



Analisis Klaster K-Means pada Data Rata-Rata Konsumsi Kalori dan Protein Menurut Provinsi dengan Metode Davies Bouldin Index

Abyan Rasyid Aryadi¹, Felix Andreas², Betha Nurina Sari ³

^{1,2,3} Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang

Email: abyan.rasyid18081@student.unsika.ac.id¹, felix.andreas18119@student.unsika.ac.id².

³betha.nurina@staff.unsika.ac.id³

Abstrak

Kualitas sumber daya manusia erat kaitannya dengan asupan makanan yang dikonsumsi untuk menentukan status gizi seseorang. Semakin banyaknya masyarakat yang mengkonsumsi makanan dengan kandungan kalori dan protein yang tinggi setiap harinya dapat memicu timbulnya beberapa penyakit seperti diabetes, stroke, penyakit jantung dan obesitas. Dalam hal ini, pemerintah perlu memiliki informasi mengenai data klasifikasi konsumsi kalori dan protein rata-rata harian menurut provinsi. Teknik clustering sangat cocok untuk pengelompokan provinsi berdasarkan tingkat konsumsi kalori dan protein. Algoritma yang digunakan dalam penelitian ini adalah K-Means. Penentuan jumlah cluster menggunakan metode elbow, dimana untuk K-Means sebanyak 2 cluster, sehingga pengelompokan provinsi dibagi menjadi 2 yaitu kategori rendah dan tinggi. Untuk kategori rendah sebanyak 7 provinsi, dan kategori tinggi sebanyak 27 provinsi. Evaluasi hasil cluster K-Means dan K-Medoids menggunakan metode Davies Bouldin Index (DBI) menghasilkan nilai K-Means sebesar 0,688.

Kata kunci: Clustering, K-Means, Elbow Method, Davies Bouldin Index

Abstract

The quality of human resources is closely related to the food intake consumed to determine a person's nutritional status. The increasing number of people who consume foods with high calorie and protein content every day can lead to several diseases such as diabetes, stroke, heart disease and obesity. In this case, the government needs to have information on data on the classification of average daily calorie and protein consumption by province. The clustering technique is very suitable in grouping provinces based on the level of calorie and protein consumption. The algorithm used in this research is K-Means. Determination of the number of clusters using the elbow method, where for K-Means as many as 2 clusters, so that the grouping of provinces is divided into 2, namely low and high categories. For the low category as many as 7 provinces, and the high category as many as 27 provinces. Evaluation of the results of the K-Means and K-Medoids clusters using the Davies Bouldin Index (DBI) method resulted in a K-Means value of 0.688.

Keywords : Clustering, K-Means, Elbow Method, Davies Bouldin Index

PENDAHULUAN

Kalori merupakan kelompok zat gizi makronutrien, yang berasal dari karbohidrat, protein dan lemak. Kalori merupakan satuan ukur untuk menyatakan nilai energi (Sri & Lilik, 2018). Sedangkan protein yaitu zat gizi penghasil energi yang terdiri dari banyak asam amino. Meningkatnya jumlah masyarakat yang mengkonsumsi makanan yang mengandung kalori dan protein tinggi setiap harinya dapat mengakibatkan timbulnya beberapa penyakit seperti penyakit diabetes, jantung serta obesitas. Mengenai hal ini pemerintah perlu memiliki informasi mengenai data pengelompokan rata-rata konsumsi kalori dan protein per hari menurut provinsi. Data tersebut sangat dibutuhkan pemerintah

sebagai informasi supaya dapat memberikan arahan kepada masyarakat untuk menjaga kesehatan serta melakukan upaya menjaga keseimbangan dalam mengkonsumsi kalori dan protein per harinya. Untuk masyarakat yang kekurangan kalori dan protein, pemerintah dapat memberikan bantuan berupa asupan makanan bergizi yang seimbang. Solusi yang tepat dalam melakukan penelitian ini menggunakan Data Mining dengan menggunakan algoritma clustering yaitu K-Means untuk mengelompokkan provinsi di Indonesia berdasarkan rata-rata konsumsi kalori dan protein (Juwita et. al, 2019).

Konsep dasar clustering yaitu mengelompokkan beberapa objek ke dalam cluster, berdasarkan tingkat kemiripan tinggi antar objek dalam cluster dan tingkat ketidaksamaan yang tinggi dengan objek cluster lainnya memperoleh kelompok cluster yang baik. Alasan penggunaan algoritma K-Means karena algoritma ini merupakan salah satu klaster paling baik. Penelitian sebelumnya dilakukan oleh (Juwitha Lovely Sweets Sinaga, Solikhun, dan Dedi Suhendro, 2019) yang berjudul "Penerapan Algoritma K-Means Dalam Mengelompokkan Rata-Rata Konsumsi Kalori Menurut Provinsi", bahwa hasil penelitian diperoleh dari pengelompokan rata-rata konsumsi kalori yang dikelompokkan dengan dua cluster yaitu cluster rendah dan tinggi, cluster rendah 13 provinsi dan cluster tinggi 21 provinsi. Penelitian yang telah dilakukan oleh (Wahyu Saputro, Moh Reza Pahlevi, Arief Wibowo, 2020) berjudul "Analisis Algoritma K-Means Untuk Klasterisasi Tindak Pidana Korupsi Di Wilayah Hukum Indonesia" bahwa dalam penelitiannya, di peroleh 3 cluster/kelompok berdasarkan tingkat kerawanan tindak pidana korupsi. Dengan demikian model terbaik adalah pada terbentuknya 3 klaster dengan nilai DBI sebesar 0,113. DBI terkecil menyatakan bahwa model memiliki bentuk klaster terbaik sebanyak 3. Hasil penelitian menunjukkan bahwa wilayah dengan kategori rawan tindak pidana korupsi berada pada pulau sumatra ada 5 wilayah, pulau jawa ada 3 wilayah dan pulau sulawesi dan kalimantan ada 1 wilayah. Dalam penelitian yang telah dilakukan oleh (Ninik Tri Hartanti, 2020) yang berjudul "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional", bahwa dalam penelitiannya dilakukan penentuan jumlah optimal cluster menggunakan metode Elbow, yang menunjukkan banyaknya cluster optimal adalah 3 cluster, kemudian dihasilkan 3 cluster yang terdiri dari cluster dengan kategori "Siap", "Cukup Siap" dan Tidak Siap". Dengan masing-masing kategori terdiri dari 7 siswa, 30 siswa, dan 29 siswa. Berdasarkan penelitian sebelumnya, penelitian yang akan dilakukan yaitu melakukan klasterisasi dengan algoritma clustering K-Medoids. Penggunaan algoritma clustering tersebut diperlukan untuk menganalisis pola pengelompokan dalam pengolahan data rata rata konsumsi kalori dan protein menurut provinsi. Penelitian ini juga menerapkan metode DBI (Davies Bouldin Index) sebagai metode analisis hasil cluster dari perhitungan yang telah diperoleh.

METODE

Metodologi penelitian yang digunakan dalam penelitian ini yaitu Knowledge Discovery in Database (KDD). Urutan metode Knowledge Discovery in Database (KDD) yaitu Data Selection, Data Preprocessing, Data Transformation, Data Mining, Evaluation, dan tahap Knowledge. Data yang digunakan Pada penelitian ini adalah data sekunder. Data tersebut diperoleh dari data Susenas Badan Pusat Statistik. Berikut merupakan gambar tahapan penelitian menggunakan metodologi KDD.



Gambar 1. Metodologi *Knowledge Discovery in Database* (KDD)

Berikut merupakan rancangan penelitian pada proses *data mining*:

1. *Data Selection*

Data yang digunakan merupakan data rata-rata konsumsi kalori dan protein menurut provinsi dari tahun 2007-2021 bersumber situs Badan Pusat Statistik. Data berjumlah 34 buah data berdasarkan provinsi di Indonesia, nilai dalam data didapat dari penjumlahan rata-rata kalori dan protein dari tahun 2007-2021, serta data tersebut memiliki 2 buah variabel yaitu kalori dan protein.

2. *Data Preprocessing*

Pada tahap ini, dilakukan proses menghitung rata-rata dari dataset mentah. Perhitungan ini dilakukan untuk mendapatkan nilai rata-rata dari masing-masing data kalori dan protein dari tahun 2007-2021, sehingga mendapatkan 2 variabel utama yaitu kalori dan protein.

3. *Data Transformation*

Data transformation dilakukan untuk mengatasi kondisi yang tidak ideal pada data rata-rata konsumsi kalori dan protein seperti adanya data yang hilang atau skala pada data yang tidak sama. Hal tersebut dilakukan dengan cara normalisasi data. Normalisasi data dilakukan untuk menstandarkan data yang ada pada data, sehingga setiap nilai yang ada pada variabel data memiliki skala yang sama. Dalam penelitian ini, proses normalisasi data menggunakan metode *min-max normalization* (Hablum et. al, 2019).

4. *Data Mining*

Pada penelitian ini pengelompokan provinsi menggunakan algoritma *K-Means* yang merupakan salah satu algoritma *clustering* dalam *data mining* bersifat *unsupervised* dengan titik pusat dalam setiap *cluster* menggunakan *centroid*. Karena algoritmanya yang tergolong sederhana dan konvergensi yang cepat, algoritma *K-Means* banyak digunakan dalam setiap penelitian untuk mengelompokkan data. Dalam *K-Means*, nilai *K* perlu diberikan terlebih dahulu dan pilihan nilai *K* secara langsung mempengaruhi hasil konvergensi. Algoritma *K-Means* merupakan algoritma yang diwakili oleh *cluster* yaitu *centroid* atau nilai rata-rata sebagai titik pusat *cluster* (Riva Arsyad Farissa et. al, 2021).

Langkah-langkah pengerjaan algoritma *K-Means* diantaranya sebagai berikut (Wahyu, Moh. Reza & Arif, 2020) :

- Tentukan *K* sebagai jumlah *cluster* yang akan dibentuk.
- Buat *K* *centroid* secara random.
- Hitung jarak setiap data ke masing-masing *centroid*.
- Setiap data memilih *centroid* yang terdekat.

- e. Tentukan posisi *centroid* yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada *centroid* yang sama.
- f. Jika posisi *centroid* baru dengan *centroid* yang lama masih belum memiliki kesamaan maka dilangkahi lagi pada langkah ke 3 sampai menemukan kesamaan.
- g. Tentukan jumlah klaster yang akan dibentuk dan k sebagai nilainya.
- h. Hitung jarak setiap inputan terhadap *centroid-centroid* menggunakan persamaan *Euclidean*, dengan rumus i

$$D(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (xi - yi)^2}$$

5. Evaluation

Dalam menentukan kualitas *cluster*, dibutuhkan evaluasi dari hasil *cluster* yang sudah dibentuk. Metode evaluasi *cluster* menggunakan metode *Davies Bouldin Index* (DBI). Dari proses evaluasi hasil *cluster* ini, nilai DBI yang didapat digunakan sebagai acuan dalam menentukan algoritma *cluster* mana yang paling baik dan akurat. Dalam evaluasi *cluster* dengan metode DBI, jika nilai DBI mendekati 0 maka semakin baik hasil *cluster* yang diperoleh (Fitriana, 2021).

6. Knowledge

Pada tahapan terakhir yaitu dimana data yang sudah diolah bisa dievaluasi kembali. Hasil yang telah didapat disimpulkan dan diubah ke dalam bentuk yang mudah dipahami dan dimengerti.

HASIL

Dataset yang digunakan dalam melakukan penelitian ini yaitu menggunakan data yang diperoleh dari dokumen BPS yang berupa data rata-rata konsumsi kalori dan protein menurut provinsi di Indonesia dari tahun 2007-2021 dengan total data sebanyak 34 data dan total 2 variabel, yaitu variabel rata-rata Kalori dan Protein. Untuk dataset yang akan diproses yaitu sebagai berikut :

Tabel 1. Data Rata-Rata Kalori dan Protein

Provinsi	Kalori	Protein
Aceh	2025,05	56,90
Sumatera Utara	2028,89	58,28
Sumatera Barat	2085,47	56,10
Riau	2016,22	56,58
Jambi	1988,54	54,36
Sumatera Selatan	2051,03	56,88
Bengkulu	2038,23	55,63
Lampung	1993,66	53,76
Kepulauan Bangka Belitung	1999,07	59,82
Kepulauan Riau	2044,54	62,16
DKI Jakarta	2001,10	63,36
Jawa Barat	2048,16	59,49
Jawa Tengah	1948,74	55,44
DI Yogyakarta	1992,92	60,03
Jawa Timur	1964,35	56,40
Banten	2085,84	61,49
Bali	2195,93	63,56
Nusa Tenggara Barat	2134,72	61,56
Nusa Tenggara Timur	1908,22	52,18
Kalimantan Barat	1928,10	54,48
Kalimantan Tengah	2080,54	60,14

Kalimantan Selatan	2134,60	62,36
Kalimantan Timur	1869,18	57,38
Kalimantan Utara	1925,36	60,41
Sulawesi Utara	2053,41	59,02
Sulawesi Tengah	1994,56	54,26
Sulawesi Selatan	2063,09	59,65
Sulawesi Tenggara	2002,84	57,60
Gorontalo	1958,95	54,80
Sulawesi Barat	2010,94	55,41
Maluku	1831,30	50,22
Maluku Utara	1778,86	46,99
Papua Barat	1835,37	51,11
Papua	1842,69	43,98

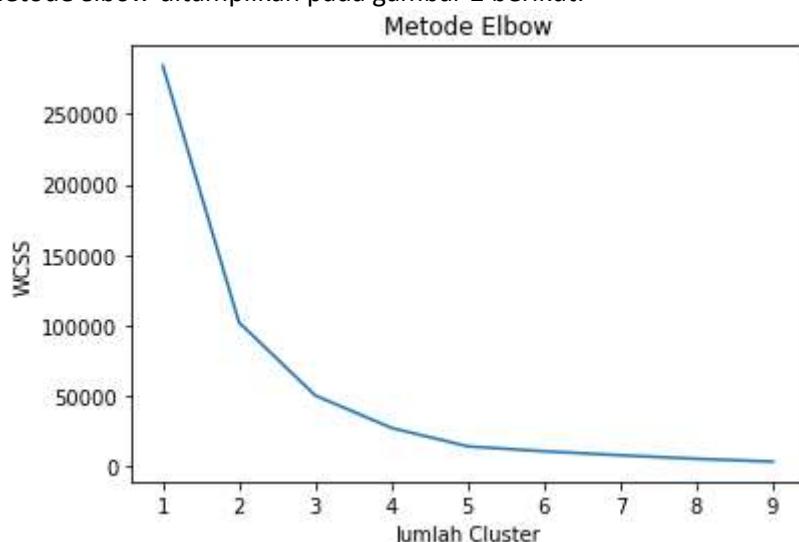
Data rata-rata konsumsi kalori dan protein yang sudah didapat melalui perhitungan dimasukkan kedalam *software Microsoft Excel*, sehingga kemudian bisa dilakukan proses *data mining* menggunakan *software Rapidminer Studio*.

Proses Data Mining

Dalam proses *data mining* terdapat tahapan yang dilalui sebelum dilakukan proses perhitungan *clustering* dengan algoritma *K-Means*. Yaitu proses penentuan jumlah *cluster* K yang paling tepat sebelum dilakukan proses *clustering*. Dalam menentukan jumlah *cluster* ini kami menggunakan metode *Elbow*, dimana untuk mendapatkan nilai *Elbow* tersebut menggunakan *Python* dengan *tools Jupyter Notebook*. Hasilnya sebagai berikut.

```
wcss = []
for i in range(1, 10):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 10), wcss)
```

Selanjutnya *output* dari program tersebut menghasilkan grafik yang menunjukkan jumlah *cluster* K yang paling ideal, sehingga dapat dilakukan proses selanjutnya yaitu *clustering* dengan algoritma *K-Means*. Grafik metode *elbow* ditampilkan pada gambar 2 berikut.



Gambar 2. Grafik metode elbow

Berdasarkan gambar 3 diatas, diketahui bahwa jumlah *cluster* K yang didapat berdasarkan perhitungan metode *elbow* dengan *Python* yaitu berjumlah 2 *cluster*. Dari hasil yang didapat, kami menyimpulkan bahwa data rata-rata konsumsi kalori dan protein tersebut akan dikelompokkan menjadi 2 *cluster* yaitu *cluster* rendah dan tinggi atau K=2. Pengelompokan dua *cluster* dilakukan agar diketahui Provinsi mana saja yang masuk *cluster* rendah dan tinggi. Jarak yang digunakan untuk menghitung jarak terpendek yaitu menggunakan *Euclidean Distance*. Selanjutnya yaitu melakukan proses *clustering* dengan algoritma *K-Means*, dimana untuk proses ini kami menggunakan *tools RapidMiner Studio*. Berikut merupakan model hasil *clustering K-Means* yang terbentuk berdasarkan perhitungan pada *tools RapidMiner Studio*.

Cluster Model

```
Cluster 0: 11 items
Cluster 1: 23 items
Total number of items: 34
```

Gambar 3. Tampilan Hasil Model *Cluster*

Berdasarkan gambar diatas, bisa dilihat bahwa dari proses *clustering* yang telah dilakukan menghasilkan 2 kelompok *cluster* dimana pada *cluster* 0 terdapat 11 provinsi, dan *cluster* 1 terdapat 23 provinsi. Dari perhitungan *clustering* yang telah dilakukan, didapatkan nilai *centroid* akhir dari 2 kelompok *cluster*. Berikut disajikan tabel nilai *centroid* akhir yang telah dilakukan menggunakan *RapidMiner Studio*.

Tabel 2. Nilai *Centroid* Akhir

Variabel	Cluster 0	Cluster 1
Kalori	1890.101	2046.493
Protein	53.036	58.626

Dari nilai akhir *centroid* yang terbentuk diatas, didapatkan 2 *cluster* pengelompokan berdasarkan provinsi dari data rata-rata kalori dan protein. Berikut tabel hasil *cluster* yang terbentuk menggunakan *RapidMiner Studio*.

Tabel 3. Hasil *Cluster* yang Terbentuk

No	Provinsi	Cluster
1	Aceh	cluster_1
2	Sumatera Utara	cluster_1
3	Sumatera Barat	cluster_1
4	Riau	cluster_1
5	Jambi	cluster_1
6	Sumatera Selatan	cluster_1
7	Bengkulu	cluster_1
8	Lampung	cluster_1
9	Kepulauan Bangka Belitung	cluster_1
10	Kepulauan Riau	cluster_1
11	DKI Jakarta	cluster_1
12	Jawa Barat	cluster_1

13	Jawa Tengah	cluster_0
14	DI Yogyakarta	cluster_1
15	Jawa Timur	cluster_0
16	Banten	cluster_1
17	Bali	cluster_1
18	Nusa Tenggara Barat	cluster_1
19	Nusa Tenggara Timur	cluster_0
20	Kalimantan Barat	cluster_0
21	Kalimantan Tengah	cluster_1
22	Kalimantan Selatan	cluster_1
23	Kalimantan Timur	cluster_0
24	Kalimantan Utara 1)	cluster_0
25	Sulawesi Utara	cluster_1
26	Sulawesi Tengah	cluster_1
27	Sulawesi Selatan	cluster_1
28	Sulawesi Tenggara	cluster_1
29	Gorontalo	cluster_0
30	Sulawesi Barat	cluster_1
31	Maluku	cluster_0
32	Maluku Utara	cluster_0
33	Papua Barat	cluster_0
34	Papua	cluster_0

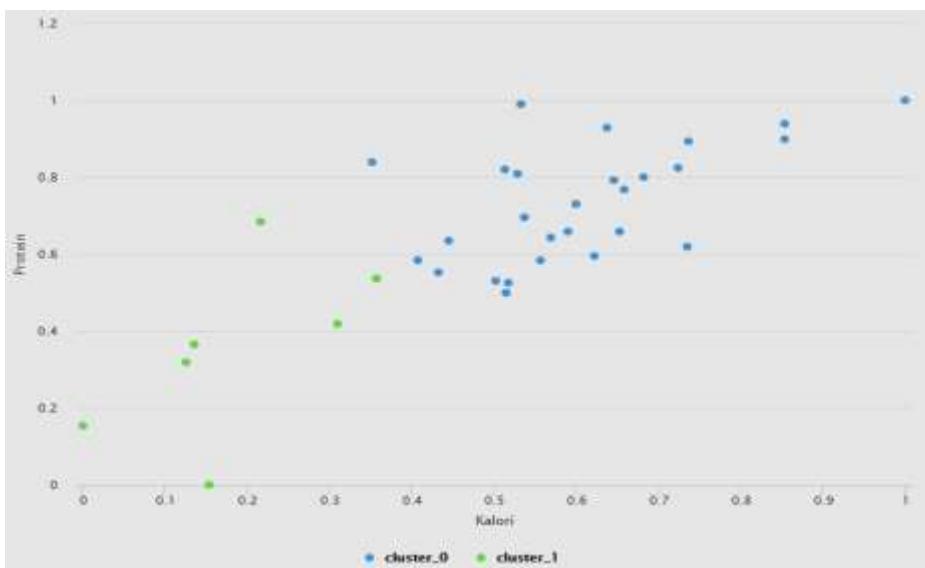
Proses evaluasi *cluster K-Means* dilakukan setelah melalui perhitungan dan didapatkan hasil dari *clustering K-Means*. Berdasarkan perhitungan sebelumnya, dihasilkan nilai K optimal yaitu 2. Dari hasil *cluster* tersebut, didapatkan nilai *Davies Bouldin Index* dengan menggunakan nilai k = 2. Berdasarkan gambar 4 dapat dilihat bahwa hasil akurasi menggunakan metode *K-Means* dengan nilai k = 2 didapatkan nilai DBI sebesar 0,688.

Davies Bouldin

Davies Bouldin: -0.688

Gambar 4. Nilai *Davies Bouldin Index K-Means*

Tahap akhir yaitu *knowledge*. Dimana pada tahap ini disusun hasil penelitian berupa laporan akhir sehingga didapatkan pengetahuan mengenai proses *data mining* supaya mudah dipahami. Pada tahap ini juga dilakukan analisis untuk menentukan kelompok provinsi berdasarkan hasil *cluster*, yang didapat dari hasil perhitungan *clustering* dengan algoritma *K-Means*. Pada tahap ini dilakukan proses analisis untuk menentukan kelompok *cluster* manakah setiap provinsi berada. Setelah mendapatkan hasil *cluster* dengan jumlah K = 2, diperoleh hasil visualisasi *cluster* dengan algoritma *K-Means*.



Gambar 5. Visualisasi Hasil Cluster K-Means

Berdasarkan hasil visualisasi *cluster*, diketahui bahwa terdapat 2 *cluster* yang terbentuk. Titik persebaran *cluster* yang berwarna biru merupakan *cluster 0* dan titik berwarna hijau merupakan *cluster 1*. Selanjutnya melakukan analisis terhadap 2 *cluster* tersebut untuk menentukan kelompok provinsi dengan tingkat konsumsi kalori dan protein rendah, dan tinggi.

Proses analisis dilakukan berdasarkan letak titik persebaran *cluster* pada hasil visualisasi *cluster K-Means*. Jika titik persebaran semakin mendekati 0 maka bisa disimpulkan bahwa *cluster* tersebut merupakan kelompok provinsi dengan tingkat konsumsi rendah. Sebaliknya jika titik persebaran semakin mendekati 1 maka bisa disimpulkan bahwa *cluster* tersebut merupakan kelompok provinsi dengan tingkat konsumsi tinggi. Berikut hasil analisis kelompok *cluster*.

- Data *cluster 0* merupakan kelompok provinsi dengan tingkat konsumsi kalori dan protein tinggi dengan jumlah 27 provinsi, yang bisa dilihat pada gambar 5 pada titik warna biru semakin mendekati nilai 1.
- Data *cluster 1* merupakan kelompok provinsi dengan tingkat konsumsi kalori dan protein rendah dengan jumlah 7 provinsi, yang bisa dilihat pada gambar 5 pada titik warna hijau semakin mendekati nilai 0.

Berdasarkan hasil analisis kelompok *cluster* tersebut, pengelompokan provinsi berdasarkan tingkat konsumsi kalori dan protein disajikan tabel 4.

Tabel 4. Kelompok Provinsi Pada Cluster K-Means

Cluster	
C0 (Tinggi)	C1 (Rendah)
Aceh	Nusa Tenggara Timur
Sumatera Utara	Kalimantan Barat
Sumatera Barat	Kalimantan Timur
Riau	Maluku
Jambi	Maluku Utara
Sumatera Selatan	Papua Barat
Bengkulu	Papua
Lampung	
Kepulauan Bangka Belitung	

Kepulauan Riau	
DKI Jakarta	
Jawa Barat	
Jawa Tengah	
DI Yogyakarta	
Jawa Timur	
Banten	
Bali	
Nusa Tenggara Barat	
Kalimantan Tengah	
Kalimantan Selatan	
Kalimantan Utara	
Sulawesi Utara	
Sulawesi Tengah	
Sulawesi Selatan	
Sulawesi Tenggara	
Gorontalo	
Sulawesi Barat	

Berdasarkan hasil kelompok *cluster* yang telah didapat, karakteristik masing-masing cluster dijabarkan dalam tabel 5, dan tabel 6.

Tabel 5. Rincian Karakteristik Cluster 0 K-Means

C0	Kalori (kkal)	Protein (gram)
Min	1778,86	43,98
Max	1928,10	57,38
Rata-Rata	1856,24	50,90

Cluster 0 merupakan kategori dengan tingkat konsumsi kalori dan protein tertinggi. *Cluster 0* terdiri dari 27 provinsi. Pada variabel kalori, angka konsumsi kalori terkecil = 1778,86 kkal, angka konsumsi terbesar = 1928,10 kkal, dan rata-rata konsumsi kalori = 1856,24 kkal. Pada variabel protein, angka konsumsi protein terkecil = 43,98 gram, angka konsumsi terbesar = 57,38 gram, dan rata-rata konsumsi protein = 50,90 gram.

Tabel 6. Rincian Karakteristik Cluster 1 K-Means

C1	Kalori (kkal)	Protein (gram)
Min	1925,36	53,76
Max	2195,93	63,56
Rata-Rata	2032,10	58,34

Cluster 1 merupakan kategori dengan tingkat konsumsi kalori dan protein terendah. *Cluster 0* terdiri dari 7 provinsi. Pada variabel kalori, angka konsumsi kalori terkecil = 1925,36 kkal, angka konsumsi terbesar = 2195,93 kkal, dan rata-rata konsumsi kalori = 2032,10 kkal. Pada variabel protein, angka konsumsi protein terkecil = 53,76 gram, angka konsumsi terbesar = 63,56 gram, dan rata-rata konsumsi protein = 58,34 gram.

Hasil merupakan bagian utama artikel ilmiah, berisi : hasil bersih tanpa proses analisis data, hasil pengujian hipotesis. Hasil dapat disajikan dengan table atau grafik, untuk memperjelas hasil secara verbal

SIMPULAN

Berdasarkan penelitian yang telah dilakukan untuk mengelompokkan provinsi berdasarkan konsumsi kalori dan protein, dapat diambil kesimpulan bahwa pada algoritma K-Means menghasilkan nilai $K = 2$ cluster. sehingga pengelompokan provinsi berdasarkan tingkat konsumsi kalori dan protein terbagi dua kategori yaitu rendah, dan tinggi. Pada algoritma K-Means dihasilkan 2 kelompok cluster berdasarkan tingkat konsumsi kalori dan protein, yaitu cluster 0 merupakan kelompok provinsi dengan tingkat konsumsi tinggi sebanyak 27 provinsi, dan cluster 1 merupakan kelompok provinsi dengan tingkat konsumsi rendah sebanyak 7 provinsi. Evaluasi algoritma K-Means menggunakan metode Davies Bouldin Index (DBI). Nilai DBI yang mendekati angka 0, maka semakin baik hasil cluster tersebut. Dari perhitungan DBI diperoleh hasil untuk K-Means sebesar 0.688.

DAFTAR PUSTAKA

- A. S. Wicaksono and A. M. Yolanda, "Pengelompokan Kabupaten / Kota di Provinsi Nusa Tenggara Timur Berdasarkan Indikator Indeks Pembangunan Manusia Menggunakan K-Medoids Clustering Penyedia Data Statistik Berkualitas untuk Indonesia Maju Pengelompokan Kabupaten / Kota di Provinsi Nusa Ten," 2021.
- E. Elmayati, "Data Mining Dengan Metode Clustering Untuk Pengolahan Informasi Persediaan Obat Pada Klinik Srikandi Medika Berbasis Web," *Pelita Inform. Inf. dan Inform.*, vol. 16, no. 4, pp. 357–362, 2017.
- N. P. Siwi and I. Paskarini, "Hubungan Asupan Karbohidrat, Lemak, dan Protein dengan Status Gizi," *Indones. J. Public Heal.*, vol. 13, no. 1, pp. 1–12, 2018, doi: 10.20473/ijph.v13i1.2018.1-12.
- N. T. Hartanti, "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
- Q. Wang, J. Liu, B. Wei, W. Chen, and S. Xu, "Investigating the construction, training, and verification methods of k-means clustering fault recognition model for rotating machinery," *IEEE Access*, vol. 8, pp. 196515–196528, 2020, doi: 10.1109/ACCESS.2020.3028146.
- R. Hablum, A. Khairan, and R. Rosihan, "Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 1, pp. 26–33, 2019, doi: 10.33387/jiko.v2i1.1053.
- S. I. Murpratiwi, I. G. Agung Indrawan, and A. Aranta, "Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail," *J. Pendidik. Teknol. dan Kejurut.*, vol. 18, no. 2, p. 152, 2021, doi: 10.23887/jptk-undiksha.v18i2.37426.
- S. Maywati and L. Hidayanti, "Peran asupan makronutrien (kalori) siang hari terhadap kelelahan pekerja wanita," *J. Siliwangi Seri Sains dan Teknol.*, vol. 4, no. 1, pp. 9–13, 2018.
- V. V Arganata, "Algoritma Partitioning Around Medoids (Pam) Dengan Metode Davies Bouldin Index Untuk Mengelompokkan Provinsi Di Indonesia ...," 2021.
- W. Saputro, M. Reza Pahlevi, and A. Wibowo, "Analisis Algoritma K-Means Untuk Klasterisasi Tindak Pidana Korupsi Di Wilayah Hukum Indonesia," *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 3, pp. 137–142, 2020, doi: 10.33387/jiko.v3i3.1960.
- Y. Septiani, E. Aribbe, and R. Diansyah, "Analisis Kualitas Layanan Sistem Informasi Akademik Universitas Abdurrahman Terhadap Kepuasan Pengguna Menggunakan Metode Sevqual (Studi Kasus : Mahasiswa Universitas Abdurrahman Pekanbaru)," *J. Teknol. Dan Open Source*, vol. 3, no. 1, pp. 131–143, 2020, doi: 10.36378/jtos.v3i1.560.