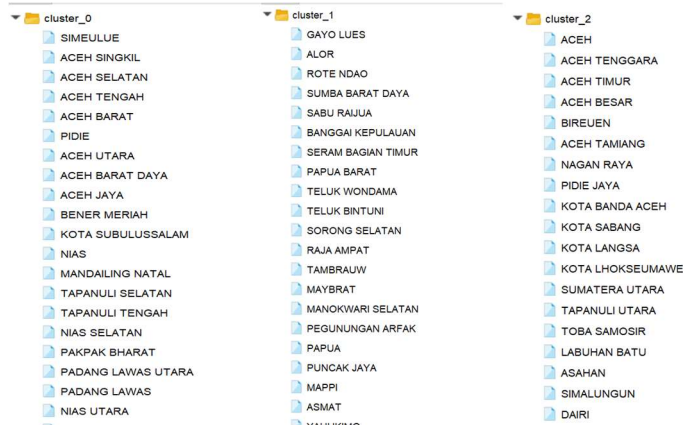
Cluster Model

Cluster 0: 124 items
 Cluster 1: 31 items
 Cluster 2: 391 items
 Total number of items: 546

Gambar 5. Tampilan Hasil Pengelompokkan Data



Gambar 6. Tampilan Hasil Visualisasi Pengelompokkan Data



Gambar 7. Tampilan Rincian Hasil Pengelompokkan Data

Dan pada hasil perhitungan rapid miner menghasilkan Davies Bouldin Indeks seperti pada gambar berikut :



Gambar 8. Nilai Davies Bouldin Indeks

Pengklusteran ini menghasilkan nilai Davies Bouldin Indeks sebesar -0,493, titik pusat Cluster 0 adalah 2,744 , titik pusat Cluster 1 adalah 6,185 dan titik pusat Cluser 2 adalah 13,756.

Attribute	cluster_0	cluster_1	cluster_2
CMR	2.744	6.185	13.756

Gambar 9. Tampilan Hasil Titik Pusat

SIMPULAN

Adapun hasil modelan ini menghasilkan 3 cluster yaitu cluster 0 (Kabupaten/Kota yang Angka Kematian Anak Sedang) sebanyak 124 data diantaranya adalah Kabupaten/Kota Aceh Singkil, Simeulue, Aceh Selatan, Pidie, Aceh Utara , cluster 1 (Kabupaten/Kota yang Angka Kematian Anak Tinggi) sebanyak 31 data diantaranya adalah Kabupaten/Kota Gayo Lues, Alor, Rote Ndao, Sumba Barat Daya dan cluster 2 (Kabupaten/Kota yang Angka Kematian Anak Rendah) 391 data diantaranya adalah Kabupaten/Kota Aceh Tenggara, Aceh Timur, Aceh Besar, Pidie Jaya .

DAFTAR PUSTAKA

Binus. (2021). *Metode Data Mining Clustering – School of Information Systems*. <https://sis.binus.ac.id/2021/10/29/metode-data-mining-clustering/>

Kautsar, A. (2023). *RI Ranking 5 Kasus Kematian Bayi Baru Lahir Tertinggi se-ASEAN*. <https://health.detik.com/berita-detikhealth/d-6877286/ri-ranking-5-kasus-kematian-bayi-baru-lahir-tertinggi-se-asean>

Magelang, P. (2021). *Pusaka Gemilang— Website Pusaka Kabupaten Magelang*. <https://pusaka.magelangkab.go.id/metadata/indikator/detailIndikator/12>

Mutia Annur, C. (2023). *Daftar Angka Kematian Balita ASEAN, Indonesia di Papan Tengah*. <https://databoks.katadata.co.id/datapublish/2023/09/05/daftar-angka-kematian-balita-asean-indonesia-di-papan-tengah>

Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100–108. <https://doi.org/10.33365/JTSL.V2I2.868>

Pujianti, A. (2023). *View of IMPLEMENTASI DATA MINING MENGGUNAKAN METODE K-MEANS CLUSTERING UNTUK MENENTUKAN STATUS KEMATIAN BAYI DI JAWA BARAT*. <https://ejournal.itn.ac.id/index.php/jati/article/view/6347/3707>

Trivusi. (2022). *K-Means Clustering: Pengertian, Cara Kerja, Kelebihan, dan Kekurangannya - Trivusi*. <https://www.trivusi.web.id/2022/06/algoritma-kmeans-clustering.html>