Penerapan Metode K-Means Clustering Status Gizi Balita Di UPT Puskesmas Barong Tongkok

Vicky Pranandika Wijaksana[1)*], Ummul Hairah [2)], Masna Wati [3)], Novianti Puspitasari [4)], Anindita Septiarini [5)]
[1,2,3,4,5)]Program Studi Informatika, Universitas Mulawarman

Received: 15 July 2025 Accepted: 21 July 2025 Published: 28 July 2025

*vickypranandika@gmail.com

Kata Kunci: Status Gizi, anak balita, K-Means clustering

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Gizi pada anak balita merupakan masalah yang sangat penting untuk diperhatikan terutama bagi orang tua dan tenaga kesehatan. Status gizi balita dapat diketahui berdasarkan indeks Berat Badan menurut Umur dan Tinggi Badan menurut Umur. Penelitian ini bertujuan untuk mengidentifikasi pola-pola yang mungkin ada dalam status gizi balita dan mengidentifikasi kelompok balita yang berisiko tinggi atau berada dalam kondisi gizi yang buruk pada balita di kecamatan Barong Tongkok dengan penerapan *K-Means*. Data yang digunakan sebanyak 300 data yang akan dicluster menjadi 3 yaitu *Underweight*, Gizi Baik dan Gizi Lebih menggunakan metode perhitungan jarak *Euclidean Distance*, *Manhattan Distance* dan *Minkowski Distance*. Hasil pengujian *Sum Squared Error* (SSE) menunjukkan metode *Minkowski Distance* lebih unggul karena mendapatkan nilai *error* terkecil sebesar 815,4409. Sebanyak 133 Balita dalam kategori Gizi Baik (C1), 83 Balita dalam kategori Gizi Lebih (C2), dan 84 Balita dalam kategori *Underweight* (C3).

PENDAHULUAN

Status Gizi Balita merupakan salah satu predictor utama dalam kelangsungan hidup seorang anak. Malnutrisi adalah masalah kesehatan masyarakat yang utama dialami oleh sebagian besar Negara berkembang. Menurut Riset Kesehatan Dasar (Riskesdas) tahun 2018, prevalensi underweight, stunting, wasting berturut-turut ialah 17,8%, 30.8%, dan 10,24% Indonesia merupakan salah satu negara dengan prevalensi stunting cukup tinggi dibandingkan negara-negara berpendapatan menengah lainnya. Riskesdas melaporkan prevalensi stunting dari tahun ke tahun berturut turut dari tahun 2007, 2010, 2013 dan 2018 adalah 36.8%, 34.6% 37.2%, dan 30.8%.

K-Means Clustering merupakan salah satu metode data *clustering* non-hirarki yang mengelompokan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil. Metode *K-Means Clustering* merupakan metode yang paling popular dan banyak digunakan pada proses klasterisasi data. Metode *K-Means Clustering* membagi data menjadi beberapa kelompok data serta dapat menerima masukan berupa data tanpa label kelas [1][2].

Algoritma *K-Means* merupakan salah satu algoritma *clustering* yang banyak digunakan diberbagai bidang seperti bidang pendidikan, kesehatan, sosial, biologi, dan ilmu komputer. Algoritma *K Means* melakukan partisi data ke dalam satu atau lebih *cluster* yang memiliki ciri-ciri yang sama. Algoritma ini mampu menghasilkan akurasi atau tingkat ketelitian yang maksimal dengan sifatnya yang linear, waktu eksekusi algoritma ini relatif cepat dan mudah untuk diadaptasikan [3][4].

Oleh sebab itu, penelitian ini mengangkat judul "Penerapan Metode K-Means Clustering Status Gizi Balita Di UPT Puskesmas Barong Tongkok" menggunakan metode *K-means* pada status gizi balita. Hal ini bertujuan agar dapat mengidentifikasi pola-pola yang mungkin ada dalam status gizi balita dan mengidentifikasi kelompok balita yang berisiko tinggi atau berada dalam kondisi gizi yang buruk. Ini memungkinkan pihak berwenang dan peneliti untuk menyediakan perhatian khusus dan intervensi yang

lebih intensif.

TINJAUAN LITERATUR

Ni Komang Sri Julyantari melakukan penelitian *K-Means* Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih). Data yang digunakan dalam penelitian ini adalah status gizi balita dimana data tersebut berisi beberapa attribut yaitu umur, jenis kelamin dan berat badan yang nantinya digunakan sebagai input kemudian diproses sehingga menghasilkan *output* berupa kelompok atau *cluster*. Data sampel diperoleh dari tiga buah posyandu yaitu Titih Kelod, Titih Tengah dan Titih Kaler pada tanggal 21 Agustus 2021. Dimana data yang diperoleh berjumlah 46 balita. Dalam penelitian ini menggunakan proses mengelompokan status gizi balita ditemukan 3 *cluster* berupa Kurang, Normal, dan Lebih. Berdasarkan hasil pengelompokan data status gizi pada banjar Titih didapatkan 3 buah *cluster* dimana gizi buruk 30.43%, gizi normal 47.83 % dan gizi lebih 21.74%. Sehingga terdapat nilai yang cukup tinggi antara gizi buruk dan gizi lebih[5].

Abdi Subayu melakukan penelitian *K-Means* dalam Analisis Stunting Gizi Pada Balita: Systematic Review. Dalam Penelitian ini pengelompokan data Clustering menggunakan teknik *data mining*, dengan algoritma *K-Means*. Algoritma tersebut dipilih karena lebih sederhana. Pengumpulan data balita di bagi menjadi 3 yaitu Tinggi badan, Berat badan dan Lingkar kepala. Penelitian tentang Status Gizi pada balita yang dilakukan di wilayah Kota Tasikmalaya, dengan jumlah presensi angka Stunting di wilayah tersebut mencapai angka 5.290 Balita. Hasil Clustering menggunakan algoritma *K-Means* memperoleh 17 data dengan kelompok yang sama, maka hasil akurasi dari algoritma *K-Means* hanya sebesar 34% Benar (17/50 = 43%). Dari hasil tersebut dapat dikatakan algoritma *K-Means* tidak cocok digunakan untuk melakukan *Clustering* status Gizi pada Balita[6].

Alia Fadilah melakukan penelitian *K-Means* dalam Pengelompokan Kabupaten/Kota Di Indonesia Berdasarkan Faktor Penyebab Stunting Pada Balita dimana hasil dari pengelompokan ini akan diperoleh karakteristik faktor penyebab stunting pada balita di setiap cluster, sehingga hasil pengelompokan ini diharapkan dapat membantu pemerintah dalam mengambil kebijakan yang sesuai terkait penurunan prevalensi stunting pada balita berdasarkan karakteristik dan permasalahan masing-masing cluster. dapat disimpulkan bahwa penggunaan metode *elbow* untuk membantu algoritma *K-Means* dalam mengelompokan kabupaten/kota di Indonesia berdasarkan faktor penyebab stunting pada balita menghasilkan 2 *cluster* sebagai *cluster* terbaik dengan nilai selisih Sum of Square Error (SSE) sebesar 1401,5156, dimana *cluster* 1 merupakan *cluster* dengan faktor penyebab stunting tinggi yang terdiri dari 324 kabupaten/kota, dan *cluster* 2 merupakan *cluster* dengan faktor penyebab stunting rendah yang terdiri dari 49 kabupaten/kota. Dengan penelitian ini memberikan wawasan baru penerapan metode prediksi di wilayah lain[1].

METODE PENELITIAN

Data mining adalah bagian dari proses penemuan pengetahuan (*knowledge discovery processs*), dimana penemuan pengetahuan didefinisikan sebagai ilmu untuk menemukaln pola dalam set data. Data mining, secara sederhana merupakan suatu langkah ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan belum diketahui. Selain itu, data mining mempunyai hubungan dengan berbagai bidang diantaranya statistik, *machine learning* (pembelajaran mesin), *pattern recognition*, *computing algorithms*, *database technology*, dan *high performance computing* [7][8].

Data mining adalah ekstraksi pola yang menarik dari besarnya jumlah data. Strategi dikatakan menarik jika tidak sepele, implisit, asing, dan bermanfaat. Strategi yang disajikan harus mudah di mengerti, dapat digunakan dalam data yang bisa diperkirakan dengan tahap ketetapan tertentu, baru dan bermanfaat [9][10].

Clustering adalah metode yang membagi Kumpulan data menjadi jumlah *subgroup* atau *cluster* yang lebih kecil dan lebih mirip, Tujuan dalam deteksi *cluster* adalah untuk menemukan kesamaaan yang sebelumnya tidak diketahui di data. Peneglompokan data adalah cara yang sangat baik untuk memulai analisis data karena dapat memberikan titik awal untuk menemukan hubungan antar kelompok [11][12]. Clustering dapat mengindentifikasi daerah yang padat, menemukan pola-pola distribusi secara keseluruhan, dan menemukan keterkaitan yang menarik antara atribut-atribut data. Dalam *Data Mining*

usaha di fokuskan pada metode-metode penemuan untuk *cluster* pada basis data berukuran besar secara efektif dan efisien Hal penting dalam pengelompokan adalah menyatakan sekumpulan pola ke kelompok yang sesuai dengan yang berguna untuk menemukan kesamaan dan perbedaan sehingga dapat menghasilkan kesimpulan yang berharga [13][14].

Metode *K-Means* merupakan salah satu metode dalam fungsi clustering atau pengelompokan. Clustering mengacu pada pengelompokan data, observasi atau kasus berdasarkan kemiripan objek yang diteliti. Metodologi Dasar Analisa *K-Means Clustering K-Means* merupakan salah satu metode dalam fungsi clustering atau pengelompokan [15].

K-Means merupakan salah satu metode non- hirarki *clustering* yang mencoba membagi data yang ada menjadi satu atau lebih *cluster*. Algoritma mengelompokkan data ke dalam suatu *cluster* berdasarkan kemiripan atau kesamaan karakteristik data sedangkan data dengan karakteristik yang tidak mirip atau berbeda dikelompokkan ke dalam *cluster* yang lain. Sementara itu, *K-Means* merupakan salah satu metode pengelompokan non-hirarki yang bertujuan mengelompokkan objek sedemikian sehingga jarak-jarak tiap objek ke pusat kelompok dalam satu kelompok adalah minimum [16][17].

1. Menentukan jumlah k (*cluster*) yang akan menjadi acuan pengelompokkan data.
2. Menentukkan titk pusat (*centroid*) dari masing masing *cluster* yang ada.
3. Menghitung jarak terdekat masing masing data ke pusat *centroid*, dengan menggunakan salah satu persamaan jarak untuk menentukan jarak terdekat ke pusat *cluster* dengan persamaan jarak *Euclidean, Manhattan,* dan *Minkowski.*
4. Menentukan nilai *centroid* baru pada setiap dengan *cluster* yang ada, dengan menghitung rata rata (*means*) dari data setiap rumus persamaan (1).
5. Mengulangi Langkah 3 sampai 5 hingga tidak ada record yang berpindah *cluster* atau konvergen.

$$Cluster\ Center = \sum \frac{ai}{n} \quad \text{.....(2.1)}$$

a. Euclidean Distance

Euclidean distance merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari dua buah titik dalam *Euclidean space* (meliputi bidang dua dimensi, tiga dimensi, atau bahkan lebih. Pengukuran jarak dilakukan dengan rumus *Euclidean distance* pada persamaan sebagai berikut:

$$d(x1, x2) = ||x1\text{-}x2|| = \sqrt{\sum_{i=1}^{n}(x1i - x2i)^2} \quad \text{.....(2.2)}$$

Dimana $x1$ data latih (*training*) dan $x2$ adalah data uji.

b. Manhattan Distance

Manhattan distance digunakan untuk menghitung perbedaan nilai absolut (mutlak) antara koordinat sepasang objek. Rumus persamaan yang dapat digunakan untuk menghitung jarak dengan *Manhattan distance* dapat dilihat pada persamaan sebagai berikut:

$$d_{manhattan}(x, y) = \sum_{i}^{n} |X_i - Y_i| \quad \text{.....(2.3)}$$

Keterangan:

d = jarak antara *centroid*

i = data ke -i

n = jumlah data

X_i = data pada setiap pusat ke i

Y_i = data pada setiap *cluster* ke i

c. Minkowski Distance

Minkowski distance merupakan sebuah metrik dalam ruang vektor dimana suatu norma didefinisikan (*normed vector space*) sekaligus dianggap sebagai generalisasi dari *Euclidean distance* dan *Manhattan distance*. Dalam pengukuran jarak objek menggunakan *Minkowski distance* umumnya digunakan nilai p adalah 1 atau 2. Berikut merupakan rumus persamaan yang dapat digunakan untuk menghitung jarak dapat dilihat pada persamaan sebagai berikut:

$$d_{minkowski}(x, y) = \left(\sum_{i}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}} \quad \text{.....(2.4)}$$

Keterangan:

d = jarak antara data dengan *centroid*

i = data ke-i

n = jumlah data
x_i = data pada setiap data ke i
y_i = data pada pusat *cluster* ke i
p = power.

Sedangkan untuk menentukan *centroid* baru dapat dilakukan dengan mengambil rata-rata dokumen yang masuk pada *cluster* awal. Iterasi terus dilakukan hingga posisi group tidak berubah. Berikut penentuan *centroid* menggunakan rumus 2.5.

$$c_i = \frac{x1+x2+x3+x4+\cdots+xi}{\sum x} \quad \text{.....(2.5)}$$

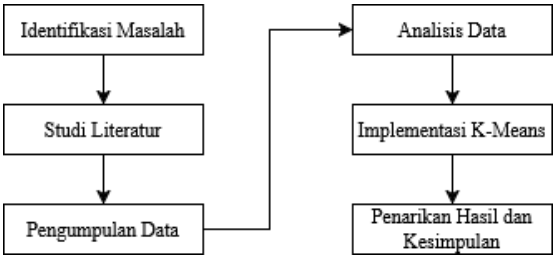
Keterangan:
c_i = jarak antara data dengan centroid
$x1$ = nilai data *record* ke 1
$x2$ = nilai data *record* ke 2
$\sum x$ = jumlah data *record*

Dalam metode *K-Means*, pengelompokkan data dilakukkan dengan cara mengelompokkan atribut data ke dalam beberapa *cluster* berdasarkan kemiripan dari data tersebut. Kemudian, kemiripan siatu kelompok data diukur menggunakan suatu metode pengukuran jarak. Metode pengukuran jarak yang digunakkan adalah *Sum of Squared Error* (SSE). Metode ini akan memberikan informasi error jarak data ke *centroid*. Jika semakin kecil nilai SSE maka semakin bagus hasil klastering yang dibuat. Berikut definisi formula untuk menghitung SSE [2].

$$SSE = \sum_{k=1}^{k} \sum_{p \in C_i} d = (p, m_i)^2 \text{.....(2.6)}$$

Keterangan:
$p \in C_i$ = tiap data poin pada *cluster* i;
m_i = *centroid* dari *cluster* i;
d = jarak / *distance* / *variance* terdekat pada masing masing *cluster* i

Tahapan penelitian merupakan langkah-langkah yang sistematis dan terstruktur yang dilakukan untuk mencapai tujuan penelitian.



Gambar 1. Tahapan Pelaksanaan Penelitian

6. Perancangan Data

Data yang digunakan adalah data status gizi balita dengan total 300 sampel data yang didapat dati UPT Puskesmas Barong Tongkok, dengan atribut yang digunakan yaitu Berat Badan (BB) dan Tinggi Badan (TB).

Tabel 1. Data Balita

NO	Nama	Jenis Kelamin	Tgl Lahir	…	Berat Badan (BB)	Tinggi Badan (TB)
1	Balita 1	L	2022-08-21	…	2,6	49
2	Balita 2	L	2020-10-04	…	9,8	83
3	Balita 3	P	2020-06-13	…	10,4	84,8
4	Balita 4	P	2018-12-19	…	14,8	95
5	Balita 5	P	2019-03-15	…	24	100
⋮	⋮	⋮	⋮	…	⋮	⋮
300	Balita 300	P	2020-07-15	…	10,2	82,1

7. Perancangan Pengujian

   Untuk perancangan pengujian dilakukan dengan data yang didapat dari perhitungan jarak *Euclidean Distance*, *Manhattan Distance*, *Minkowski Distance*. Data tersebut kemudian diuji menggunakan *Sum Squared Error* (SSE) untuk mengukur kualitas cluster dimana pengelompokkan data dilakukan dengan cara mengelompokkan atribut data ke dalam beberapa *cluster* berdasarkan kemiripan dari data tersebut.

K-Means bekerja optimal pada data numerik seperti pengukuran antropometri (berat badan, tinggi badan). Data ini menjadi dasar untuk menentukan kelompok berdasarkan rentang tertentu sehingga output *K-Means* memberikan informasi tentang *centroid* (pusat klaster), yang dapat diartikan sebagai rata-rata karakteristik balita dalam satu klaster. Ini mempermudah identifikasi kategori gizi.

a. Pengumpulan data

   Pengumpulan data diperlukkan sebagai bahan pemecahan masalah yang telah dirumuskan sebelumnya. Data yang digunakan berupa data hasil observasi UPT Puskesmas Barong Tongkok yang berupa data status gizi pada balita.

b. Analisis data

   Analisa dan mengidentifikasi data yang telah dikumpulkan agar dapat lebih mudah untuk diproses sesuai dengan metode K-Means Clustering sehingga data tersebut dapat digunakan untuk hasil perhitungan yang lebih baik.

c. Implementasi K - Means

   Pengolahan data pada penelitian ini akan diola menggunakan *data mining*, *data mining* adalah teknologi yang merupakan campuran teknik-teknik analisis data dengan algoritma- algoritma untuk memproses data berukuran besar.

d. *Pre-Processing*

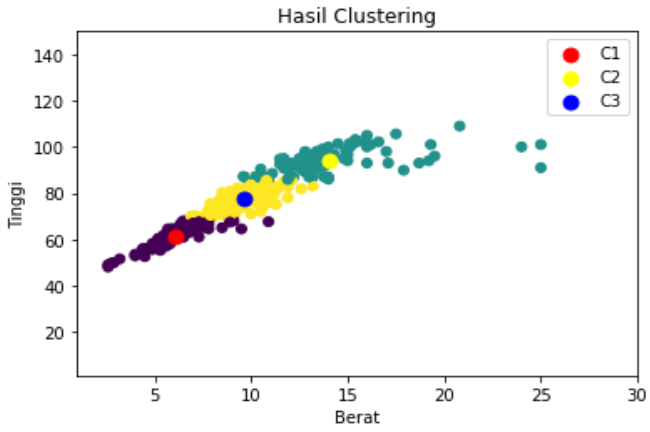
   Pre-processing berupa selection data, transformasi data dan melakukan normalisasi Z-*score*. Dari data status gizi pada balita di Barong tongkok perlu dilakukan seleksi atribut apa saja yang diggunakan yaitu berupa Tinggi Badan (TB) dan Berat Badan (BB) pada balita.

e. Clustering *K-means*

   Tahap selanjutnya melakukkan analisis dengan menggunakan metode *K-means Clustering* dengan menggunakan metode pengukuran jarak *Euclidean Distance*, *Manhattan Distance* dan *Minkowski Distance*. *Clustering* dilakukkan dengan *Microsoft Excel* dan juga menggunakkan *software* untuk mengurangi dampak *human error* saat perhitungan manual.

8. Perancangan Tampilan

   Perancangan Tampilan Status gizi balita yang sudah di cluster menggunakan bahasa pemrograman python dengan library dari sklearn dan pandas bisa dilihat di Gambar 2 sebagai berikut.



Gambar 2. Tampilan Hasil *Clustering* Status Gizi Balita

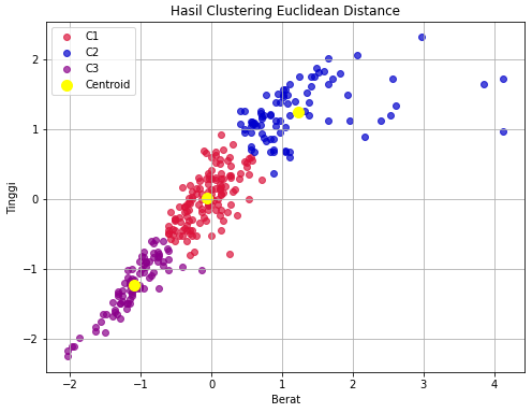
Pada Gambar 2 menampilkan sebuah data plot hasil clustering status gizi balita menggunakan aplikasi Spyder dimana status gizi balita dibagi menjadi 3 cluster yaitu *Underweight*, Gizi Baik dan Gizi Lebih.

HASIL PENELITIAN

Berdasarkan perancangan data, data yang digunakan adalah data Balita Kecamatan Barong Tongkok yang bersumber dari UPT Puskesmas Barong Tongkok. Data diolah melalui beberapa tahap *Pre-Processing* agar data dapat digunakan dalam penerapan *Clustering K-Means*.

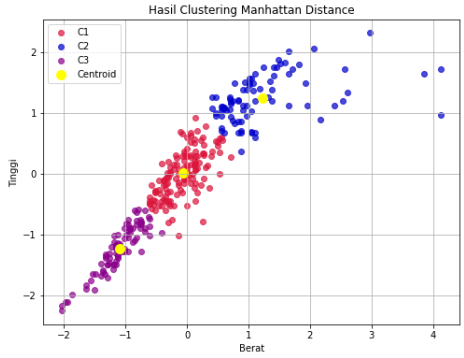
1. Hasil Pengolahan Data

Hasil *clustering* yang didapatkan dan perhitungan jarak menggunakan metode *Euclidean Distance*, *Manhattan Distance*, dan *Mankowski Distance* ditampilkan dalam bentuk plot.



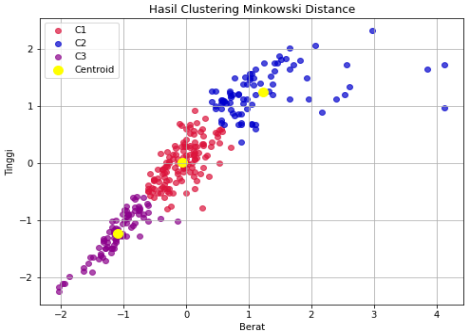
Gambar 3. Hasil *Clustering Euclidean Distance*

Hasil *clustering* yang didapatkan dari perhitungan jarak menggunakan metode *Euclidean Distance* yaitu terdapat 133 balita dalam kategori Gizi Baik (C1), 83 balita dalam kategori Gizi Lebih (C2), dan 84 balita dalam kategori *Underweight* (C3).



Gambar 4. *Hasil Clustering Manhattan Distance*

Hasil *clustering* yang didapatkan dan perhitungan jarak menggunakan metode *Manhattan Distance* yaitu terdapat 134 balita dalam kategori Gizi Baik (C1), 83 balita dalam kategori Gizi Lebih (C2), dan 83 balita dalam kategori *Underweight* (C3).



Gambar 5. Hasil *Clustering Minkowski Distance*

Hasil *clustering* yang didapatkan dan perhitungan jarak menggunakan metode *Mankowski Distance* yaitu terdapat 133 balita dalam kategori Gizi Baik (C1), 83 balita dalam kategori Gizi Lebih (C2), dan 84 balita dalam kategori *Underweight* (C3).

a. Perhitungan Euclidean Distance

Perhitungan SSE metode *Eucidean Distance* yang dilakukan adalah mencari nilai rata-rata hasil iterasi akhir penentuan jarak C1, C2, C3.

$$\bar{X}_1 = \frac{2{,}9379 + 0{,}3616 + 0{,}5278 + \cdots + 0{,}3196}{300} = 1{,}1536$$

$$\bar{X}_2 = \frac{4{,}7199 + 1{,}5455 + 1{,}3334 + \cdots + 1{,}4971}{300} = 1{,}9419$$

$$\bar{X}_3 = \frac{1{,}3226 + 1{,}9184 + 2{,}1224 + \cdots + 1{,}9265}{300} = 1{,}7921$$

Didapatkan hasil nilai rata-rata pada tiap kolom indicator SSE yaitu C1 = 1,1536, C2 = 1,9419, C3 = 1,7921. Langkah selanjutnya yaitu mencari SSE tiap cluster jarak seperti berikut:

$d_{11} = (2{,}9379 - 1{,}1536)^2 + (0{,}3616 - 1{,}1536)^2 + \cdots + (0{,}3196 - 1{,}1536)^2 = 202{,}1411$

$d_{12} = (4{,}7199 - 1{,}9387)^2 + (1{,}5455 - 1{,}9387)^2 + \cdots + (1{,}4971 - 1{,}9387)^2 = 379{,}2023$

$d_{13} = (1{,}3226 - 1{,}7948)^2 + (1{,}9184 - 1{,}7948)^2 + \cdots + (1{,}9265 - 1{,}7948)^2 = 459{,}8156$

Nilai SSE dari metode *Euclidean Distance* yang didapatkan yaitu:

$$SSE = 202{,}1411 + 379{,}2023 + 459{,}8156 = 1041{,}159$$

b. Perhitungan *Manhattan Distance*

Perhitungan SSE metode *Manhattan Distance* yang dilakukan adalah mencari nilai rata-rata hasil iterasi akhir penentuan jarak C1, C2, C3.

$$\bar{X}_1 = \frac{4{,}1491 + 0{,}3766 + 0{,}6764 + \cdots + 0{,}4193}{300} = 1{,}5884$$

$$\bar{X}_2 = \frac{6{,}6732 + 2{,}1474 + 1{,}8476 + \cdots + 2{,}1048}{300} = 2{,}7038$$

$$\bar{X}_3 = \frac{1{,}8705 + 2{,}6552 + 2{,}955 + \cdots + 2{,}697}{300} = 2{,}5026$$

Didapatkan hasil nilai rata-rata pada tiap kolom indicator SSE yaitu C1 = 1,5884, C2 = 2,7038, C3 = 2,5026. Langkah selanjutnya yaitu mencari SSE tiap cluster jarak seperti berikut:

$d_{11} = (4{,}1491 - 1{,}5884)^2 + (0{,}3766 - 1{,}5884)^2 + \cdots + (0{,}4193 - 1{,}5884)^2$
$= 390{,}6499$

$d_{12} = (6{,}6732 - 2{,}7038)^2 + (2{,}1474 - 2{,}7038)^2 + \cdots + (2{,}1048 - 2{,}7038)^2$
$= 778{,}5292$

$d_{13} = (1{,}8705 - 2{,}5026)^2 + (2{,}6552 - 2{,}5026)^2 + \cdots + (2{,}697 - 2{,}5026)^2$
$= 904{,}7346$

Nilai SSE dari metode *Manhattan Distance* yang didapatkan yaitu:

$$SSE = 390{,}6499 + 778{,}5292 + 904{,}7346 = 2073{,}9137$$

c. Perhitungan *Minkowski Distance*

Perhitungan SSE metode *Minkowski Distance* yang dilakukan adalah mencari nilai rata-rata hasil iterasi akhir penentuan jarak C1, C2, C3.

$$\bar{X}_1 = \frac{2{,}6209 + 0{,}3613 + 0{,}5038 + \cdots + 0{,}3014}{300} = 1{,}0481$$

$$\bar{X}_2 = \frac{4{,}2061 + 1{,}3998 + 1{,}2104 + \cdots + 1{,}3415}{300} = 1{,}7482$$

$$\bar{X}_3 = \frac{1{,}1784 + 1{,}7432 + 1{,}9189 + \cdots + 1{,}7326}{300} = 1{,}6143$$

Didapatkan hasil nilai rata-rata pada tiap kolom indicator SSE yaitu C1 = 1,0481, C2 = 1,7482, C3 = 1,6143. Langkah selanjutnya yaitu mencari SSE tiap *cluster* jarak seperti berikut:

$d_{11} = (2{,}6209 - 1{,}0481)^2 + (0{,}3613 - 1{,}0481)^2 + \cdots + (0{,}3014 - 1{,}0481)^2$
$= 166{,}3098$

$d_{12} = (4{,}2061 - 1{,}7364)^2 + (1{,}3998 - 1{,}7364)^2 + \cdots + (1{,}3415 - 1{,}7364)^2$
$= 297{,}0656$

$d_{13} = (1{,}1784 - 1{,}7482)^2 + (1{,}7432 - 1{,}7482)^2 + \cdots + (1{,}7326 - 1{,}7482)^2$
$= 371{,}3622$

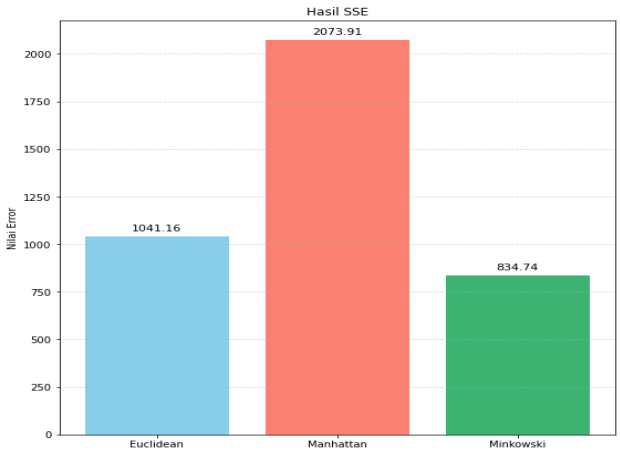
Nilai SSE dari metode *Minkowski Distance* yang didapatkan yaitu:

$$SSE = 166{,}3098 + 297{,}0656 + 371{,}3622 = 834{,}7376$$

Tabel 2. Tabel Hasil SSE.

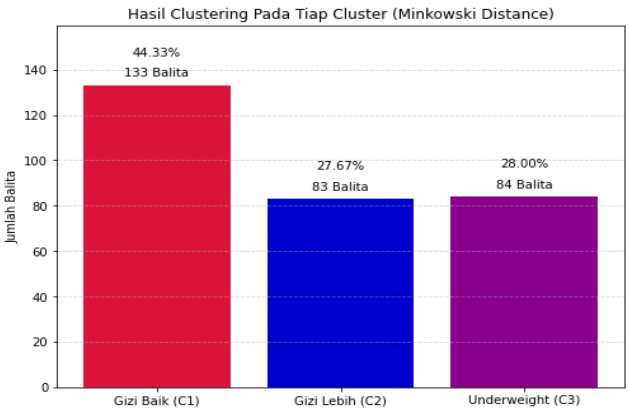
Distance	*Euclidean*	*Manhattan*	*Minkowski*
Hasil SSE	1041,159	2073,9137	834,7376

Dapat dilihat pada tabel 2 di atas, didapatkan nilai SSE Euclidean 1041,159, Manhattan 2073,9137, dan Minkowski 834,7376.



Gambar 6. Hasil *Clustering Minkowski Distance*

Pada Gambar 6 didapatkan nilai SSE paling rendah berada pada metode penentuan jarak *Minkowski Distance*, yang dimana pada pada pengujian SSE semakin kecil nilai yang dihasilkan atau mendekati 0 menandakan bahwa model tersebut memiliki nilai kesalahan acak terkecil. Kemudian hasil *clustering* akan menggunakan *Minkowski Distance* sebagai hasil pengujian.



Gambar 7. *Hasil Clustering* Tiap Cluster *Manhattan Distance*

Pada Gambar 7 diperoleh bahwa perhitungan menggunakan metode *Minkowski Distance* menghasilkan

jumlah 133 atau 44.33% balita dalam kategori Gizi Baik (C1), 83 atau 27.67% balita dalam kategori Gizi Lebih (C2), dan 84 atau 28% balita dalam kategori *Underweight* (C3).

PEMBAHASAN

Berdasarkan Penelitian yang telah dilakukan pada sub-bab sebelumnya dengan 3 metode penentuan jarak yang terdiri dari *Euclidean distance*, *Manhattan Distance*, dan *Minkowski Distance* untuk mencari 3 kategori *cluster* dari Gizi Baik (C1), Gizi Lebih (C2), dan *Underweight* (C3), diperoleh bahwa perhitungan menggunakan metode *Euclidean Distance* menghasilkan jumlah 133 atau 44.3% balita dalam kategori Gizi Baik (C1), 83 atau 27.7% balita dalam kategori Gizi Lebih (C2), dan 84 atau 28% balita dalam kategori *Underweight* (C3). Sementara untuk perhitungan *Manhattan Distance* menghasilkan jumlah 134 atau 44.6% balita dalam kategori Gizi Baik (C1), 83 atau 27.7% balita dalam kategori Gizi Lebih (C2), dan 83 atau 27.7% balita dalam kategori *Underweight* (C3). Sementara untuk perhitungan *Minkowski Distance* menghasilkan jumlah 133 atau 44.3% balita dalam kategori Gizi Baik (C1), 83 atau 27.7% balita dalam kategori Gizi Lebih (C2), dan 84 atau 28% balita dalam kategori *Underweight* (C3).

Pada penelitian ini, juga dilakukan hasil pengujian menggunakan metode *Sum Squared Error* untuk mengetahui jumlah *cluster* optimal dari *distance measure* yang telah terpilih dari hasil pengujian akurasi yang dilakukan sebelumnya, untuk mengethui metode perhitungan jarak terbaik yang optimal. Pada pengujian SSE nilai pengujian yang mendekati nol atau semakin kecil menandakan bahwa midel tersebut mempunyai nilai kesalahan acak terkecil.

Hasil pengujian SSE cluster untuk metode *Euclidean Distance* bernilai 1037,6888, *Manhattan Distance* bernilai 2078,833, dan *Minkowski Distance* 815,4409. Dari hasil yang telah didapatkan diambil kesimpulan bahwa untuk pengujiaan SSE, metode *Minkowski Distance* lebih unggul karena mendapatkan nilai *error* terkecil. Dengan demikian, berdasarkan hasil penelitian yang telah dilakukan dari 3 metode pengukuran jarak yang direkomendasikan adalah *Minkowski Distance* dengan nilai akurasi yang lebih unggul dari *Euclidean Distance* dan *Manhattan Distance*.

KESIMPULAN

Cluster optimal untuk Algoritma *K-Means clustering* adalah 3 Cluster. Hasil *clustering* menggunakan *Minkowski Distance* terdapat jumlah 133 atau 44.3% balita dalam kategori Gizi Baik (C1), 83 atau 27.7% balita dalam kategori Gizi Lebih (C2), dan 84 atau 28% balita dalam kategori *Underweight* (C3). Hasil perhitungan *Sum Squared Error* dari ketiga metode penentuan jarak yang digunakan, nilai hasil error terkecil terdapat pada metode penentuan jarak *Minkowski Distance* yang bernilai 815,4409.

REFERENCES

[1] 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, 2022, doi: 10.26798/jiko.v6i2.581.

[2] F. Naufal, Y. Herry Chrisnanto, and A. Kania Ningsih, "Sistem Rekomendasi Penawaran Produk Pada Online Shop Menggunakan K-Means Clustering," *Informatics Digit. Expert*, vol. 4, no. 1, 2022, doi: 10.36423/index.v4i1.879.

[3] M. Mursalim, P. Purwanto, and M. A. Soeleman, "Penentuan Centroid Awal Pada Algoritma K-Means Dengan Dynamic Artificial Chromosomes Genetic Algorithm Untuk Tuberculosis Dataset," *Techno.Com*, vol. 20, no. 1, 2021, doi: 10.33633/tc.v20i1.4230.

[4] R. G. Solechati and A. Jananto, "PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA DATA BRAIN STROKE UNTUK PENGELOMPOKAN PROFILE PASIEN," *semanTIK*, vol. 9, no. 1, 2023, doi: 10.55679/semantik.v9i1.29446.

[5] Ni Komang Sri Julyantari, I. K. Budiarta, and N. M. D. K. Putri, "Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih)," *J. Janitra Inform. dan Sist. Inf.*, vol. 1, no. 2, 2021, doi: 10.25008/janitra.v1i2.134.

[6] A. Subayu, "PENERAPAN METODE K-MEANS UNTUK ANALISIS STUNTING GIZI PADA BALITA: SYSTEMATIC REVIEW," *J. Sains, Nalar, dan Apl. Teknol. Inf.*, vol. 2, no. 1, 2022, doi: 10.20885/snati.v2i1.18.

[7] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, "Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 16, no. 1, 2021, doi: 10.30872/jim.v16i1.5182.

[8] Ramadhana, Islamiyah, and A. P. A. Masa, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara," *Adopsi Teknol. dan Sist. Inf.*, vol. 2, no. 1, 2023, doi: 10.30872/atasi.v2i1.595.

[9] E. Ramadanti and M. Muslih, "PENERAPAN DATA MINING ALGORITMA K-MEANS CLUSTERING PADA POPULASI AYAM PETELUR DI INDONESIA," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 1, 2022, doi: 10.36341/rabit.v7i1.2155.

[10] T. B. Pamungkas, S. Maesaroh, and P. Ardiansyah, "IMPLEMENTASI DATA MINING PADA STOK PENGGUNAAN BARANG DI GMF AEROASIA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," *J. Ilm. Sains dan Teknol.*, vol. 7, no. 2, 2023, doi: 10.47080/saintek.v7i2.2697.

[11] I. N. M. Adiputra, "Clustering Penyakit DBD pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means," *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 2, no. 2, 2022, doi: 10.23887/insert.v2i2.41673.

[12] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," *Techno.Com*, vol. 20, no. 2, 2021, doi: 10.33633/tc.v20i2.4556.

[13] I. Pii, N. Suarna, and N. Rahaningsih, "PENERAPAN DATA MINING PADA PENJUALAN PRODUK PAKAIAN DAMEYRA FASHION MENGGUNAKAN METODE K-MEANS CLUSTERING," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, 2023, doi: 10.36040/jati.v7i1.6336.

[14] M. Fithriyah, M. A. Yaqin, and S. Zaman, "K-Means Clustering Untuk Segmentasi Produk Berdasarkan Analisis Recency, Frequency, Monetary (RFM) Pada Data Transaksi Penjualan," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 3, no. 2, 2021, doi: 10.28926/ilkomnika.v3i2.284.

[15] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Cigugur Tengah," *J. Inf. Technol.*, vol. 2, no. 2, 2020, doi: 10.47292/joint.v2i2.30.

[16] A. Septiarini, I. A. Thaher, and N. Puspitasari, "Pengelompokan Kualitas Kinerja Pegawai Menggunakan Metode K-Means Clustering," *Komputika J. Sist. Komput.*, vol. 11, no. 2, 2022, doi: 10.34010/komputika.v11i2.5518.

[17] F. P. Ferdy Pangestu, N. Y. Nur Yasin, R. C. Ronald Chistover Hasugian, and Y. Yunita, "Penerapan Algoritma K-Means Untuk Mengklasifikasi Data Obat," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, 2023, doi: 10.32736/sisfokom.v12i1.1461.