



IMPLEMENTASI ALGORITMA *K-MEANS CLUSTERING* DALAM KLASTERISASI KABUPATEN/KOTA PROVINSI JAWA BARAT BERDASARKAN FAKTOR PEMICU *STUNTING* PADA BALITA

Marwa Fadlila Amalia¹, Dede Brahma Arianto²

¹marwafadlila5@gmail.com, ²dede.brahma2@gmail.com

¹Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

²Program Studi Informatika Program Magister, Universitas Islam Indonesia

Abstrak

Stunting merupakan kondisi pertumbuhan serta perkembangan seorang balita yang mengalami kegagalan jika dibandingkan dengan umur yang semestinya. Pada hakikatnya, untuk mempermudah melihat karakteristik faktor pemicu balita *stunting* di suatu wilayah dapat dilakukan proses klasterisasi. Pada penelitian ini bertujuan untuk mengelompokkan wilayah kabupaten/kota Provinsi Jawa barat berdasarkan faktor pemicu balita *stunting* dengan metode *silhouette* sebagai metode penentu jumlah *cluster* optimumnya serta mengetahui karakteristik pada setiap *cluster*-nya. Hasil penelitian menunjukkan bahwa penggunaan metode *silhouette* dapat membantu algoritma *k-means* dalam proses pengelompokan kabupaten/kota di Provinsi Jawa Barat, dihasilkan 3 *cluster* sebagai *cluster* optimum dengan nilai *silhouette* yang sekitar 0,37. Karakteristik kondisi rumah tangga terhadap akses sanitasi layak yang relatif tinggi dimiliki oleh *cluster* 1 dengan 6 wilayah kabupaten/kota dan *cluster* 2 dengan 11 wilayah kabupaten/kota. Sedangkan *cluster* 3 dengan 10 wilayah kabupaten/kota ini memiliki karakteristik persentase bayi menerima inisiasi menyusui dini yang cukup tinggi dibandingkan dengan faktor lainnya.

Kata Kunci: *Stunting*, *K-Means Clustering*, Metode *Silhouette*

Abstract

Stunting is a condition of growth and development of a toddler who has failed when compared to the appropriate age. In essence, to make it easier to see the characteristics of factors that trigger *stunting* in toddlers in an area, a clustering process can be carried out. This study aims to cluster the districts/cities of West Java Province based on factors that trigger *stunting* in children under five with the *silhouette* method as a method of determining the optimum number of clusters and knowing the characteristics of each cluster. The results showed that the use of the *silhouette* method can help the *k-means* algorithm in the process of clustering districts/cities in West Java Province, resulting in 3 clusters as the optimum cluster with a *silhouette* value of around 0.37. The characteristics of household conditions with relatively high access to proper sanitation are owned by cluster 1 with 6 districts/cities and cluster 2 with 11 districts/cities. Cluster 3 with 10 districts/municipalities is characterized by a relatively high percentage of infants receiving early breastfeeding initiation compared to other factors.

Keywords: *Stunting*, *K-Means Clustering*, *Silhouette Method*

1. Pendahuluan

Permasalahan kesehatan global menjadi hal yang perlu diperhatikan untuk mendukung terciptanya *Sustainable Development Goals* (SDGs) di dunia, khususnya Negara Indonesia. Salah satu masalah kesehatan yang sedang populer di Indonesia adalah kasus *stunting* pada balita. *Stunting* merupakan kondisi pertumbuhan serta perkembangan seorang balita (dibawah 5 tahun) yang mengalami kegagalan jika dibandingkan dengan umur yang semestinya. Hal ini diakibatkan oleh kebutuhan gizi yang kurang dalam jangka waktu yang panjang, yang mana berawal ketika janin masih dalam kandungan hingga mencapai usia dua tahun [1].

Menurut informasi dari WHO (*World Health Organization*) tahun 2017, kondisi *stunting* berdampak pada sekitar 150,8 juta anak di seluruh dunia yang berusia di bawah 5 tahun atau sekitar 22,2%. Dari jumlah balita yang mengalami *stunting* pada tahun tersebut, 55% di antaranya berasal dari Asia. Sedangkan 39% dari jumlah balita yang mengalami *stunting* berasal dari Afrika. Berdasarkan data

yang telah dikumpulkan oleh WHO tahun 2017 yang lalu, Indonesia menjadi negara ketiga dengan tingkat prevalensi tertinggi di Asia Tenggara setelah Timor Leste dan India [2]. Hal ini sejalan dengan adanya peningkatan angka prevalensi dari tahun 2016 sebesar 27,5% menjadi 26,9% pada tahun 2017. Fakta *stunting* juga bisa dilihat berdasarkan hasil dari Survei Status Gizi Balita Indonesia (SSGI) pada tahun 2021, dimana angka prevalensi *stunting* di Indonesia mengalami penurunan dari tahun 2019 sampai 2021, yaitu sebesar 27,7% menjadi 24,4% [3]. Berdasarkan hasil Forum Nasional *Stunting*, Dante Saksono Harbuwono mengatakan bahwa masih terdapat 12 provinsi di Indonesia yang perlu mendapatkan penanganan dalam upaya untuk menangani penurunan *stunting* [4]. Salah satu provinsi di Indonesia yang mempunyai kasus *stunting* terbesar adalah Provinsi Jawa Barat. Hasil Survei Status Gizi Indonesia (SSGI) tahun 2021 menunjukkan bahwa angka prevalensi *stunting* di Jawa Barat mencapai 24,5%, yang berarti bahwa 2 dari 10 anak di Provinsi Jawa Barat berisiko terjadi *stunting*. Angka tersebut masih belum memenuhi standar yang ditetapkan WHO, yaitu dibawah 20% [4] [5]. Sejalan dengan Perpres nomor 72 tahun 2021, pemerintah telah melakukan cara untuk mempercepat penurunan angka *stunting* melalui 5 pilar [6]. Kelima pilar tersebut yaitu komitmen, pencegahan *stunting*, harus bisa melakukan konvergensi, menyediakan pangan yang baik dan melakukan inovasi dan data yang baik. Pada pilar kelima ini, pemerintah Indonesia masih belum bisa melakukan distribusi pelayanan dengan baik. Hal ini dikarenakan setiap daerah memiliki kondisi demografi yang berbeda sehingga perlu dilakukannya pengelompokan wilayah berdasarkan tingkat kemiripannya supaya pemerintah lebih mudah untuk memonitoring pelayanan sesuai dengan keadaan wilayahnya. Pengelompokan wilayah berdasarkan kemiripannya bisa dilakukan dengan menggunakan metode *machine learning* berupa *clustering*.

Clustering merupakan salah satu metode *machine learning unsupervised learning* yang berguna dalam proses klasterisasi data menjadi beberapa *cluster* atau kelompok berdasarkan kesamaan atau perbedaan antara data tersebut. Metode ini akan mengelompokkan objek berdasarkan ukuran kedekatan dari suatu karakteristik sampel yang berdekatan. Salah satu *machine learning* yang digunakan untuk mengungkap pola atau struktur tersembunyi dari data, adalah *k-means clustering*. Dalam algoritma *unsupervised learning*, *k-means* memiliki peran penting dalam mengklasifikasikan objek-objek dalam data yang tidak memiliki label melalui pemahaman struktur data dan pengelompokkan berdasarkan kemiripan fitur. Algoritma ini menjadi salah satu algoritma yang paling mudah untuk diimplementasikan dalam memecahkan suatu permasalahan karena memerlukan waktu yang cukup singkat. Algoritma ini bisa digunakan untuk data dengan skala besar sehingga lebih mudah untuk mendeskripsikan karakteristik pada setiap *cluster* [7].

Penelitian terkait algoritma *k-means clustering* sudah banyak dilakukan sebelumnya, baik di bidang kesehatan, lingkungan, maupun bidang lainnya. Penelitian Azizah, et al. tahun 2022 menunjukkan bahwa metode *k-means clustering* dengan metode kernel terbaik untuk data kualitas air Sungai di Kota Semarang adalah dengan 2 atau 4 *cluster*. Dengan 4 *cluster*, sungai di Kota Semarang di bagi menjadi 4 kelompok, yaitu *cluster* 1 terdiri dari 12 sungai, 14 sungai pada *cluster* 2, 11 sungai mengelompok pada *cluster* 3 sedangkan *cluster* 4 terdiri dari 10 sungai. Kedua *cluster* pertama, yaitu *cluster* 1 dan 2 memiliki faktor pencemaran air yang lebih tinggi dari *cluster* lainnya. Hal ini dikarenakan terdapat 6 variabel yang memiliki nilai tinggi di kedua *cluster* tersebut [8]. Berdasarkan hasil penelitian Sari, et al tahun 2023 dapat diketahui bahwa pengelompokan data sesuai kriteria dapat dilakukan menggunakan teknik penambangan data *k-means*. Dengan menggunakan algoritma tersebut dihasilkan jumlah *cluster* terbaik dengan nilai k sebesar 2. Dari 395 kumpulan data telah dikelompokkan menjadi 2 *cluster* dimana 287 balita kerdil dan 108 balita dengan perkembangan normal, nilai DBI paling ideal yaitu sebesar 0,007. Nilai tersebut mendekati 0 sehingga dapat menunjukkan bahwa *cluster* yang sedang ditinjau menyediakan *cluster* yang berkualitas [9]. Sartika, et al. tahun 2020 membahas tentang proses pengelompokan menggunakan metode *k-means* dan evaluasi struktur *cluster* menggunakan siluet. Penelitian ini diperoleh nilai evaluasi 34 provinsi yang dibagi menjadi 2 kelompok yaitu nilai yang baik untuk *cluster* 1 dengan anggota 9 provinsi dan nilai yang kurang untuk *cluster* 2 dengan anggota 25 provinsi [10]. Penelitian terdahulu yang dilakukan oleh Fadilah, et al. tahun 2022 didapatkan kesimpulan bahwa dengan menggunakan bantuan metode *elbow*, jumlah *cluster* optimal yang dihasilkan yaitu sebesar 2 *cluster* dengan rincian *cluster* 1 terdiri dari 324 kabupaten/kota, sedangkan *cluster* 2 terdiri 49 kabupaten/kota [11]. Algoritma *k-means* juga dapat digunakan untuk mengklasifikasikan provinsi di Indonesia berdasarkan paket pelayanan *stunting*. Hal ini sesuai dengan penelitian yang dilakukan oleh Matdoan, et al tahun 2022 bahwa klasifikasi provinsi menurut pelayanan *stunting* di Indonesia dibagi dalam 3 *cluster*

yaitu *cluster* 0 terdiri dari Provinsi Riau, Bengkulu, Jambi, Kepulauan Bangka Belitung, Sumatera Barat, Banten, Bali, Kalbar, Sulawesi Utara, Nusa Tenggara Barat, Sulawesi Tengah, Kalimantan Utara, Gorontalo, DKI Jakarta, Kepulauan Riau, Kalimantan Timur, Sulawesi Tenggara, DI Yogyakarta, Kalimantan Tengah, Kalimantan Selatan, dan Kalimantan Barat. Provinsi Jawa Barat, Maluku, Sulawesi Barat, Jawa Tengah, Maluku Utara, Jawa Timur, Papua, Papua Barat serta Aceh termasuk anggota dari *cluster* 1. Sementara Provinsi Sulawesi Selatan, Sumatera Selatan, Lampung, Nusa Tenggara Timur, dan Sumatera Utara menjadi anggota *cluster* 2 [12]. Terdapat juga penelitian yang dilakukan oleh Apriyani, et al. tahun 2023 dengan judul Penerapan Algoritma *K-Means* dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi dapat disimpulkan bahwa jumlah *cluster* optimal yang didapatkan sebanyak 2 *cluster* dengan rincian 392 balita merupakan *cluster* 0 dan 3 balita merupakan *cluster* 1. Pada pengujian, nilai evaluasi performa optimal didapatkan sebesar 0,007 ketika $k = 2$. Berdasarkan nilai Davies Bouldin Index (DBI) dengan menggunakan bantuan tools RapidMiner, nilai evaluasi performa tersebut mendekati 0 sehingga dapat diartikan bahwa *cluster* yang dievaluasi menghasilkan jumlah *cluster* yang baik [13]. Hasil penelitian Winarta & Kurniawan tahun 2021 tentang Optimasi Cluster *K-Means* Menggunakan Metode Elbow pada Data Pengguna Narkoba dengan Pemrograman Python menunjukkan bahwa *k-means* mampu membentuk kelompok dengan penentuan anggota *cluster* yang tepat melalui eksekusi pemrograman *python*. Informasi yang didapatkan bahwa k sebanyak 3 menjadi jumlah *cluster* optimal dengan perhitungan nilai selisih tertinggi pada SSE [14].

Berdasarkan referensi penelitian terdahulu, beberapa penelitian mengadopsi metode pada *k-means* yang masih menjadi bias dalam penentuan jumlah *cluster* optimumnya. Salah satu metode tersebut adalah metode *elbow*. Metode ini masih dikatakan bias karena hasil yang diberikan tidak selalu akurat dan terkadang masih terdapat data pada beberapa *cluster* yang dianggap tidak berada di *cluster* yang sesuai. Oleh karena itu, untuk menutupi kekurangan metode tersebut serta mengacu pada latar belakang, pada penelitian ini akan dilakukan klasterisasi wilayah kabupaten/kota di Provinsi Jawa Barat berdasarkan kesamaan faktor pemicu *stunting* pada balita dengan menggunakan metode *silhouette*. Metode *silhouette* ini sebagai dasar dalam penentuan jumlah *cluster* terbaiknya. Adapun variabel yang diambil sebagai faktor pemicu *stunting* antara lain, persentase rumah tangga yang memiliki akses rumah tangga layak, persentase pemberian ASI (Air Susu Ibu) eksklusif, jumlah bayi mendapat imunisasi dasar lengkap, persentase bayi menerima inisiasi menyusui dini, dan jumlah pengeluaran per kapita [15].

2. Metode

Data yang digunakan pada penelitian ini merupakan data yang diperoleh dari hasil pencarian langsung melalui *website* resmi milik Badan Pusat Statistik (BPS) Provinsi Jawa Barat [16] dan OpenData Provinsi Jawa Barat [17] [18]. Data yang diambil pada *website* sebanyak 27 berdasarkan kabupaten/kota di Provinsi Jawa Barat pada tahun 2021. Selain itu, metode analisis data yang digunakan dalam penelitian ini adalah metode algoritma *k-means clustering* dengan pemilihan jumlah *cluster* menggunakan metode *Silhouette*. *K-means clustering* adalah salah satu algoritma *clustering* yang tergolong cukup sederhana dalam prosesnya dengan cara mempartisi atau mengelompokkan *dataset* menjadi beberapa *cluster* yang telah ditentukan. Jumlah *cluster* optimum ditentukan berdasarkan nilai *silhouette* tertinggi. Di bawah ini merupakan tahapan analisis untuk mengelompokkan wilayah kabupaten/kota di Provinsi Jawa Barat berdasarkan faktor pemicu terjadinya *stunting* pada balita menggunakan algoritma *k-means clustering* dengan metode *Silhouette* beserta diagram alirnya.

1) Melakukan Pengumpulan Data (*Data Collecting*)

Pengumpulan data (*data collection*) adalah langkah awal yang penting dalam proses pembelajaran mesin (*machine learning*). Kualitas data yang dikumpulkan harus sesuai dengan tujuan dari model *machine learning* yang akan dibangun karena sangat mempengaruhi kualitas model yang dihasilkan. Pengumpulan data pada penelitian ini melalui *website* resmi dari BPS dan OpenData Provinsi Jawa Barat kemudian data yang didapatkan akan disimpan dalam format CSV (*Comma Separated Values*).

2) Mengimpor Data (*Data Importing*)

Data yang sudah dikumpulkan dan disimpan dalam bentuk CSV akan diimpor melalui *files* pada *R studio*.

3) Melakukan Pengecekan Tipe Data (*Data Understanding*)

Pada langkah ini akan dilakukan pengecekan data untuk menentukan tipe data pada setiap variabelnya.

4) Melakukan Pembersihan Data (*Data Cleaning*)

Pembersihan data merupakan proses mengidentifikasi dan mengoreksi kesalahan, inkonsistensi, atau ketidaklengkapan dalam data yang sudah dikumpulkan. Tahap ini bertujuan untuk meningkatkan kualitas data sehingga model *machine learning* yang dihasilkan mampu memiliki kinerja yang optimal. Data yang sudah dikumpulkan pada penelitian ini masih terdapat *missing values* pada kolom jumlah pengeluaran rumah tangga per kapita baris terakhir. Oleh karena itu, akan dilakukan imputasi data dengan menggunakan rata-rata untuk mengatasi data yang hilang tersebut.

5) Melakukan Uji Kecukupan Sampel

Sebelum masuk ke tahapan algoritma *k-means clustering*, uji kecukupan sampel perlu dilakukan untuk mengetahui apakah sampel mewakili populasi atau tidak. Uji yang bisa digunakan untuk mengetahui apakah sampel mewakili populasi adalah uji KMO (*Kaiser Meyer Olkin*). Apabila ingin memenuhi asumsi kecukupan sampel, nilai KMO yang dihasilkan harus lebih besar dari 0,5 [19].

6) Melakukan Uji Non-multikolinearitas

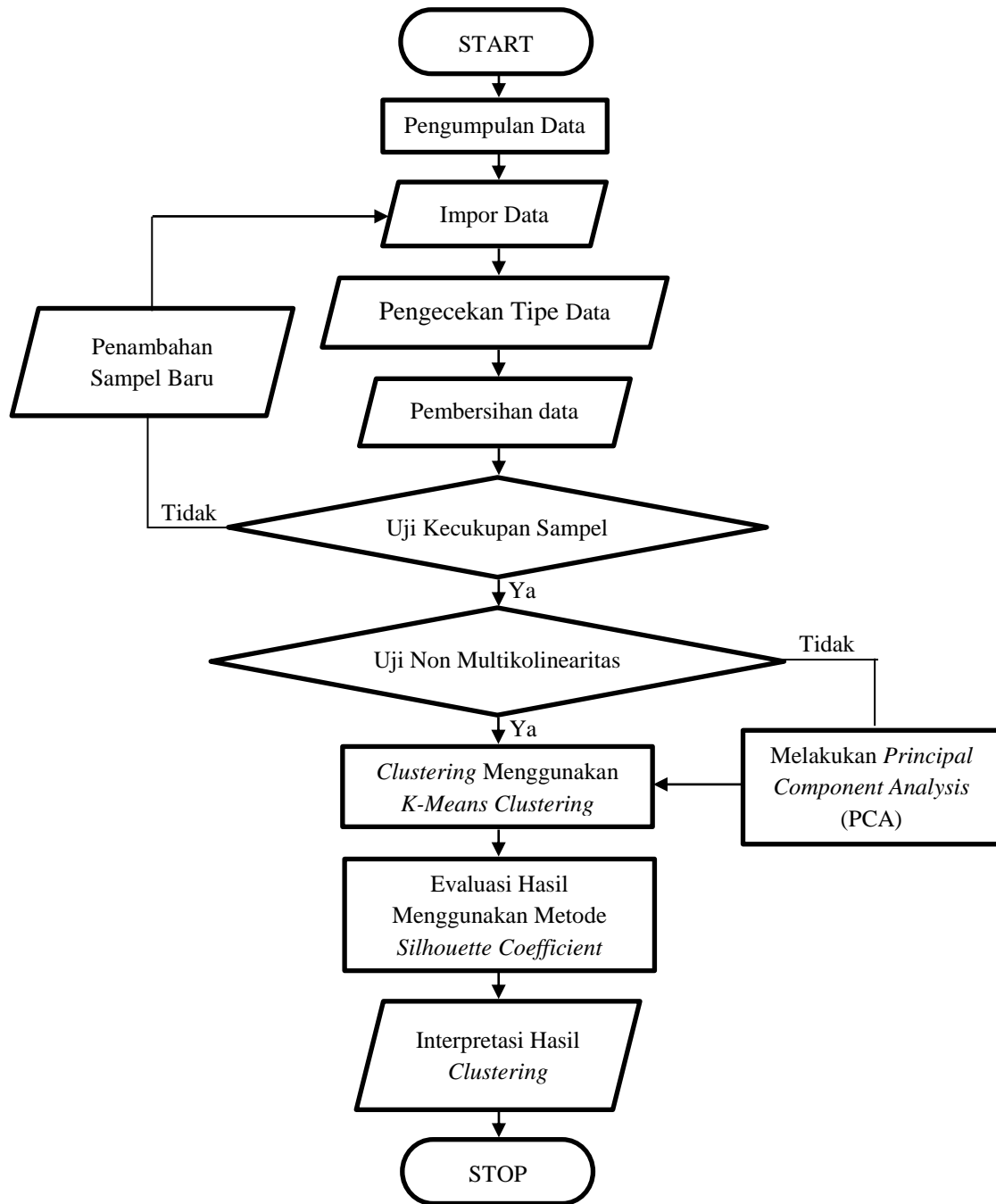
Pada algoritma *clustering*, uji non-multikolinearitas menjadi asumsi kedua setelah uji kecukupan sampel. Uji ini dapat ditentukan melalui nilai VIF (*Variance Inflation Function*) pada setiap variabelnya. Suatu variabel dikatakan tidak terjadi multikolinearitas apabila memiliki nilai VIF (*Variance Inflation Function*) kurang dari 10 [19].

7) Melakukan Pengolahan Metode *Data Mining*

Pada tahap ini, metode *data mining* yang digunakan adalah *k-means clustering* untuk mencari *cluster* terbaik dalam pengelompokan wilayah kabupaten/kota di Provinsi Jawa Barat dengan memanfaatkan metode *silhouette*.

8) Mengevaluasi Jumlah *Cluster* Optimum

Evaluasi jumlah *cluster* optimum dilakukan dengan memanfaatkan metode *Silhouette* dengan menghitung rata-rata skor *Silhouette* pada setiap *cluster* yang berkisar antara -1 dan 1. Skor tertinggi menjadi penentu jumlah *cluster* yang akan terbentuk. Jika sudah menentukan jumlah *cluster*, akan dilakukan pengelompokan ulang wilayah kabupaten/kota Provinsi Jawa Barat berdasarkan faktor pemicu *stunting* pada balita.



Gambar 1. Diagram Alir Algoritma *K-Means Clustering*

3. Hasil dan Pembahasan

3.1 Pengecekan Tipe Data

Pada tahap ini, data sekunder yang sudah dikumpulkan oleh peneliti akan dilihat jenis tipe datanya pada setiap variabel. Adapun deskripsi data tersebut dapat dilihat melalui tabel 1 di bawah ini.

Tabel 1. Deskripsi Data

Variabel	Tipe Data
Nama Kabupaten/Kota	Objek
X1 (Persentase rumah tangga yang memiliki akses sanitasi layak)	Float
X2 (Persentase pemberian ASI (Air Susu Ibu) eksklusif)	Float
X3 (Persentase bayi menerima inisiasi menyusui dini)	Float

Berdasarkan tabel di atas dapat dilihat beberapa informasi mengenai tipe data pada setiap variabelnya. Terlihat bahwa ada dua tipe data, diantaranya data kuantitatif berupa float dan data kualitatif berupa objek. Tipe data float merupakan tipe data kuantitatif yang menyatakan bahwa bilangan tersebut berupa bilangan desimal. Sedangkan tipe data objek merupakan tipe data yang digunakan untuk merepresentasikan objek dalam pemrograman. Sesuai dengan *output* di atas, variabel nama kabupaten/kota termasuk ke dalam tipe data objek, sementara variabel persentase rumah tangga yang memiliki akses sanitasi layak, persentase pemberian ASI (Air Susu Ibu) eksklusif, dan persentase bayi menerima inisiasi menyusui dini termasuk ke dalam tipe data kuantitatif berupa float.

3.2 Persiapan Data

Pada tahap persiapan data, data yang sudah dikumpulkan akan dilakukan pengintegrasian data, sehingga didapatkan sebanyak 81 *record* data yang terdiri dari 3 faktor pemicu *stunting* di mana setiap faktornya terdapat 27 data sesuai dengan jumlah kabupaten/kota di Provinsi Jawa Barat. Pada tabel 2 menunjukkan dataset faktor pemicu balita *stunting* Provinsi Jawa Barat tahun 2021, terdiri dari X1 mewakili variabel persentase rumah tangga yang memiliki akses sanitasi layak, variabel persentase pemberian ASI (Air Susu Ibu) diwakili oleh X2 sedangkan X3 mewakili variabel persentase bayi menerima inisiasi menyusui dini.

Tabel 2. Dataset Faktor Pemicu Balita *Stunting*

Nama Kabupaten/Kota	X1	X2	X3
Bogor	63,91	48,58	52,95
Sukabumi	64,35	72,76	82,26
Cianjur	58,52	71,85	87,23
Bandung	64,51	66,16	95,44
Garut	43,73	74,01	89,12
Tasikmalaya	49,35	51,04	90,46
Ciamis	66,59	80,39	93,79
Kuningan	80,98	76,29	77
Cirebon	83,41	70,86	81,43
Majalengka	80,59	80,32	92,88
Sumedang	87,87	81,26	88,11
Indramayu	89,75	55,46	85,94
Subang	85,82	52,57	82,1
Purwakarta	72,61	63,19	78,68
Karawang	77,19	68,15	78,47
Bekasi	81,02	72,62	88,85
Bandung Barat	63,78	75,17	78,46
Pangandaran	81,49	100	84,27
Kota Bogor	75,35	42,52	95,85
Kota Sukabumi	39,64	67,5	100
Kota Bandung	48,9	72,64	78,84

Kota Cirebon	92,71	71,81	81,69
Kota Bekasi	97,54	50,29	78,99
Kota Depok	97,06	73,61	96,15
Kota Cimahi	78,67	77,05	85,18
Kota Tasikmalaya	52,62	70,38	85
Kota Banjar	81,92	74,29	84,59

Langkah selanjutnya adalah melakukan pengecekan duplikasi. Setelah dilakukan pengecekan ternyata tidak terdapat duplikasi pada dataset. Pada data ini juga tidak perlu dilakukan transformasi karena antar datanya sudah memiliki skala yang sama.

3.3 Analisis Deskriptif

Suatu penelitian, analisis deskriptif dilakukan untuk mendapatkan gambaran secara general tentang data penelitian yang digunakan. Adapun informasi yang diperlukan meliputi jumlah data, nilai minimum, maksimum, rata-rata, dan standar deviasi. Statistik deskriptif faktor pemicu *stunting* pada balita disajikan dalam tabel 3.

Tabel 3. Statistik Deskriptif Faktor Pemicu Stunting Provinsi Jawa Barat Tahun 2021

Variabel	N	Minimum	Maksimum	Mean	Standar Deviasi
X1	27	39,64	97,54	72,59	16,20
X2	27	42,52	100	68,92	12,33
X3	27	52,95	100	84,95	9,04

Tabel 3 di atas memberikan informasi berupa statistik deskriptif untuk setiap variabelnya. Ketiga variabel memiliki jumlah data yang sama yaitu sebanyak 27 data. Pada variabel X1 diketahui nilai minimum sebesar 39,64, sedangkan nilai maksimum sebesar 97,54. Selain itu, dapat diketahui juga rata-rata pada variabel X1 sebesar 72,59 serta nilai standar deviasi sebesar 16,20. Nilai minimum sebesar 42,52 dan maksimum sebesar 100 dimiliki oleh variabel X2. Selain itu, dapat dilihat juga nilai standar deviasi dan rata-ratanya sebesar 12,33 dan 68,92. Sementara untuk variabel X3 memiliki nilai minimum, maksimum, rata-rata, standar deviasi yaitu sebesar 52,95, 100, 84,95, dan 9,04.

3.4 Pengelompokan dengan Algoritma *K-Means Clustering*

3.4.1 Pengujian Asumsi

Pada pengelompokan dengan metode *k-means clustering* langkah pertama yang dilakukan adalah menguji asumsi apakah variabel-variabel yang digunakan sudah tepat untuk digunakan.

1) Asumsi Sampel Representatif

Pada penelitian yang menggunakan metode *k-means clustering* variabel-variabel yang akan digunakan dilakukan uji asumsi apakah sampel yang digunakan dapat merepresentasikan populasi data terlebih dahulu. Uji yang dapat digunakan untuk mengetahui apakah sampel dapat mewakili populasi adalah uji KMO (*Kaiser Meyer Olkin*). Berikut merupakan hasil pengujian asumsi sampel representatif.

Tabel 4. Asumsi Sampel Representatif

KMO Test	
KMO (<i>Kaiser Meyer Olkin</i>)	0,551

Berdasarkan tabel 4 di atas, nilai KMO yang diperoleh sebesar 0,551. Nilai tersebut lebih besar dari 0,5. Hal ini menunjukkan bahwa sampel yang diambil telah merepresentasikan atau mewakili populasi yang ada.

2) Asumsi Non-Multikolinearitas

Uji asumsi non-multikolinearitas dilakukan untuk menguji apakah antar variabel independen yang digunakan pada penelitian ini memiliki korelasi atau tidak. Asumsi ini terpenuhi jika multikolinearitas tidak terjadi. Berikut merupakan ringkasan nilai VIF (*Variance Inflation Function*) variabel penelitian.

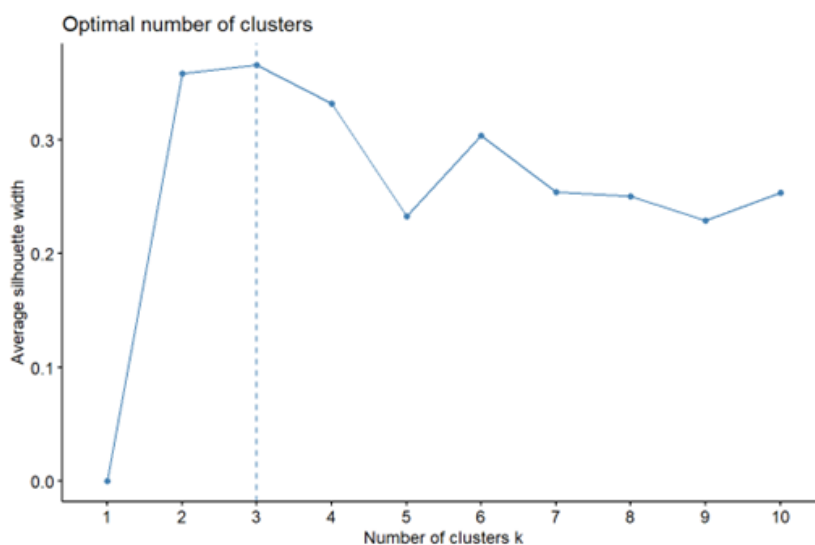
Tabel 5. Ringkasan Nilai VIF Variabel Penelitian

Variabel bebas	Nilai VIF
X1	1,010
X2	1,041
X3	1,048

Berdasarkan tabel 5 tersebut, setiap variabel bebas memiliki nilai VIF (*Variance Inflation Function*) di bawah 10. Hal ini membuktikan bahwa asumsi kedua dalam *k-means clustering* terpenuhi yaitu tidak terjadi multikolinearitas atau antar variabel bebas tidak saling berkorelasi.

3.4.2 Hasil Pengelompokan *K-Means Clustering*

Penentuan jumlah *cluster* (k) dan maksimum iterasi menjadi parameter yang digunakan pada proses *clustering* menggunakan algoritma *k-means*. Penentuan jumlah *cluster* terbaik pada penelitian ini didasari dengan nilai *silhouette coefficient* pada metode *silhouette*. Hasil pengelompokan dilakukan dengan menentukan jumlah k *cluster* sebanyak 2 sampai 10. Berikut merupakan bentuk plot antara jumlah *cluster* dengan nilai *silhouette* dari 3 faktor pemicu *stunting* berdasarkan wilayah kabupaten/kota di Provinsi Jawa Barat tahun 2021.

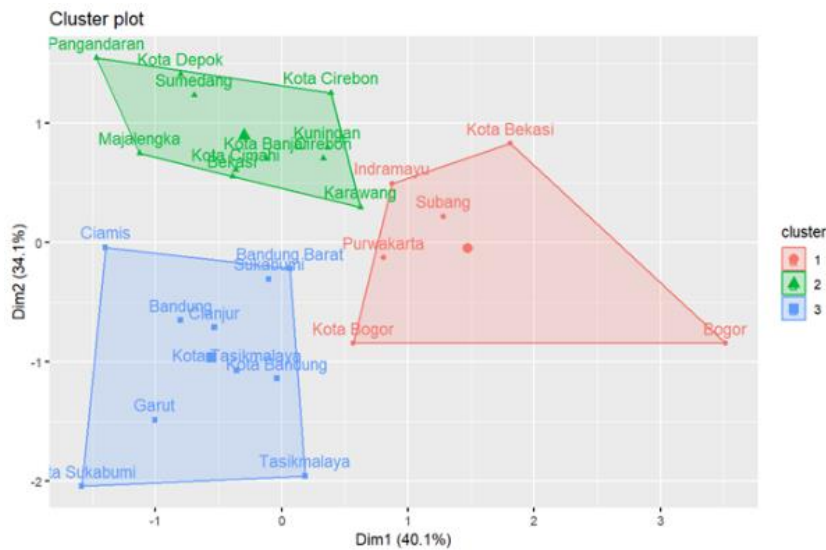


Gambar 2. Jumlah Cluster Optimum

Berdasarkan gambar 2 di atas dapat diketahui bahwa nilai *silhouette* tertinggi diperoleh ketika k sebesar 3 dengan nilai *silhouette* sekitar 0,37. Hal ini ditandai juga dengan tergamabarnya garis putus-putus vertikal yang menarik jumlah cluster 3 ke nilai *silhouette* sekitar 0,37. Jika dibandingkan dengan jumlah cluster yang lain, jumlah inilah yang lebih sesuai untuk dijadikan sebagai jumlah cluster optimum karena nilai *silhouette* yang dimiliki lebih mendekati angka 1. Pada penelitian ini, dipilih jumlah cluster sebanyak 3 cluster dengan tujuan untuk melihat pengelompokan wilayah kabupaten/kota di Provinsi Jawa Barat berdasarkan faktor pemicu balita *stunting* serta melihat karakteristik pada setiap kelompoknya. Adapun hasil pengelompokan wilayah kabupaten/kota di Provinsi Jawa Barat ke dalam 3 cluster dituliskan pada tabel 6 serta dibuktikan melalui gambar scatter plot di bawah ini.

Tabel 6. Daftar Wilayah Setiap Cluster

Cluster	Wilayah
1	Bogor, Indramayu, Subang, Purwakarta, Kota Bogor, Kota Bekasi
2	Kuningan, Cirebon, Majalengka, Sumedang, Karawang, Bekasi, Pangandaran, Kota Cirebon, Kota Depok, Kota Cimahi, Kota Banjar
3	Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Bandung Barat, Kota Sukabumi, Kota Bandung, Kota Tasikmalaya



Gambar 3. Plotting Cluster

Berdasarkan hasil *plotting cluster* pada gambar 3 dapat diketahui bahwa terdapat 3 *cluster* pada pengelompokan kabupaten/kota faktor pemicu *stunting* pada balita di Provinsi Jawa Barat tahun 2021. *Cluster* 1 ditandai dengan *plot* berwarna merah yang terdiri dari 6 wilayah kabupaten/kota. *Plot* warna hijau sebagai penanda pada *cluster* 2 dengan wilayah sebanyak 11 kabupaten/kota. Sedangkan untuk *cluster* 3 ditandai dengan *plot* berwarna biru yang terdiri dari 10 wilayah kabupaten/kota. Setelah pengelompokan wilayah, tahap selanjutnya adalah memberikan karakteristik khusus yang menggambarkan isi pada setiap *cluster*. Karakteristik akan diidentifikasi berdasarkan hasil analisis deskriptif berupa rata-rata setiap variabel faktor pemicu *stunting* pada balita di setiap *cluster*-nya.

Tabel 7. Karakteristik Setiap Cluster

Variabel	Cluster		
	1	2	3
X1	80,8	83,9	55,2
X2	52,1	76,9	70,2
X3	79,1	85,3	88,1

Berdasarkan tabel 7 dapat diketahui karakteristik pada masing-masing *cluster*-nya. Pada *cluster* 1 cenderung memiliki karakteristik kondisi rumah tangga yang memiliki akses sanitasi layak yang relatif tinggi dengan rata-rata sebesar 80,8. Akan tetapi tingkat pemberian ASI eksklusif kepada balita masih relatif rendah dengan nilai rata-rata sekitar 52,1. Diketahui juga nilai rata-rata pada variabel X3, bayi menerima inisiasi menyusui dini, yaitu sebesar 79,1. Lain halnya dengan anggota wilayah kabupaten/kota yang berada di *cluster* 2 yang mana karakteristik persentase bayi menerima inisiasi menyusui dini cukup tinggi dibandingkan dengan faktor lainnya. Hal ini dapat dilihat pada nilai rata-rata pada variabel X3 yaitu sebesar 85,3, untuk nilai rata-rata variabel X1 dan X2 sebesar 83,9 dan 76,9. Sementara, pada *cluster* 3, wilayah kabupaten/kotanya cenderung memiliki karakteristik yang sama dengan *cluster* 2 dimana nilai rata-rata tertinggi dimiliki oleh variabel X3, persentase bayi menerima inisiasi menyusui dini. Selain itu juga terlihat nilai rata-rata untuk persentase rumah tangga yang memiliki akses sanitasi layak atau X1 dan pemberian ASI (Air Susu Ibu) eksklusif atau X2, yaitu sebesar 55,2 dan 70,2.

4. Kesimpulan dan Saran

Berdasarkan hasil dan pembahasan penelitian di atas, didapatkan kesimpulan bahwa penggunaan metode *silhouette* dapat membantu algoritma *k-means* dalam proses klasterisasi kabupaten/kota di Provinsi Jawa Barat. Pada penelitian tersebut dihasilkan 3 *cluster* sebagai *cluster* optimum dengan nilai *silhouette* yang didapatkan sekitar 0,37. Pada *cluster* 1 terdiri dari 6 wilayah kabupaten/kota, antara lain Kabupaten Bogor, Indramayu, Subang, Purwakarta, Kota Bogor, dan Kota Bekasi dengan karakteristik kondisi rumah tangga yang memiliki akses sanitasi layak yang relatif tinggi. Kabupaten Kuningan, Cirebon, Majalengka, Sumedang, Karawang, Bekasi, Pangandaran, Kota Cirebon, Kota Depok, Kota Cimahi, dan Kota Banjar menjadi 11 kabupaten/kota yang berada pada *cluster* 2 dengan karakteristik persentase bayi menerima inisiasi menyusui dini cukup tinggi dibandingkan dengan faktor lainnya. *Cluster* 3 dengan 10 wilayah kabupaten/kota antara lain Kabupaten Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Bandung Barat, Kota Sukabumi, Kota Bandung, dan Kota Tasikmalaya, memiliki karakteristik seperti *cluster* 2 dimana persentase bayi menerima inisiasi menyusui dini cukup tinggi dibandingkan dengan faktor lainnya yaitu sebesar 88,1. Sedangkan untuk nilai rata-rata untuk variabel X1 dan X3 yaitu sebesar 55,2 dan 70,2

Berdasarkan penelitian ini, saran membangun yang bisa disampaikan adalah pemodifikasian algoritma *k-means* dengan algoritma lainnya. Selain itu, penelitian selanjutnya diharapkan tidak hanya menentukan jumlah *cluster* optimum melainkan juga dapat menentukan *cluster* mana yang menjadi *cluster* terbaik. Diperlukan juga informasi mengenai perbandingan hasil *output* dari *software R studio* dengan perhitungan secara manual pada hasil pengelompokan menggunakan algoritma *k-means* metode *silhouette*.

Daftar Pustaka

- [1] Kementerian Kesehatan Republik Indonesia, “Stunting,” ayosehat.kemkes.go.id. Accessed: Oct. 25, 2023. [Online]. Available: <https://ayosehat.kemkes.go.id/topik-penyakit/covid-19/stunting>
- [2] A. Hidayah, Y. Siswanto, and K. D. Pertiwi, “Riwayat Pemberian MP-ASI dan Sosial Ekonomi dengan Kejadian Stunting pada Balita,” *J. Penelit. dan Pengemb. Kesehat. Masy. Indones.*, vol. 2, no. 1, pp. 76–83, 2021, doi: 10.15294/jppkmi.v2i1.47526.
- [3] Rokom, “Penurunan Prevalensi Stunting tahun 2021 sebagai Modal Menuju Generasi Emas Indonesia 2045,” sehatnegeriku.kemkes.go.id. Accessed: Oct. 24, 2023. [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/umum/20211227/4339063/penurunan-prevalensi-stunting-tahun-2021-sebagai-modal-menuju-generasi-emas-indonesia-2045/>
- [4] Malini, “Angka Stunting di Indonesia Masih Tinggi, Ini 5 Provinsi dengan Kasus Stunting Terbanyak,” nasional.tempo.co. Accessed: Oct. 24, 2023. [Online]. Available: <https://nasional.tempo.co/read/1683885/angka-stunting-di-indonesia-masih-tinggi-ini-5-provinsi-dengan-kasus-stunting-terbanyak>
- [5] Mikhael Gewati, “Upaya Dinkes Jawa Barat Mewujudkan Zero Stunting,” regional.kompas.com. Accessed: Oct. 24, 2023. [Online]. Available: <https://regional.kompas.com/read/2022/11/23/08213781/upaya-dinkes-jawa-barat-mewujudkan-zero-stunting>
- [6] Kementerian Sekretariat Negara RI Sekretariat Wakil Presiden, “Perpres Nomor 72 tahun 2021

-
- Tentang Percepatan Penurunan Stunting,” stunting.go.id. Accessed: Oct. 24, 2023. [Online]. Available: <https://stunting.go.id/perpres-nomor-72-tahun-2021-tentang-percepatan-penurunan-stunting/>
- [7] S. Paembonan and H. Abduh, “Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat,” *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.
- [8] A. N. Azizah, T. Widiari, and A. R. Hakim, “Kernel K-Means Clustering untuk Pengelompokan Sungai di Kota Semarang Berdasarkan Faktor Pencemaran Air,” *J. Gaussian*, vol. 11, no. 2, pp. 228–236, 2022, doi: 10.14710/j.gauss.v11i2.35470.
- [9] I. Technology, “Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases,” *J. Comput. Sci. Inf. Technol. Telecommun. Eng.*, vol. 4, no. 2, pp. 402–412, 2023, doi: 10.30596/jcositte.v4i2.15765.
- [10] W. Sartika, S. Suryono, and A. Wibowo, “Information System for Evaluating Specific Interventions of Stunting Case Using K-means Clustering,” *E3S Web Conf.*, vol. 202, 2020, doi: 10.1051/e3sconf/202020213003.
- [11] A. Fadilah, M. N. Pangestu, S. Lumbanbatu, and S. Defiyanti, “Pengelompokan Kabupaten/Kota Di Indonesia Berdasarkan Faktor Penyebab Stunting Pada Balita Menggunakan Algoritma K-Means,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, p. 223, 2022, doi: 10.26798/jiko.v6i2.581.
- [12] M. Y. Matdoan, U. A. Matdoan, and M. Saleh Far-Far, “Algoritma K-Means Untuk Klasifikasi Provinsi di Indonesia Berdasarkan Paket Pelayanan Stunting,” *PANRITA J. Sci. Technol. Arts*, vol. 1, no. 2, pp. 41–46, 2022, [Online]. Available: <https://journal.dedikasi.org/pjsta>
- [13] P. Apriyani, A. R. Dikananda, and I. Ali, “Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi,” *Hello World J. Ilmu Komput.*, vol. 2, no. 1, pp. 20–33, 2023, doi: 10.56211/helloworld.v2i1.230.
- [14] A. Winarta and W. J. Kurniawan, “Optimasi Cluster K-means Menggunakan Metode Elbow pada Data Pengguna Narkoba dengan Pemrograman Python,” *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- [15] Novita Agustina, “Faktor-faktor Penyebab Kejadian Stunting pada Balita,” yankes.kemkes.go.id. Accessed: Oct. 24, 2023. [Online]. Available: https://yankes.kemkes.go.id/view_artikel/1529/faktor-faktor-penyebab-kejadian-stunting-pada-balita
- [16] BPS, “Persentase Rumah Tangga yang Memiliki Akses Terhadap Sanitasi Layak (Persen), 2020-2022,” jabar.bps.go.id. Accessed: Oct. 24, 2023. [Online]. Available: <https://jabar.bps.go.id/indicator/29/728/1/persentase-rumah-tangga-yang-memiliki-akses-terhadap-sanitasi-layak.html>
- [17] Dinas Kesehatan, “Persentase Pemberian Air Susu Ibu (ASI) Eksklusif pada Bayi <6 Bulan Berdasarkan Kabupaten/Kota di Jawa Barat,” opendata.jabarprov.go.id. Accessed: Oct. 24, 2023. [Online]. Available: <https://opendata.jabarprov.go.id/id/dataset/persentase-pemberian-air-susu-ibu-asi-eksklusif-pada-bayi-6-bulan-berdasarkan-kabupatenkota-di-jawa-barat>
- [18] Dinas Kesehatan, “Persentase Bayi Menerima Inisiasi Menyusui Dini (IMD) Berdasarkan Kabupaten/Kota di Jawa Barat,” opendata.jabarprov.go.id. Accessed: Oct. 24, 2023. [Online]. Available: <https://opendata.jabarprov.go.id/id/dataset/persentase-bayi-menerima-inisiasi-menyusui-dini-imd-berdasarkan-kabupatenkota-di-jawa-barat>
- [19] D. Satriawan, “Pengelompokan Provinsi Di Indonesia Berdasarkan Faktor Penyebab Balita Stunting,” *Bul. Penelit. Sist. Kesehat.*, vol. 24, no. 4, pp. 308–317, 2021, doi: 10.22435/hsr.v24i4.4341.