Pengelompokan Makanan Berdasarkan Komposisi Zat Gizi Menggunakan Metode *Agglomerative Hierarchical Clustering*

Firda Siti Hanifah *, Marizsa Herlina

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

firdasth@email.com, marizsa.herlina@unisba.co.id

Abstract. To meet daily nutritional needs, humans require a balanced diet. According to the 2023 Indonesian Health Survey, 82.4% of Indonesian adults have a non-ideal nutritional status (based on body mass index), with 7.8% categorized as wasting, 14.4% as overweight, 23.4% as obese, and 36.8% as centrally obese. This study aims to group food data based on nutrient composition to make it easier for the Indonesian public to understand, thereby helping them achieve a more ideal nutritional status. The study was conducted using the Agglomerative Hierarchical Clustering method with five linkage techniques. The evaluation of the best clustering model using the cophenetic correlation coefficient showed a correlation of 0.907, and the representative number of clusters based on the silhouette coefficient was two clusters. The result is cluster 1 consists of foods rich in protein and carbohydrates but low in fat and calories, while cluster 2 consists of foods high in fat and calories but low in protein and carbohydrates.

Keywords: *Agglomerative Hierarchical Clustering, Food Composition, Nutritious Food.*

Abstrak. Agar dapat memenuhi asupan gizi harian, manusia membutuhkan gizi yang seimbang. Menurut Survei Kesehatan Indonesia tahun 2023, 82,4% masyarakat Indonesia yang berusia dewasa memiliki status gizi tidak ideal (berdasarkan indeks massa tubuh), dengan 7,8% dikategorikan sebagai *wasting*, 14,4% sebagai *overweight*, 23,4% sebagai obesitas, dan 36,8% sebagai obesitas sentral. Penelitian ini bertujuan untuk mengelompokkan data makanan berdasarkan komposisi zat gizi agar mudah dipahami oleh masyarakat Indonesia sehingga masyarakat Indonesia dapat memiliki status gizi yang lebih ideal. Penelitian ini dilakukan dengan menggunakan metode *Agglomerative Hierarchical Clustering* dengan 5 metode *linkage*. Hasil evaluasi model *clustering* terbaik menggunakan koefisien korelasi *cophenetic* menunjukkan korelasi sebesar 0,907 dan jumlah *cluster* yang representatif berdasarkan koefisien *silhouette* sebanyak 2 *cluster*. Hasilnya adalah *cluster* 1 terdiri dari makanan kaya protein dan karbohidrat namun rendah lemak dan kalori, sedangkan *cluster* 2 terdiri dari makanan yang tinggi lemak dan kalori namun rendah protein dan karbohidrat.

Kata Kunci: *Agglomerative Hierarchical Clustering, Komposisi Pangan, Makanan Bergizi.*

*firdasth@email.com

A. Pendahuluan

Pangan adalah bahan yang dikonsumsi sehari-hari oleh manusia, berasal dari sumber hayati maupun air, dan menjadi kebutuhan dasar utama bagi manusia [1]. Manusia membutuhkan makanan yang bergizi seimbang setiap harinya untuk memenuhi asupan gizi harian, yaitu mencakup kebutuhan nutrisi makro dan kebutuhan nutrisi mikro.

Menurut Survei Kesehatan Indonesia tahun 2023, status gizi masyarakat Indonesia pada usia dewasa berdasarkan indeks massa tubuh sebesar 7,8% pada kategori *wasting*, kemudian sebesar 14,4% pada kategori *overweight*, dan sebesar 23,4% pada kategori obesitas. Sementara itu, 36,8% masyarakat Indonesia pada usia dewasa mengalami obesitas sentral [2]. Sebagai upaya untuk meningkatkan kesadaran dan pemahaman masyarakat Indonesia tentang seberapa penting asupan gizi yang benar dan membantu masyarakat Indonesia memenuhi kebutuhan gizi harian dengan lebih mudah, Kementerian Kesehatan Republik Indonesia mengeluarkan panduan kebutuhan gizi harian yang seimbang, yaitu "isi piringku". Panduan ini memberikan panduan tentang jenis makanan dan minuman yang dikonsumsi sekali makan dengan takaran atau porsi makanan yang dapat memenuhi kebutuhan asupan gizi harian [3]. Berdasarkan hal tersebut, diperlukan pengelompokan komposisi gizi pada makanan yang mudah dipahami oleh masyarakat Indonesia.

Pada penelitian pendahulu telah mengimplementasikan metode K-Means untuk pengelompokan data makanan berdasarkan nilai nutrisi yang serupa(Herliani & Kudus, 2023) Penelitian tersebut menghasilkan 3 cluster yang berbeda. Cluster 0 terdiri dari makanan dengan kandungan kalori serta karbohidrat yang tinggi, namun rendah protein dan lemak. Cluster 1 berisi makanan dengan kalori tinggi, protein cukup tinggi, serta kadar lemak dan karbohidrat yang rendah. Sementara itu, cluster 2 mencakup makanan dengan kandungan nutrisi yang rendah [4].

Pada penelitian ini, peneliti menggunakan metode *Agglomerative Hierarchical Clustering* (AHC) yang terdiri dari *single linkage*, *complete linkage*, *average linkage*, *centroid linkage*, dan *ward's minimum variance*. Metode ini memiliki suatu pendekatan yang disebut *bottom-up* yakni proses pengelompokan data dimulai dari seluruh objek menjadi masing-masing *cluster* hingga seluruh objek berada dalam 1 *cluster* besar. Berikutnya, tujuan dari penelitian ini diuraikan dalam pokok-pokok sebagai berikut:

1. Melakukan pengelompokan terhadap data makanan berdasarkan komposisi zat gizi menggunakan metode *Agglomerative Hierarchical Clustering* (AHC).
2. Mengetahui makanan yang termasuk aman untuk dikonsumsi sehari-hari berdasarkan hasil metode terbaik.

B. Metode

Data yang digunakan dalam penelitian ini adalah data sekunder yakni data Tabel Komposisi Pangan Indonesia (TKPI) tahun 2017 yang dipublikasikan oleh Kementerian Kesehatan Republik Indonesia [5]. Variabel penelitian yang digunakan adalah kandungan nutrisi makro yakni energi, protein, lemak, dan karbohidrat dan analisis data dilakukan menggunakan *software* Microsoft Excel dan R Studio.

Langkah Analisis Data

1. Input data kandungan nutrisi makro dari data Komposisi Pangan Indonesia.
2. Melakukan standarisasi data karena data yang digunakan memiliki satuan yang berbeda dengan *z-score* [6].

$$Z = \frac{x_{ij} - \bar{x}_i}{s_i} \quad (1)$$

3. Melakukan perhitungan ukuran jarak untuk mengukur kemiripan antar objek menggunakan jarak *Euclidean* [7].

$$d_{ik} = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{kj})^2} \quad (2)$$

4. Menggabungkan objek x dan y yang memiliki jarak terdekat antar objek menjadi satu kelompok. Kemudian, dari cluster xy yang sudah terbentuk, dihitung jarak dengan objek lain

yang belum bergabung [8].

a. *Single linkage*

$$d_{(UV)W} = min(d_{UW}, d_{VW}) \quad (3)$$

b. *Complete linkage*

$$d_{(UV)W} = max(d_{UW}, d_{VW}) \quad (4)$$

c. *Average linkage*

$$d_{(UV)W} = \frac{\sum_i \sum_j d_{ij}}{n_{(UV)} n_W} \quad (5)$$

d. *Centroid linkage*

$$d_{(UV)W} = \frac{n_U d_{(UW)} + n_V d_{(VW)}}{n_{(UV)}} - \frac{n_U n_V d_{(UV)}}{n_{(UV)}^2} \quad (6)$$

e. *Ward's minimum variance*

$$d_{(UV)W} = SSE(d_{UW}, d_{VW}) = \frac{[(n_W + n_U) d_{(UW)} + (n_W + n_V) d_{(VW)}] - n_W d_{(UV)}}{n_W + n_{(UV)}} \quad (7)$$

Memperbaharui matriks ukuran jarak setiap mendapatkan pengelompokan baru. Melakukan tahapan tersebut secara berulang hingga semua objek berada pada cluster yang sama.

5. Mengevaluasi kualitas dari masing-masing hasil pengelompokan untuk mendapatkan metode terbaik yang dapat dikatakan metode tersebut dapat membentuk dendrogram yang paling optimal dengan hasil pengelompokan lebih baik dalam merepresentasikan data asli menggunakan koefisien korelasi *Cophenetic* [9].

$$r_{coph} = \frac{\sum_{i<k} (d_{ik} - \bar{d})(d_{C_{ik}} - \bar{d}_C)}{\sqrt{[\sum_{i<k} (d_{ik} - \bar{d})^2][\sum_{i<k} (d_{C_{ik}} - \bar{d}_C)^2]}} \quad (8)$$

6. Melakukan validitas *cluster* dari metode yang telah dipilih untuk mengetahui jumlah *cluster* yang optimal dari metode tersebut menggunakan koefisien *Silhouette* [10].

$$S_i = \frac{(b_i - a_i)}{max(a_i, b_i)} \quad (9)$$

Menurut Kaufman dan Rousseeuw (1990) interpretasi subjektif kualitas dari koefisien *silhouette* ditunjukkan pada Tabel 1.

Tabel 1. Interpretasi Subjektif Kualitas dari Koefisien *Silhouette*

Koefisien *Silhouette*	Interpretasi
0,71–1,00	*Cluster* yang kuat
0,51–0,70	*Cluster* yang baik
0,26–0,50	*Cluster* yang lemah
≤0,25	*Cluster* yang buruk

7. Melakukan profilisasi dari hasil pengelompokan terbaik.

C. Hasil Penelitian dan Pembahasan

Melakukan Standarisasi Data

Dikarenakan data yang digunakan memiliki satuan yang berbeda maka diperlukan standarisasi data. Sebagai contoh perhitungan *z-score* setiap variabel pada data beras giling (Z_1) sebagai berikut:

$$Z_{1,energi} = \frac{357 - 198{,}4}{162{,}4} = 0{,}98$$

$$Z_{1,protein} = \frac{357 - 9{,}67}{11{,}5} = -0{,}11$$

$$Z_{1,lemak} = \frac{357 - 7{,}39}{13{,}7} = -0{,}42$$

$$Z_{1,karbohidrat} = \frac{357 - 23{,}76}{25} = 2{,}13$$

Perhitungan *z-score* dengan data beras giling pada setiap variabel menghasilkan nilai *z-score* seperti pada perhitungan diatas. Ringkasan hasil standarisasi data sebagai berikut:

Tabel 2. Hasil Standarisasi Data

No.	Energi	Protein	Lemak	Karbohidrat
1.	0,98	-0,11	-0,42	2,13
2.	1,05	-0,01	-0,44	2,13
3.	0,98	-0,11	-0,42	2,13
4.	0,94	-0,14	-0,44	2,13
5.	0,98	-0,36	-0,53	2,36
⋮	⋮	⋮	⋮	⋮
1146.	-1,12	-0,82	-0,53	-0,80

Melakukan Perhitungan Ukuran Jarak

Mengukur kedekatan antar objek memerlukan perhitungan ukuran jarak. Pada penelitian ini, pengukuran tersebut menggunakan jarak *Euclidean*. Sebagai contoh, perhitungan jarak kedekatan antara data makanan beras giling (objek kesatu) dengan beras giling varian pelita (objek kedua) sebagai berikut:

$$d_{1,2} = \sqrt{(x_{11} - x_{21})^2 + \ldots + (x_{14} - x_{24})^2}$$
$$d_{1,2} = \sqrt{(0{,}98 - 1{,}05)^2 + \ldots + (2{,}13 - 2{,}13)^2}$$
$$d_{1,2} = 0{,}1228001$$

Perhitungan jarak *Euclidean* antara beras giling dengan beras giling varian pelita sebesar 0,1228001. Perhitungan jarak *Euclidean* terus dilakukan hingga data terakhir. Ringkasan matriks dari hasil perhitungan jarak *Euclidean* sebagai berikut:

Tabel 3. Matriks Jarak *Euclidean*

$\boldsymbol{d_{euc}}$	$\boldsymbol{d_{i,1}}$	$\boldsymbol{d_{i,2}}$	$\boldsymbol{d_{i,3}}$	$\boldsymbol{d_{i,4}}$	$\boldsymbol{d_{i,5}}$	…	$\boldsymbol{d_{i,1146}}$
$d_{1,k}$	0	0,1228001138	0	0,05907809608	0,356900339	…	3,674465409
$d_{2,k}$	0,1228001138	0	0,1228001138	0,1714614122	0,4296785638	…	3,73595525
$d_{3,k}$	0	0,1228001138	0	0,05907809608	0,356900339	…	3,674465409
$d_{4,k}$	0,05907809608	0,1714614122	0,05907809608	0	0,3325100955	…	3,639668409
$d_{5,k}$	0,356900339	0,4296785638	0,356900339	0,3325100955	0	…	3,81861903
⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮
$d_{1146,k}$	3,674465409	3,73595525	3,674465409	3,639668409	3,81861903	…	0

Hasil Pengelompokan Agglomerative Hierarchical Clustering (AHC)

Setelah melakukan perhitungan jarak kedekatan menggunakan *Euclidean*, langkah berikutnya, yaitu menemukan data terdekat dalam matriks jarak *Euclidean* dan menggabungkannya menjadi *cluster* baru. Berdasarkan hasil matriks jarak, maka didapatkan jarak terdekat, yaitu terdapat pada data ke-1 (beras giling) dan data ke-3 (beras giling varian rojolele).

$$\boldsymbol{min(d_{euc}) = min(d_{1,3}) = 0}$$

Didapatkan jarak terdekat pada data ke-1 (beras giling) dengan data ke-3 (beras giling varian rojolele) sebesar 0. Setelah itu, kedua data tersebut digabungkan menjadi satu *cluster*. Hal ini merupakan tahap awal untuk setiap metode yang akan diterapkan pada penelitian ini.

Hasil Cluster Single Linkage

Pada tahap awal sudah didapatkan jarak terdekat, yaitu data ke-1 (beras giling) dengan data

ke-3 (beras giling varian rojolele) sebesar 0 sehingga membentuk *cluster* 1. Kemudian, menghitung jarak antar *cluster* (1 dan 3) dengan *cluster* lain yang tersisa. Didapatkan hasil pengelompokan data makanan berdasarkan komposisi zat gizi sebanyak 2 *cluster*. Ringkasan anggota setiap *cluster* sebagai berikut:

Tabel 4. Hasil Pengelompokan *Single Linkage*

Cluster	Nama Makanan	Banyak Anggota
1	Beras giling (mentah), Beras giling var rojolele (mentah), Beras giling var pelita (mentah), Beras menir (mentah), Beras (tumbuk, mentah), Beras jagung kuning (kering, mentah), Jagung muda kuning (mentah), …, Kelapa muda (air, segar).	1144
2	Dendeng ikan mujair (mentah), dendeng ikan mujair (goreng).	2

Hasil Cluster Complete Linkage

Pada tahap awal sudah didapatkan jarak terdekat, yaitu data ke-1 (beras giling) dengan data ke-3 (beras giling varian rojolele) sebesar 0 sehingga membentuk *cluster* 1. Kemudian, menghitung jarak antar *cluster* (1 dan 3) dengan *cluster* lain yang tersisa. Didapatkan hasil pengelompokan data makanan berdasarkan komposisi zat gizi sebanyak 2 *cluster*. Ringkasan anggota setiap *cluster* sebagai berikut:

Tabel 5. Hasil Pengelompokan *Complete Linkage*

Cluster	Nama Makanan	Banyak Anggota
1	Beras giling (mentah), Beras giling var rojolele (mentah), Beras giling var pelita (mentah), Beras menir (mentah), Beras (tumbuk, mentah), Beras jagung kuning (kering, mentah), Jagung muda kuning (mentah), …, Kelapa muda (air, segar).	1090
2	Kereput, Kue kelapa, Kacang mete/biji jambu monyet (segar), Kacang mete/biji jambu monyet (goreng), Kacang tanah (goreng), Kacang atom, Kacang goyang, Keripik oncom, …, Pala (biji, kering).	56

Hasil Cluster Average Linkage dan Centroid Linkage

Pada tahap awal sudah didapatkan jarak terdekat, yaitu data ke-1 (beras giling) dengan data ke-3 (beras giling varian rojolele) sebesar 0 sehingga membentuk *cluster* 1. Kemudian, menghitung jarak antar *cluster* (1 dan 3) dengan *cluster* lain yang tersisa. Didapatkan hasil pengelompokan data makanan berdasarkan komposisi zat gizi sebanyak 2 *cluster*. Ringkasan anggota setiap *cluster* sebagai berikut:

Tabel 6. Hasil Pengelompokan Average Linkage dan Centroid Linkage

Cluster	Nama Makanan	Banyak Anggota
1	Beras giling (mentah), Beras giling var rojolele (mentah), Beras giling var pelita (mentah), Beras menir (mentah), Beras (tumbuk, mentah),	1134

2	Beras jagung kuning (kering, mentah), Jagung muda kuning (mentah), …, Kelapa muda (air, segar). Lemak kerbau (lemak sapi), Minyak hiu (hati), Minyak ikan, Minyak kedelai, Minyak wijen, Minyak kacang tanah, Minyak kelapa, Minyak zaitun, Minyak kelapa sawit, Margarin, Mentega.	12

Hasil Cluster Ward's Minimum Variance

Pada tahap awal sudah didapatkan jarak terdekat, yaitu data ke-1 (beras giling) dengan data ke-3 (beras giling varian rojolele) sebesar 0 sehingga membentuk *cluster* 1. Kemudian, menghitung jarak antar *cluster* (1 dan 3) dengan *cluster* lain yang tersisa. Didapatkan hasil pengelompokan data makanan berdasarkan komposisi zat gizi sebanyak 2 *cluster*. Ringkasan anggota setiap *cluster* sebagai berikut:

Tabel 7. Hasil Pengelompokan *Ward's Minimum Variance*

Cluster	Nama Makanan	Banyak Anggota
1	Beras giling (mentah), Beras giling var rojolele (mentah), Beras giling var pelita (mentah), Maizena (tepung), Makaroni (mentah), Misoa, Roti putih, Tepung terigu, …, Terasi merah.	425
2	Jagung muda kuning (mentah), Nasi, Ketan (ketupat), Apem (kue), Bubur tinotuan (Manado), Klepon (kue), Ketoprak, Bengkuang (segar), …, Kelapa muda (air, segar).	721

Melakukan Evaluasi *Cluster*

Setelah pengelompokan diperoleh dari hasil analisis cluster menggunakan metode agglomerative hierarchical clustering (AHC) dengan metode complete linkage, single linkage, average linkage, centroid linkage, dan ward's minimum variance pada data komposisi zat gizi pada data Tabel Komposisi Pangan Indonesia, berikutnya perlu melakukan evaluasi cluster dengan menggunakan koefisien cophenetic. Berikut contoh perhitungan koefisien korelasi cophenetic pada 5 data awal matriks Euclidean dengan metode single linkage:

Tabel 8. Perhitungan Koefisien Korelasi *Cophenetic*

(i,k)	$d_{(i,k)}$	$d_{C_{(i,k)}}$	$d_{(i,k)} - \bar{d}$	$d_{C_{(i,k)}} - \bar{d}_C$	$(d_{(i,k)} - \bar{d})(d_{C_{(i,k)}} - \bar{d}_C)$	$(d_{(i,k)} - \bar{d})^2$	$(d_{C_{(i,k)}} - \bar{d}_C)^2$
1,2	0,1228	0,1228	-0,07832	-0,05886	0,0046099	0,006134	0,003464
1,3	0	0	-0,20112	-0,18166	0,0365355	0,04045	0,033
1,4	0,059078	0,059078	-0,14204	-0,12258	0,0174118	0,020176	0,015026
1,5	0,3569	0,33251	0,15578	0,15085	0,0234994	0,024267	0,022756
2,3	0,1228	0,1228	-0,07832	-0,05886	0,0046099	0,006134	0,003464

2,4	0,171461	0,1228	-0,02966	-0,05886	0,0017457	0,00088	0,003464
2,5	0,429679	0,33251	0,228558	0,15085	0,034478	0,052239	0,022756
3,4	0,059078	0,059078	-0,14204	-0,12258	0,0174118	0,020176	0,015026
3,5	0,3569	0,33251	0,15578	0,15085	0,0234994	0,024267	0,022756
4,5	0,33251	0,33251	0,131389	0,15085	0,0198201	0,017263	0,022756
Jumlah	-	-	-	-	0,1836217	0,211986	0,164469

Berdasarkan tabel perhitungan di atas, nilai koefisien korelasi *cophenetic* (r_{coph}) dapat dihitung sebagai berikut:

$$r_{coph} = \frac{\sum_{i<k}(d_{ik} - \bar{d})(d_{C_{ik}} - \bar{d}_C)}{\sqrt{[\sum_{i<k}(d_{ik} - \bar{d})^2][\sum_{i<k}(d_{C_{ik}} - \bar{d}_C)^2]}}$$

$$r_{coph} = \frac{0{,}1836217}{\sqrt{(0{,}211986)(0{,}164469)}} = 0{,}983394$$

Maka, didapatkan koefisien korelasi *cophenetic* untuk metode *single linkage* dari 5 data matriks jarak *Euclidean* sebesar 0,983394. Perhitungan yang dilakukan pada data keseluruhan untuk metode *linkage* yang digunakan dalam penelitian ini sama seperti contoh yang telah diberikan. Hasil dari perhitungan koefisien korelasi *cophenetic* pada data keseluruhan dengan metode *linkage* yang digunakan dalam penelitian ini sebagai berikut:

Tabel 9. Nilai Koefisien Korelasi *Cophenetic*

Metode	Koefisien Korelasi *Cophenetic*
Single linkage	0,788
Complete linkage	0,822
Average linkage	0,907
Centroid linkage	0,805
Ward's minimum variance	0,715

Metode analisis *cluster* yang paling baik dalam merepresentasikan data asli, yaitu ketika nilai koefisien korelasi *cophenetic* semakin mendekati 1. Jika melihat pada tabel di atas, nilai koefisien korelasi *cophenetic* menggunakan metode *average linkage* memiliki nilai koefisien korelasi *cophenetic* terbesar, yaitu sebesar 0,907. Hal tersebut membuktikan bahwa metode *average linkage* adalah metode yang paling baik dalam merepresentasikan data asli.

Melakukan Validitas *Cluster*

Setelah melakukan evaluasi *cluster* didapatkan metode pengelompokan paling baik, yaitu metode *average linkage*. Langkah berikutnya, melakukan validitas *cluster* untuk menentukan jumlah *cluster* optimal. Berikut contoh perhitungan koefisien *silhouette* pada objek ke-1 (beras giling) untuk *cluster* berjumlah 2 (k = 2):

a_1 = avg(jarak objek 1 dengan semua objek pada cluster 1)

$$= \frac{0{,}122800114 + 0 + 0{,}059078096 + \ldots + 3{,}674465409}{1134}$$

$$= 2{,}808414604$$

b_1 = avg(jarak objek 1 dengan semua objek pada cluster 2)

$$= \frac{7{,}711857981 + 8{,}531879498 + 8{,}531879498 + \ldots + 7{,}019688685}{12}$$

$$= 8{,}171658015$$

$$S_1 = \frac{(b_1 - a_1)}{\max(a_1, b_1)} = \frac{8{,}171658015 - 2{,}808414604}{8{,}171658015} = 0{,}656322548$$

⋮

S_i, dengan $i = 1, 2, \ldots, 1146$.

Perhitungan tersebut akan terus dilakukan hingga seluruh nilai S_i dari seluruh data makanan

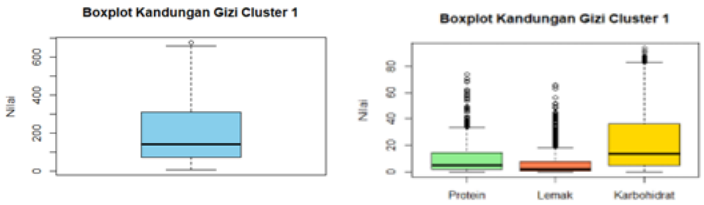
didapatkan. Setelah itu, nilai S_i dari setiap data makanan akan dijumlahkan dan dibagi dengan total data makanan untuk mencari nilai rata-rata yang kemudian nilai tersebut akan menjadi nilai koefisien *silhouette* untuk *cluster* berjumlah 2 (k = 2). Begitupun cara untuk mencari nilai koefisien *silhouette* untuk cluster berjumlah 3 dan seterusnya. Maka, hasil dari perhitungan koefisien *silhouette* sebagai berikut:

Tabel 10. Nilai Koefisien *Silhouette*

Jumlah *Cluster*	Koefisien *Silhouette*
2	0,7098
3	0,5284
4	0,4930
5	0,4818

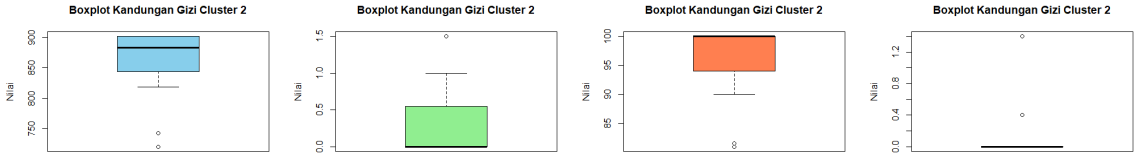
Jika melihat pada tabel di atas, jumlah *cluster* yang paling optimal berdasarkan nilai koefisien *silhouette*, yaitu *cluster* yang berjumlah 2 sebab memiliki nilai koefisien *silhouette* sebesar 0,7098 karena nilai tersebut merupakan yang paling besar atau paling mendekati 1. Hal tersebut membuktikan bahwa metode jumlah *cluster* 2 adalah jumlah yang paling optimal. Berdasarkan pada Tabel 1, nilai koefisien *silhouette* yang diperoleh termasuk ke dalam kategori *cluster* yang kuat.

Profilisasi Hasil *Cluster* Terbaik



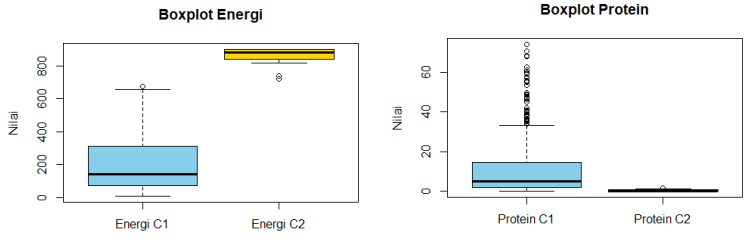
Gambar 1. Boxplot 4 Variabel *Cluster* 1

Berdasarkan Gambar 1, dapat dilihat bahwa pada *cluster* ini, variabel kandungan energi memiliki nilai yang paling besar jika dibandingkan variabel kandungan gizi lainnya. Diikuti dengan variabel kandungan karbohidrat, protein, kemudian lemak yang paling kecil.



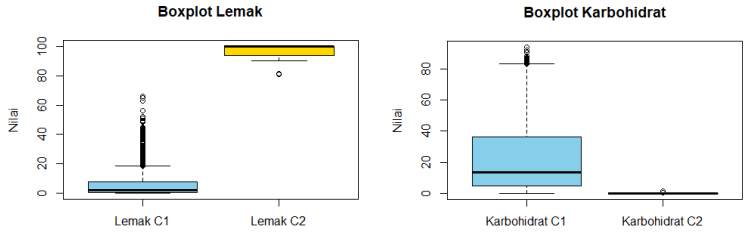
Gambar 2. Boxplot 4 Variabel *Cluster* 2

Sementara itu, berdasarkan Gambar 2, dapat dilihat bahwa pada *cluster* ini, variabel kandungan energi memiliki nilai yang paling besar jika dibandingkan variabel kandungan gizi lainnya. Diikuti dengan variabel kandungan lemak, protein, kemudian karbohidrat yang paling kecil.



Gambar 3. *Boxplot* Variabel Energi dan Protein dari Kedua *Cluster*

Berdasarkan Gambar 3, pada *boxplot* energi diketahui bahwa nilai variabel kandungan energi dari *cluster* 2 lebih besar dibandingkan nilai variabel kandungan energi dari *cluster* 1. Sementara itu, pada *boxplot* protein diketahui bahwa nilai variabel kandungan protein dari *cluster* 1 lebih besar dibandingkan nilai variabel kandungan protein dari *cluster* 2.



Gambar 4. *Boxplot* Variabel Lemak dan Karbohidrat dari Kedua *Cluster*

Berdasarkan Gambar 4, pada *boxplot* lemak diketahui bahwa nilai variabel kandungan lemak dari *cluster* 2 lebih besar dibandingkan nilai variabel kandungan lemak dari *cluster* 1. Sedangkan, pada *boxplot* karbohidrat diketahui bahwa nilai variabel kandungan karbohidrat dari *cluster* 1 lebih besar dibandingkan nilai variabel kandungan karbohidrat dari *cluster* 2.

Berdasarkan gambar-gambar diatas, diketahui bahwa nilai kandungan energi tertinggi terdapat dalam *cluster* 2, nilai kandungan protein tertinggi terdapat dalam *cluster* 1, nilai kandungan lemak tertinggi terdapat dalam *cluster* 2, dan nilai kandungan karbohidrat tertinggi terdapat dalam *cluster* 1. Oleh karena itu, makanan pada *cluster* 1 lebih aman untuk dikonsumsi sehari-hari karena memiliki kandungan protein yang tinggi serta kandungan lemak yang rendah. Makanan pada *cluster* 2 tidak disarankan untuk dikonsumsi sehari-hari atau dikonsumsi dengan jumlah yang terbatas karena kandungan lemak dan energi yang tinggi serta kandungan protein yang rendah sehingga jika dikonsumsi dalam jumlah yang banyak dapat mengakibatkan berbagai masalah kesehatan.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Berdasarkan hasil pengelompokan terhadap data makanan berdasarkan komposisi zat gizi menggunakan metode *Agglomerative Hierarchical Clustering* (AHC) pada metode *single linkage*, *complete linkage*, *average linkage*, *centroid linkage*, dan *ward's minimum variance* didapatkan metode pengelompokan terbaik, yaitu metode *average linkage* dengan jumlah *cluster* optimal sebanyak 2 *cluster*.
2. Berdasarkan metode pengelompokan terbaik, yaitu metode *average linkage*, diketahui bahwa makanan pada *cluster* 1 lebih aman untuk dikonsumsi sehari-hari karena memiliki kandungan protein yang tinggi serta kandungan lemak yang rendah. Sedangkan, makanan pada *cluster* 2 tidak disarankan untuk dikonsumsi sehari-hari karena kandungan lemak dan energi yang tinggi serta kandungan protein yang rendah sehingga jika dikonsumsi dalam jumlah yang banyak dapat mengakibatkan berbagai masalah kesehatan.

Ucapan Terimakasih

Peneliti mengucapkan terima kasih kepada semua pihak yang telah membantu dalam menyelesaikan dan memberikan saran untuk penelitian ini.

Daftar Pustaka

Aisha Kusuma Putri, & Suliadi. (2023). Rekomendasi Destinasi Wisata di Indonesia Menggunakan Metode Item2Vec. *Jurnal Riset Statistika*, 11–18. https://doi.org/10.29313/jrs.v3i1.1770

Herliani, F. D., & Kudus, A. (2023). Penanganan Data Missing dengan Algoritma Multivariate Imputation By Chained Equations (MICE). *DataMath: Journal of Statistics and Mathematics*, *1*(1), 35–42. https://doi.org/10.29313/datamath.v1i1.25

Dani, A. T. R., Wahyuningsih, S., & Rizki, N. A. (2019). Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu. *Jambura Journal of Mathematics*, 1(2), 64-78. https://doi.org/10.34312/jjom.v1i2.2354

Direktorat Jenderal Kesehatan Masyarakat. (2018). *Tabel Komposisi Pangan Indonesia*. Jakarta: Direktorat Jenderal Kesehatan Masyarakat. https://repository.kemkes.go.id/book/777

Fadliana, A., & Rozi, F. (2015). Penerapan Metode Agglomerative Hierarchical Clustering untuk Klasifikasi Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana. *Jurnal Matematika Murni dan Aplikasi*, 4(1), 35-40. https://doi.org/10.18860/ca.v4i1.3172

Fauzi, M., Kastaman, R., & Pujianto, T. (2019). Pemetaan Ketahanan Pangan pada Badan Koordinasi Wilayah I Jawa Barat. *Jurnal Industri Pertanian*, 1(1), 1-10. https://jurnal.unpad.ac.id/justin/article/view/21143

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Group in Data: An Introduction to Cluster Analysis*. New Jersey: John Wiley & Sons Inc Publication. https://www.researchgate.net/profile/Peter-Rousseeuw/publication/220695963_Finding_Groups_in_Data_An_Introduction_To_Cluster_Analysis/links/60fbbe85169a1a0103b20e91/Finding-Groups-in-Data-An-Introduction-To-Cluster-Analysis.pdf

Kementerian Kesehatan Republik Indonesia. (2019). *Leaflet: Informasi Isi Piringku*. https://ayosehat.kemkes.go.id/leaflet-informasi-isi-piringku

Kementerian Kesehatan. (2023). *Survei Kesehatan Indonesia (SKI) 2023 dalam Angka*. Laporan Penelitian. Jakarta: Badan Kebijakan Pembangunan Kesehatan. https://www.badankebijakan.kemkes.go.id/ski-2023-dalam-angka/

Johnson, R. A., & Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis*. *Fifth Edition*. New Jersey: Prentice Hall Inc.

Saracli, S., Dogan, N., & Dogan, I. (2013). Comparison of Hierarchical Cluster Analysis Methods by Cophenetic Correlation. *Journal of Inequalities and Applications*. http://dx.doi.org/10.1186/1029-242X-2013-203

Vredizon, P.A., Firmansyah, H., Salsabila, N. S., & Nugroho, W. E. (2024). Penerapan Algoritma *K-Means* untuk Mengelompokkan Makanan Berdasarkan Nilai Nutrisi. *Journal of*

Technology and Informatics (JoTI), 5(2), 108-115. https://doi.org/10.37802/joti.v5i2.577

Khoeriyah, R. Y., & Hajarisman, N. (2021). Regresi Terboboti Geografis Semiparametrik (RTG-S) untuk Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota di Sumatera Utara. *Jurnal Riset Statistika*, *1*(1), 43–50. https://doi.org/10.29313/jrs.v1i1.145