

Segmentasi Produk Minuman Tidak Termasuk Produk Susu Berdasarkan Informasi Nilai Gizi Menggunakan Metode DBSCAN

Baiq Wita Rachmatia, Arum Handini Primandari*

Fakultas Matematika dan Ilmu Pengetahuan Alam, Statistika, Universitas Islam Indonesia, Sleman, Indonesia

Email: ¹baiq.rachmatia@students.uii.ac.id, ^{2,*}primandari.arum@uui.ac.id

Email Penulis Korespondensi: primandari.arum@uui.ac.id

Submitted: 02/05/2025; Accepted: 31/05/2025; Published: 01/06/2025

Abstrak—Sebanyak 28.7% masyarakat Indonesia mengonsumsi gula, garam, dan lemak (GGL) melebihi batas yang direkomendasikan oleh Kementerian Kesehatan. Konsumsi MBDK (minuman berpemanis dalam kemasan) meningkat pesat dalam dua dekade terakhir, menjadikan Indonesia sebagai negara tertinggi ketiga dengan konsumsi MBDK di Asia Tenggara. Untuk mengurangi konsumsi MBDK, konsumen memerlukan informasi yang efektif terkait kandungan GGL dalam produk yang akan dibeli. Sementara informasi nilai gizi yang tercantum dalam kemasan belum dimanfaatkan secara optimal karena kompleksitas penyajiannya. Oleh karena itu, diperlukan segmentasi produk sebagai acuan dalam memilih minuman yang dikonsumsi, serta mendukung intervensi kesehatan, dan mengendalikan konsumsi produk berisiko. Data nilai GGL diambil dari MBDK yang dijual pada tiga jenis toko dengan memfoto tabel informasi nilai gizi. Data dianalisis menggunakan metode DBSCAN, yang mampu menangani keragaman data dan mengatasi outlier tanpa menentukan jumlah *cluster* sebelumnya. Evaluasi *clustering* menggunakan *silhouette coefficient* dan *dunn index* memberikan gambaran kualitas *cluster*. Hasil statistika deskriptif menunjukkan bahwa sebagian besar produk memiliki kandungan lemak rendah, namun kadar gula lebih tinggi, mendekati 15 gram. Standarisasi data dengan *z-score* dilakukan untuk mengatasi perbedaan skala variabel. *Clustering* DBSCAN membagi data menjadi dua *cluster* dan noise. Evaluasi *clustering* menunjukkan performa yang cukup baik dengan *silhouette coefficient* 0.396 dan *dunn index* 0.137. Hasil uji-t menunjukkan terdapat *perbedaan* yang signifikan antara *cluster* 1 dan 2.

Kata Kunci: DBSCAN; Informasi Nilai Gizi GGL; Klasterisasi Produk Minuman Tidak Termasuk Susu; Klasterisasi *Outlier*

Abstract—Approximately 28.7% of Indonesians consume sugar, salt, and fat (SSF) in amounts that exceed the Ministry of Health's recommended limits. Over the past two decades, sweetened drink (MBDK: *minuman berpemanis dalam kemasan*) consumption has surged, making Indonesia the third highest in Southeast Asia for MBDK consumption. To mitigate this, consumers need clear information about GGL content, but nutritional labels are often complex and underutilized. Product segmentation can help consumers make healthier drink choices and support health interventions aimed at reducing risky consumption. Data on GGL values were collected from MBDK sold in three store types and analyzed using the DBSCAN method, which handles diversity and outliers without predefining cluster numbers. Descriptive statistics showed most products had low fat but higher sugar content, nearing 15 grams. After standardizing the data using *z-scores*, the DBSCAN clustering revealed two clusters and some noise. The evaluation indicated a silhouette coefficient of 0.396 and a Dunn index of 0.137, with t-tests showing significant differences between the clusters.

Keywords: DBSCAN; Nutrition Information on SSF; Clustering Beverage Products Excluding Milk; Outliers Clustering

1. PENDAHULUAN

Salah satu indikator utama dari penyakit metabolik kronis yang dikenal sebagai diabetes adalah meningkatnya kadar glukosa dalam darah. Kondisi ini dapat menyebabkan kerusakan serius dan progresif pada berbagai organ tubuh, termasuk pembuluh darah, mata, ginjal, jantung, serta sistem saraf. Tipe diabetes yang paling umum adalah diabetes tipe 2, yang umumnya terjadi pada orang dewasa dan ditandai oleh resistensi tubuh terhadap insulin atau ketidakmampuan pankreas untuk memproduksi insulin dalam jumlah yang memadai [1]. Menurut IDF Diabetes Atlas edisi ke-10, diabetes telah menjadi masalah besar bagi kesehatan dan kesejahteraan individu, keluarga, dan masyarakat di seluruh dunia. Laporan ini menunjukkan peningkatan yang berkelanjutan dalam prevalensi diabetes di seluruh dunia. Saat ini, 537 juta orang di dunia menderita diabetes. Pada tahun 2045, diperkirakan 700 juta orang di seluruh dunia akan mengidap penyakit ini. Lebih dari 4 juta orang meninggal setiap tahun karena diabetes, yang berdampak buruk pada individu, masyarakat, negara, dan wilayah [2].

Menurut data Kementerian Kesehatan, saat ini ada 19.5 juta orang di Indonesia yang menderita diabetes, dan angka ini diperkirakan akan meningkat menjadi 28.5 juta orang pada tahun 2045. Menurut Wakil Menteri Kesehatan Dante Saksono Harbuwono, penyebabnya adalah kegemaran masyarakat terhadap makanan manis, termasuk minuman manis yang dikemas, yang telah terbukti meningkatkan risiko diabetes [3]. Berdasarkan data Badan Pusat Statistik (BPS) 2021, konsumsi gula putih per kapita di Indonesia mencapai 1123 gram mingguan, melebihi tiga kali rekomendasi Kementerian Kesehatan (Kemenkes) atau enam kali lipat anjuran Organisasi Kesehatan Dunia (WHO). Preferensi masyarakat terhadap minuman manis tercermin dalam tingginya konsumsi minuman berpemanis dalam kemasan (MBDK). Laporan dari *Center for Indonesia Strategic Development Initiatives* (CISDI) mencatatkan bahwa konsumsi MBDK mengalami kenaikan dramatis, meningkat hingga 15 kali lipat dalam dua dekade terakhir. Konsumsi yang awalnya tercatat 51 juta liter pada 1996, melonjak menjadi 780 juta liter pada 2014. Pada 2020, Indonesia tercatat sebagai negara dengan konsumsi MBDK tertinggi ketiga di Asia Tenggara [4].

Data yang dihimpun oleh Kementerian Kesehatan menunjukkan bahwa sebanyak 28.7% penduduk Indonesia mengonsumsi gula, garam, dan lemak (GGL) melebihi batas yang direkomendasikan. Sebagai respons terhadap kondisi tersebut, pemerintah telah mengembangkan berbagai strategi intervensi guna mengendalikan asupan GGL,

yang mencakup penguatan regulasi, reformulasi produk pangan, penerapan kebijakan pajak atau cukai, pelaksanaan kegiatan penelitian, serta penyelenggaraan kampanye edukatif. Salah satu kebijakan penting dalam kerangka ini adalah Peraturan Menteri Kesehatan No. 30 Tahun 2013 yang kemudian diperbarui melalui Permenkes No. 63 Tahun 2015. Regulasi tersebut mewajibkan produsen mencantumkan informasi kandungan GGL dalam label nilai gizi pada kemasan, iklan, dan seluruh media promosi produk pangan olahan dan siap saji. Selain itu, pemerintah juga menganjurkan masyarakat untuk mengikuti pedoman gizi seimbang dan membatasi konsumsi GGL masing-masing tidak melebihi 50 gram gula, 2 gram garam, dan 67 gram lemak per hari sesuai dengan rekomendasi yang berlaku [5]. Adapun produk minuman kemasan biasanya disajikan dalam bentuk air minum dan serbuk. Penyajian yang praktis tersebut menyebabkan konsumen sangat mudah dalam mengonsumsinya, tanpa melihat informasi nilai gizi dalam produk.

Segmentasi produk merupakan strategi penting yang membagi produk ke dalam beberapa kelompok berdasarkan karakteristik dan kebutuhan yang sama. Oleh karena itu, segmentasi produk sangat dibutuhkan saat ini sebagai acuan pengambilan keputusan dalam memilih dan mengontrol produk minuman yang akan dikonsumsi. Proses segmentasi ini dilakukan dengan menggunakan metode *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) yang menentukan *cluster* berdasarkan kepadatan data dan mampu mendeteksi *noise* dari data yang mengandung *noise* dan *outlier*.

Berdasarkan penelitian yang dilakukan oleh Qadrini dkk. dengan judul “Metode *K-Means* dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017” menunjukkan bahwa metode DBSCAN memberikan hasil yang lebih baik. Metode DBSCAN menghasilkan nilai *Silhouette Coefficient* sebesar 0,72 pada parameter epsilon 5,5 dan MinPts 1, yang menunjukkan kualitas kluster yang baik dan lebih tinggi dibandingkan dengan hasil pengelompokan menggunakan metode *K-Means*. Kedua metode sama-sama menghasilkan 4 kelompok, namun nilai *silhouette* DBSCAN yang lebih mendekati 1 mengindikasikan bahwa pengelompokan kompetensi Laboratorium telah dilakukan dengan lebih tepat dan optimal [6]. Selanjutnya, penelitian oleh Mutiah dkk. yang berjudul “Perbandingan Metode *Clustering K-Means* dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat” menghasilkan temuan berbeda. *K-Means* menunjukkan performa lebih baik dengan nilai *I* 0.69, menandakan pengelompokan yang lebih terstruktur untuk data dengan karakteristik seragam. Sementara DBSCAN, meski lebih baik dalam menangani *Outlier*, menghasilkan skor *silhouette* lebih rendah (0.398) dengan 7 *cluster* dan 248 *noise points*. Sementara metode DBSCAN lebih cocok untuk data dengan pola distribusi atau variasi yang tidak teratur, temuan menunjukkan bahwa metode *K-Means* lebih baik digunakan pada area dengan akses fasilitas yang relatif merata [7].

Kemudian penelitian yang membahas tentang diabetes dilakukan oleh Razkiazka dan Hasan dengan judul “Analisis Dampak Konsumsi Makanan dan Minuman Berpemanis terhadap Tingkat Diabetes pada Remaja Indonesia” menunjukkan adanya peningkatan signifikan kasus Diabetes Mellitus tipe II. Obesitas dan gaya hidup tidak sehat adalah beberapa penyebab utama. Tiga faktor utama berkontribusi pada peningkatan konsumsi minuman berpemanis: regulasi yang lemah, harga yang terjangkau, dan taktik pemasaran yang agresif. Untuk mengatasi masalah ini dan mengurangi prevalensi Penyakit Tidak Menular (PTM) di kalangan remaja, regulasi harus diperbaiki, pendidikan kesehatan harus ditingkatkan, dan kebijakan harus diubah untuk mengurangi konsumsi gula berlebih [8]. Penelitian yang dilakukan oleh Syafrianto dan Riswanto dengan judul “Pengelompokan Jumlah Kunjungan Mahasiswa ke Perpustakaan Kampus Menggunakan Algoritma DBSCAN” penerapan DBSCAN menggunakan parameter epsilon sebesar 0.4 dan MinPts sebanyak 2 berhasil mengelompokkan mahasiswa berdasarkan jumlah kunjungan perpustakaan dan IPK, menghasilkan 7 kluster serta 5 data yang dikategorikan sebagai *noise*. Hasil analisis menunjukkan bahwa pola pengelompokan lebih dipengaruhi oleh frekuensi kunjungan ke perpustakaan, di mana mahasiswa yang memiliki frekuensi kunjungan 2 hingga 4 kali cenderung memiliki IPK yang lebih tinggi dibandingkan kelompok lainnya [9]. Selain itu, penelitian oleh Nasution dan Rakhmawati yang berjudul “Segmentasi Pengguna *E-Wallet* dengan Menggunakan Metode DBSCAN (*Density Based Spatial Clustering Application with Noise*) di Kota Medan” bertujuan mengelompokkan pengguna dompet digital di Kota Medan menggunakan metode DBSCAN. Menggunakan metode analisis dengan teknik Quota sampling, data dikumpulkan melalui kuesioner *online* dari responden di 10 kecamatan. Proses analisis meliputi pengumpulan data, analisis dengan DBSCAN, dan penarikan kesimpulan. Hasil visualisasi DBSCAN berhasil mengelompokkan pengguna *e-wallet* berdasarkan jenis *e-wallet*, media *top-up*, dan jenis penggunaan. Metode DBSCAN terbukti tepat untuk mengidentifikasi jumlah pengguna *e-wallet* di Kota Medan [10].

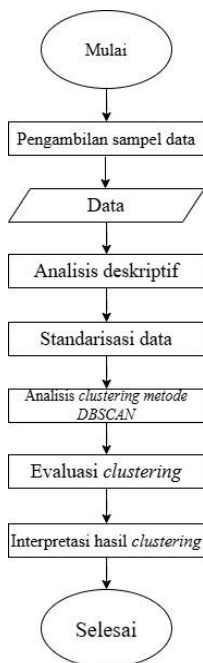
Berdasarkan beberapa studi terdahulu telah melakukan perbandingan antara metode *K-Means* dan DBSCAN untuk berbagai kasus pengelompokan, termasuk data laboratorium, kondisi sosial ekonomi, perilaku pengguna perpustakaan, dan pengguna *e-wallet*. Namun, terdapat celah penelitian yang signifikan terkait penerapan teknik *clustering* dalam mengidentifikasi pola konsumsi minuman yang berhubungan dengan risiko diabetes. Meski studi sebelumnya telah mengkaji hubungan antara konsumsi makanan/minuman manis dengan kejadian diabetes, belum ada yang memanfaatkan metode pembelajaran tanpa pengawasan seperti DBSCAN untuk memetakan profil risiko berdasarkan pola konsumsi. Penelitian ini menjembatani kesenjangan tersebut dengan menerapkan teknik *clustering* untuk mengidentifikasi produk minuman tidak termasuk susu berdasarkan informasi nilai gizi, sehingga dapat menyediakan landasan bagi penentuan keputusan dan kebijakan kesehatan. Nilai gizi yang akan digunakan antara lain lemak jenuh, lemak total, gula, dan garam. Adapun informasi nilai gizi dari beberapa minuman memiliki perbedaan yang signifikan misalnya terdapat minuman yang memiliki kandungan gula rendah bahkan tidak ada, sedangkan

minuman lainnya memiliki kandungan gula tinggi. Hal tersebut dapat menyebabkan terjadinya nilai ekstrim yang dapat memungkinkan data *outlier*. Maka, peneliti menggunakan metode DBSCAN dalam menangani kasus data tersebut.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Berikut merupakan tahapan penelitian menggunakan metode DBSCAN yang disajikan dalam bentuk *flowchart* pada Gambar 1.



Gambar 1. Tahapan Penelitian

- Penelitian dimulai dengan melakukan pengambilan sampel produk minuman bukan termasuk produk susu di tiga toko. Pengambilan data dilakukan dengan memfoto informasi nilai gizi yang terdapat dalam kemasan.
- Kemudian, memasukkan data nilai gizi ke *Microsoft Excel*. Data yang dimasukkan hanya 4 variabel saja yaitu lemak jenuh, lemak total, gula, dan garam. Data tersebut dimasukkan ke RStudio untuk memulai tahap analisis.
- Selanjutnya, melakukan analisis deskriptif menggunakan nilai minimum, maksimum, *mean*, dan *median*.
- Berikutnya, standarisasi data guna mengatasi perbedaan skala antar variabel yang dilakukan dengan menerapkan metode *z-score*.
- Setelah itu, melakukan analisis *clustering* menggunakan metode DBSCAN.
- Melakukan evaluasi *cluster* dari hasil analisis menggunakan *silhouette coefficient* dan *dunn index*.
- Langkah terakhir, menginterpretasikan hasil *clustering* yang diperoleh menggunakan metode DBSCAN.

2.2 Sumber Data

Penelitian ini menggunakan populasi berupa seluruh produk minuman non-susu yang diklasifikasikan dalam kategori pangan sesuai dengan standar yang berlaku. Sampel penelitian merupakan produk minuman tidak termasuk produk susu yang diambil dari tiga toko dengan pertimbangan dapat mewakili segmen pasar yang berbeda-beda. Pengambilan sampel dilakukan menggunakan *purposive sampling*, karena peneliti menentukan toko-toko tempat pengambilan sampel. Toko tersebut antara lain Mirota Kampus Palagan Yogyakarta yang merupakan swalayan (supermarket), Toko Agung Grosir Yogyakarta yang merupakan toko grosir, dan KN Putra Toserba Magelang merupakan toko kelontong. Adapun sampel yang diperoleh dari ketiga toko tersebut sebanyak 355 produk.

2.3 Definisi Operasional Variabel Penelitian

Studi ini menggunakan empat variabel yang termasuk dalam data nilai gizi, yang disajikan dalam Tabel 1.

Tabel 1. Definisi Operasional Variabel

No.	Nama Variabel	Definisi Variabel	Satuan	Skala Data
-----	---------------	-------------------	--------	------------

1.	Lemak Jenuh	Jenis lemak yang umumnya ditemukan dalam produk hewani dan beberapa minyak nabati.		
2.	Lemak Total	Jumlah keseluruhan lemak dalam makanan, termasuk lemak baik (seperti lemak tak jenuh) dan lemak kurang baik (seperti lemak jenuh dan trans).	Gram (g)	Rasio
3.	Gula	Zat pemanis yang dapat meningkatkan kadar gula darah dengan cepat.		
4.	Garam	Bahan mineral yang mengandung natrium.		

2.4 Kategori Pangan

Pengelompokan pangan dilakukan dengan mempertimbangkan sumber bahan utama, metode pengolahan, serta sasaran konsumsinya. Klasifikasi ini memungkinkan identifikasi jenis pangan berdasarkan asal-usulnya, cara pemrosesannya, dan fungsi penggunaannya dalam pola konsumsi masyarakat. Kategori pangan berperan penting dalam penetapan standar serta pengaturan terkait keamanan, kualitas, kandungan gizi, dan pelabelan pada produk pangan olahan. Kategori pangan terdiri atas 16 kategori, salah satunya kategori produk minuman tidak termasuk produk susu yang terdapat dalam kategori 14.0. Menurut Badan Pengawas Obat dan Makanan Republik Indonesia, kategori produk minuman tidak termasuk produk susu terbagi menjadi dua kategori yaitu minuman (ringan) tidak beralkohol dan minuman beralkohol. Namun, minuman beralkohol tidak ditemukan informasi nilai gizi dalam kemasan. Lima kategori minuman (ringan) tidak beralkohol: air minum dan air berkarbonat, sari buah dan sayur, nektar buah dan sayur, minuman berbasis air berperisa, dan minuman yang disiapkan dengan penyeduhan atau perendaman, seperti teh dan kopi [11].

2.5 Informasi Nilai Gizi

Informasi nilai gizi (*nutrition facts*) merupakan label yang tercantum pada kemasan produk pangan atau minuman, berisi rincian mengenai kandungan zat gizi di dalam produk tersebut. Label ini berfungsi sebagai acuan bagi konsumen dalam membuat keputusan pembelian yang lebih bijak, khususnya bagi individu yang sedang mengontrol asupan kalori atau memiliki kondisi kesehatan tertentu. Meskipun informasi yang disediakan cukup lengkap, tidak sedikit konsumen yang mengalami kesulitan dalam memahami label tersebut. Hal ini diperumit oleh fakta bahwa satu kemasan produk, seperti minuman, umumnya mengandung lebih dari satu porsi, sementara informasi gizi yang ditampilkan hanya merujuk pada satu porsi saja. Beberapa zat gizi seperti lemak jenuh, lemak total, gula tambahan, dan natrium merupakan komponen yang perlu dikendalikan konsumsinya dari makanan dan minuman kemasan. Hal ini karena kebutuhan harian terhadap zat-zat tersebut umumnya sudah tercukupi dari konsumsi minuman atau makanan harian lainnya, sehingga asupan tambahan dari produk kemasan sebaiknya dibatasi [12].

2.6 Standarisasi Data

Data sering kali memiliki variasi yang besar dalam satuan pengukurannya. Perbedaan skala yang signifikan antar variabel dapat menyebabkan bias dalam proses analisis data dan mempengaruhi hasil. Oleh karena itu, sebelum melakukan analisis data, perlu dilakukan proses standarisasi data terlebih dahulu. Salah satu metode standarisasi data yang umum digunakan adalah transformasi *z-score*. *Z-score* mengubah data menjadi unit-unit standar deviasi, sehingga semua variabel memiliki rata-rata 0 dan standar deviasi konstan [13]. Berikut merupakan rumus untuk mengetahui nilai *z-score*.

$$z_i = \frac{x_i - \bar{x}}{s} \quad (1)$$

dimana i merupakan data ke-1 hingga n , x_i menunjukkan nilai data ke- i , \bar{x} merupakan nilai rata-rata, dan s adalah nilai standar deviasi.

2.7 Density-Based Spatial Clustering of Application with Noise (DBSCAN)

DBSCAN adalah teknik klastering yang mengelompokkan data berdasarkan kepadatan titik yang berdekatan, dengan mendeteksi wilayah yang memiliki konsentrasi data tinggi sebagai klaster dan mengabaikan titik yang terpisah sebagai *noise*. Metode ini membentuk *cluster* dengan mensyaratkan setiap objek dalam radius tertentu harus memiliki jumlah minimum data yang ditentukan. Dalam prosesnya, DBSCAN secara otomatis akan mengidentifikasi data-data yang tidak memenuhi kriteria *cluster* sebagai *noise*, sehingga memungkinkan pemisahan yang jelas antara data yang termasuk dalam *cluster* dan yang tidak termasuk [14]. Berikut komponen-komponen yang terdapat dalam DBSCAN untuk proses analisis *clustering* [15].

- a. Epsilon
Epsilon merupakan parameter yang menentukan jarak maksimum antara dua titik serta menentukan seberapa jauh jangkauan pencarian tetangga.
- b. Minimum Points (MinPts)
MinPts adalah nilai minimal atau batas minimum yang menentukan berapa banyak titik data yang harus ada dalam suatu lingkungan (*neighborhood*) agar bisa membentuk sebuah kelompok dalam algoritma DBSCAN.
- c. Core Point

Core Point merupakan titik data yang berada di pusat kepadatan dengan jumlah tetangga di sekitarnya sama atau melebihi nilai *MinPts* yang telah ditentukan [16].

d. *Border Point*

Border Point adalah titik data yang berada di pinggir kelompok. Titik ini memiliki beberapa tetangga, namun jumlahnya tidak mencukupi syarat *MinPts*.

e. *Noise*

Noise adalah titik data yang terisolasi, artinya tidak memiliki tetangga sama sekali dalam radius yang ditentukan.

Penentuan parameter epsilon dan *MinPts* yang optimal dapat dilakukan melalui metode *grid search*, yakni dengan mengevaluasi kombinasi nilai parameter secara sistematis untuk menemukan konfigurasi yang menghasilkan kualitas *cluster* terbaik berdasarkan hasil evaluasi seperti *Silhouette Coefficient* atau lainnya. Pendekatan *grid search* memungkinkan identifikasi parameter optimal secara objektif, sehingga menghasilkan klusterisasi yang lebih *robust* dan terukur. Penelitian ini melakukan percobaan kombinasi dengan epsilon dari 0.2 hingga 1 dan *MinPts* dari 2 hingga 5. Berikut merupakan langkah-langkah perhitungan manual dalam melakukan analisis menggunakan metode DBSCAN [17].

a. Tentukan nilai awal untuk parameter *MinPts* dan Eps.

b. Pilih secara acak satu titik awal *z*.

c. Hitung jarak Eps dengan menggunakan rumus jarak *Euclidean* berikut.

$$d_{ij} = \sqrt{(x_{iLT} - x_{jLT})^2 + (x_{iLJ} - x_{jLJ})^2 + (x_{iGaram} - x_{jGaram})^2 + (x_{iGula} - x_{jGula})^2} \quad (2)$$

dimana d_{ij} artinya nilai jarak *euclidean*, x_{iLT} merupakan variabel lemak total dari objek i ($i = 1, 2, \dots, n$), x_{jLT} merupakan variabel lemak total dari objek j ($j = 1, 2, \dots, n$), x_{iLJ} menunjukkan variabel lemak jenuh dari objek i ($i = 1, 2, \dots, n$), dan x_{jLJ} artinya variabel lemak jenuh dari objek j ($j = 1, 2, \dots, n$).

d. Jika jumlah titik dalam Eps lebih dari *MinPts*, maka titik z adalah *core point* dan membentuk sebuah *cluster*.

e. Langkah 3 dan 4 diulangi hingga semua titik diproses. Jika z adalah *border point* dan tidak ada titik lain yang *density reachable* terhadap z , lanjutkan ke titik selanjutnya.

2.8 Evaluasi Clustering

2.8.1 Silhouette Coefficient

Silhouette coefficient merupakan teknik evaluasi untuk menilai efektivitas dan ketepatan pengelompokan dalam *cluster*. Pendekatan ini mengkombinasikan dua aspek pengukuran: kohesi, yang mengevaluasi tingkat kedekatan relasi antar objek dalam satu *cluster*, dan separasi, yang mengukur tingkat pemisahan antar *cluster*. Adapun langkah-langkah untuk melakukan perhitungan *silhouette coefficient* sebagai berikut [18].

a. Hitung jarak rata-rata antara objek dengan semua objek lain yang berada dalam satu *cluster*.

$$a(i) = \frac{1}{|C_i|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (3)$$

dengan $a(i)$ adalah jarak rata-rata antara setiap objek dalam *cluster* dengan semua objek lainnya, j artinya objek dalam *cluster*, $|C_i|$ merupakan nilai absolut dari seluruh data yang tergolong dalam kluster ke- i , dan $d(i, j)$ menunjukkan jarak antara data i dengan j .

b. Hitung rata-rata jarak setiap objek terhadap seluruh objek yang berada dalam kluster lain, kemudian mengambil nilai rata-rata terkecil sebagai representasi kedekatan objek tersebut dengan kluster terdekat di luar klusternya sendiri.

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j) \quad (4)$$

dengan $b(i)$ merupakan rata-rata jarak objek dengan objek lain yang berada di *cluster* yang berbeda dan C_k merupakan *cluster* yang berbeda dengan *cluster* yang mengandung objek ke- i .

c. Hitung nilai koefisien *silhouette*.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

dengan $s(i)$ merupakan seluruh rata-rata pada *dataset*.

Hasil perhitungan koefisien *silhouette* memiliki rentang nilai dari -1 hingga 1. Suatu *clustering* dikatakan berkualitas baik jika nilai koefisien *silhouette* positif $a(i) < b(i)$ dan $a(i)$ mendekati 0. Nilai maksimal koefisien *silhouette* yaitu 1, dicapai ketika $a(i) = 0$, yang menunjukkan bahwa objek berada dalam *cluster* yang paling sesuai [19].

2.8.2 Dunn Index

Indeks dunn (DI), yang pertama kali diperkenalkan oleh JC Dunn pada tahun 1974, adalah metrik untuk mengevaluasi algoritma *clusterisasi* dengan mengandalkan data *cluster* itu sendiri. Sebagai metode evaluasi internal, Indeks Dunn

bertujuan untuk menemukan *cluster* yang kompak, dimana titik-titik dalam *cluster* memiliki variasi kecil, dan *cluster-cluster* tersebut terpisah dengan jelas, dengan jarak antar *cluster* yang jauh dibandingkan dengan variasi di dalam *cluster*. Indeks dunn dapat didefinisikan dengan rumus berikut [20].

$$D = \left\{ \frac{\min_{1 \leq i \leq k} \left(d(c_i, c_j) \right)}{\max_{1 \leq l \leq q} d(c_l)} \right\} \tag{6}$$

dengan $d(c_i, c_j)$ yang mengukur kedekatan antar *cluster* i dengan *cluster* j , dan $d(c_l)$ yang mengukur kedekatan antar anggota dalam *cluster* l .

2.9 Uji t

Analisis uji-t merupakan teknik statistik untuk membandingkan rata-rata. Metode ini membandingkan nilai tengah disertai interval kepercayaan spesifik antara dua kelompok populasi. Esensi dari pengujian ini adalah evaluasi variasi perbedaan di antara dua kelompok data tersebut. Kesetaraan atau perbedaan varian antar kelompok menjadi informasi krusial dalam proses pengujian. Karakteristik varian dari masing-masing kelompok data berdampak pada perhitungan standar eror, yang kemudian menjadi faktor penentu dalam pemilihan formula pengujian yang sesuai. [21]. Uji t pada dasarnya mengambil variasi data dari masing-masing kelompok untuk membandingkan perbedaan rata-rata. Uji t bertujuan untuk menentukan apakah perbedaan yang diamati antara dua kelompok itu signifikan secara statistik atau hanya kebetulan [22].

3. HASIL DAN PEMBAHASAN

DBSCAN adalah salah satu algoritma pengelompokan yang efektif dalam mengidentifikasi pola dan struktur data berdasarkan kepadatan, dengan kemampuan untuk mengelompokkan data yang memiliki distribusi tidak teratur dan mengabaikan data pencilan. Metode ini dipilih karena kemampuannya dalam menangani data dengan kepadatan yang berbeda serta kesederhanaannya dalam mengidentifikasi kelompok yang relevan. Pembahasan ini akan mencakup tahapan-tahapan penerapan DBSCAN, hasil yang diperoleh, serta interpretasi terhadap pola-pola yang muncul dalam data yang dianalisis.

3.1 Deskripsi Data

Data produk minuman yang digunakan dalam penelitian ini merupakan produk minuman tidak termasuk susu sesuai dengan kategori produk 14.0 berdasarkan Peraturan Badan Pengawas Obat dan Makanan (BPOM) Nomor 13 Tahun 2023 tentang Kategori Pangan. Data yang digunakan sebanyak 355 data, yang disajikan dalam **Tabel 2**.

Tabel 2. Data Produk Minuman

No	Produk	Lemak Total (gram)	Lemak Jenuh (gram)	Gula (gram)	Garam (gram)
1	Pocari Sweat	0	0	14	0.12
2	Isoplus	0	0	10	0.085
3	Le Minerale	0	0	0	0.02
4	MIZONE White Isotonic Active	0	0	8	0.065
...
350	Excelso Robusta Gold	1	0	0	0
351	Excelso Arabica Gold	1.5	0.5	0	0
352	Indocafe Original Bland	0	0	0	0
353	Kapal Api Special	0.5	0	0	0
354	Kapal Api Lampung	0.5	0	0	0
355	Kapal Api One Kopi + Gula	0.5	0	17	0

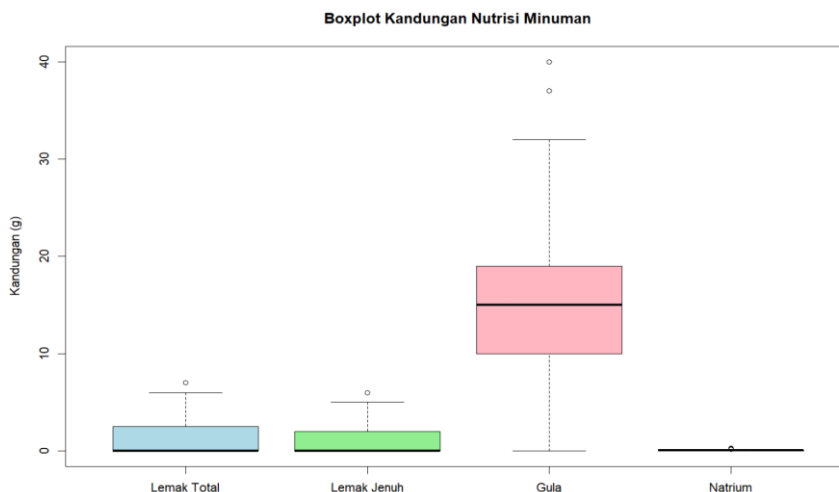
Analisis deskriptif merupakan metode dalam mengorganisasi dan menyajikan data secara sistematis melalui tabel, grafik, dan visualisasi lainnya. Dalam penelitian ini, analisis deskriptif dilakukan untuk memberikan gambaran umum mengenai informasi nilai gizi (ING) dari produk minuman, yang divisualisasikan dalam bentuk *boxplot*.

Tabel 3. Statistik Deskriptif

Variabel	Minimum (g)	Maksimum (g)	Mean (g)	Median (g)
Lemak Total	0	7	1.216	0
Lemak Jenuh	0	6	0.9648	0
Gula	0	40	14.71	15
Garam	0	0.29	0.0514	0.0350

Pada Tabel 3 statistik deskriptif terhadap nilai gizi minuman menunjukkan bahwa kandungan lemak total dalam produk minuman memiliki nilai minimum 0 gram dan maksimum 7 gram, dengan rata-rata 1.216 gram dan median 0 gram. Hal ini mengindikasikan bahwa sebagian besar produk minuman memiliki kadar lemak total yang sangat rendah atau tidak mengandung lemak sama sekali, jauh di bawah batas konsumsi harian yang disarankan. Kandungan lemak jenuh juga menunjukkan pola serupa dengan nilai minimum 0 gram dan maksimum 6 gram, rata-rata 0.9648 gram, serta median 0 gram, yang mencerminkan rendahnya kandungan lemak jenuh dalam sebagian besar minuman.

Kandungan gula dalam produk minuman bervariasi, dengan nilai minimum 0 gram dan maksimum 40 gram, rata-rata 14.71 gram, dan median 15 gram. Produk dengan kandungan gula paling tinggi tersebut adalah Orange Water. Meskipun masih di bawah batas konsumsi harian gula yang disarankan, nilai median yang lebih tinggi dari rata-rata menunjukkan bahwa lebih dari setengah jumlah produk mengandung kadar gula yang relatif tinggi, dengan beberapa produk yang memiliki kadar gula rendah atau tanpa gula. Kandungan garam dalam produk minuman memiliki nilai minimum 0 gram dan maksimum 0.29 gram, dengan rata-rata 0.0514 gram dan median 0.035 gram, menunjukkan bahwa kandungan garam dalam produk minuman cenderung rendah dan konsisten, jauh di bawah batas konsumsi harian yang disarankan oleh Kemkes. Produk minuman disajikan dalam bentuk visualisasi *boxplot* pada **Gambar 2** berikut.



Gambar 2. Visualisasi *Boxplot*

Berdasarkan Gambar 2 visualisasi *boxplot* yang disajikan, terdapat variasi nilai gizi GGL pada produk minuman yang dianalisis. Kandungan lemak total dan lemak jenuh terlihat cenderung rendah, dengan sebagian besar data mendekati nilai 0 gram. Lemak total dan lemak jenuh memiliki *outlier* masing-masing 1 dan 2 data, yang mengindikasikan adanya produk dengan kandungan lemak yang lebih tinggi dibandingkan mayoritas produk lainnya. Kandungan gula menunjukkan distribusi yang lebih lebar dengan median sekitar 15 gram, dan terdapat *outlier* sebanyak 3 data yang mendekati 40 gram. Hal ini menunjukkan bahwa sebagian besar produk memiliki kadar gula yang cukup tinggi, meskipun terdapat produk dengan kadar gula sangat rendah atau bahkan 0 gram. Kandungan natrium (garam) dalam produk minuman terlihat sangat rendah dan konsisten, dengan sebagian besar nilai mendekati 0 gram, serta terdapat 16 *outlier*. Secara keseluruhan, visualisasi ini mengindikasikan bahwa *dataset* yang digunakan mengandung beberapa *outlier*, serta nilai nol yang cukup besar khususnya pada lemak dan natrium.

3.2 Standarisasi Data

Visualisasi *boxplot* menunjukkan variabel gula memiliki distribusi lebih lebar dibandingkan lemak total, lemak jenuh, dan natrium yang nilainya mendekati nol. Perbedaan skala antar variabel dapat menyebabkan bias analisis, dengan variabel berskala besar mendominasi hasil. Standarisasi data diperlukan untuk menyelaraskan skala variabel, memastikan kontribusi setara dalam analisis lanjutan, dan mengurangi dampak *outlier*.

Tabel 4. Hasil Standarisasi Kandungan Gizi Produk Minuman

No	Produk	Lemak Total (gram)	Lemak Jenuh (gram)	Gula (gram)	Garam (gram)
1	Pocari Sweat	-0.75537	-0.70856	-0.10440	1.41574
2	Isoplus	-0.75537	-0.70856	-0.69267	0.69304
3	Le Minerale	-0.75537	-0.70856	-2.16334	-0.64913
...
300	Luwak White Koffie Premium	0.79790	1.12750	-1.42800	-0.75237
301	ABC Kopi Klepon	1.41922	1.86192	0.04267	-0.75237
302	Good Day Coffee Freeze Mocafrio	1.41922	1.86192	0.92507	-0.95886



No	Produk	Lemak Total (gram)	Lemak Jenuh (gram)	Gula (gram)	Garam (gram)
...
352	Indocafe Original Bland	-0.75537	-0.70856	-2.16334	-1.06210
353	Kapal Api Special	-0.44472	-0.70856	-2.16334	-1.06210
354	Kapal Api Lampung	-0.44472	-0.70856	-2.16334	-1.06210
355	Kapal Api One Kopi + Gula	-0.44472	-0.70856	0.33680	-1.06210

Hasil standarisasi data pada Tabel 4 menunjukkan bahwa GGL seluruh produk minuman telah diubah menjadi unit-unit standar deviasi menggunakan transformasi *z-score*. Oleh karena itu, seluruh variabel memiliki rata-rata 0 dan standar deviasi konstan.

3.3 Analisis Metode DBSCAN

Berdasarkan hasil statistik deskriptif yang menunjukkan adanya variasi signifikan pada GGL produk minuman memperlihatkan bahwa keberadaan *outlier* dan distribusi data yang tidak merata, diperlukan analisis *clustering* yang dapat menangani data dengan kepadatan yang bervariasi. Analisis menggunakan metode DBSCAN menjadi pilihan yang tepat karena mampu mengelompokkan data berdasarkan kepadatan tanpa harus menentukan jumlah *cluster* di awal, serta dapat menangani *noise* dengan efektif. Sebelum melakukan analisis menggunakan metode DBSCAN, proses *tuning* parameter dilakukan untuk mempermudah dalam memperoleh nilai epsilon dan minPts yang optimal. Teknik *tuning* parameter adalah *grid search* yang bekerja dengan mencari dan menentukan parameter optimal.

Tabel 5. Hasil Tuning Parameter

Epsilon	MinPts	Jumlah Cluster	Silhouette Coefficient	Dunn Index
0.2	2	31	0.277259	0.128552
0.2	3	11	0.273821	0.128552
0.2	4	8	0.420122	0.067355
...
0.7	2	10	0.07302	0.125278
0.7	3	6	0.083941	0.125278
0.7	4	2	0.39553	0.137292
...
1	4	2	0.378428	0.27216
1	5	2	0.383192	0.281385

Hasil *tuning* parameter yang disajikan pada Tabel 5, menunjukkan bahwa dilakukan kombinasi parameter yang menghasilkan kualitas *cluster* terbaik dalam algoritma DBSCAN dengan kombinasi nilai epsilon dari 0.2 hingga 1 dan minPts dari 2 hingga 5. Berdasarkan hasil tersebut diperoleh bahwa nilai epsilon sebesar 0.7 dan nilai batas minimum dalam kelompok atau minPts sebanyak 4 titik (ditandai dengan warna hijau), menghasilkan 2 *cluster* dengan nilai *Silhouette Coefficient* sebesar 0.39553 dan *Dunn Index* sebesar 0.137292. Kombinasi tersebut menunjukkan keseimbangan yang baik antara jumlah *cluster* dan kualitas *cluster*. Nilai epsilon 0.7 artinya dua produk minuman akan dianggap berdekatan atau bertetangga jika jarak di antara keduanya tidak lebih dari 0.7. Sementara itu, nilai minPts sebanyak 4 artinya suatu produk minuman akan menjadi *core point* dan membentuk *cluster* apabila minimal terdapat 4 produk (termasuk produk itu sendiri) terletak dalam radius 0.7 tersebut. Apabila produk tidak memenuhi nilai epsilon dan minPts, maka dianggap sebagai *noise*. Adapun perhitungan jarak dilakukan menggunakan jarak *euclidean*, dengan contoh perhitungannya menggunakan produk Pocari Sweat dan Isoplus berikut.

$$d_{PI} = \sqrt{(-0.755 - (-0.755))^2 + (-0.709 - (-0.709))^2 + (-0.104 - (-0.693))^2 + (1.416 - 0.693)^2} = 0.933 \tag{7}$$

Berdasarkan jarak *euclidean*, produk Pocari Sweat dan Isoplus menunjukkan nilai kedekatan sebesar 0.933 yang mengindikasikan bahwa kedua produk tersebut berada dalam radius epsilon yang telah ditetapkan. Proses perhitungan jarak kemudian diimplementasikan pada seluruh produk untuk mengevaluasi kondisi densitas berdasarkan parameter minPts, berikut diperoleh hasil pembentukan *cluster* setelah melalui proses tersebut dalam Gambar 3.



Gambar 3. Visualisasi Persebaran Cluster

Pada Gambar 3 ditampilkan hasil *clustering* DBSCAN dengan parameter epsilon 0.7 dan minPts 4, yang menghasilkan 2 *cluster*. *Cluster* 1 ditandai dengan warna merah dan memiliki jumlah data lebih besar dengan area cakupan yang luas. *Cluster* 2 ditandai dengan warna biru, memiliki data lebih sedikit dengan area cakupan yang lebih sempit. Titik-titik hitam yang berada di luar kedua *cluster* tersebut merupakan *noise*. Hasil *clustering* tersebut diuraikan dalam Tabel 6 berikut.

Tabel 6. Hasil *Clustering* DBSCAN

Cluster	Jumlah Produk	Produk
0/Noise	42	Orange Water, Thai Milk Coffee, Ichitan Brown Sugar Milk, Ichitan Thai Milk Tea, Nutriboost Stroberi, Nutriboost Jeruk, Good Day Original Cappuccino, Golda Coffee Cappuccino, dan lain-lain.
1	287	Pocari Sweat, Isoplus, Le Minerale, MIZONE White Isotonic Active, You C1000 Mango, Cool Time Coconut, Adem Sari Herbal Lemon, Buavita Jambu, Buavita Mangga, Buavita Jeruk, Buavita Leci, dan lain-lain.
2	26	NU Milk Tea, NU Teh Tarik, Luwak White Koffie Original, Kopi Kenangan Kopi Susu Alpukat, Kopi Kenangan Cappuccino Mantancino, Nescafe Latte, Nescafe Caramel Macchiato, Nescafe Cappuccino, Teh Poci Teh Tarik Gula Aren, dan lain-lain.

Hasil *clustering* secara keseluruhan dapat diakses pada [link https://bit.ly/ProdukMinumanTidakTermasukProdukSusu](https://bit.ly/ProdukMinumanTidakTermasukProdukSusu), agar memudahkan dalam pencarian produk.

3.4 Evaluasi Hasil *Clustering*

Evaluasi hasil *clustering* merupakan tahap untuk menilai kualitas *cluster* yang dihasilkan. Metode evaluasi yang digunakan adalah *silhouette coefficient* dan *dunn index*. *Silhouette coefficient* memiliki rentang nilai dari -1 hingga 1, sedangkan dalam *dunn index* semakin tinggi nilainya, semakin baik kualitas *cluster* yang terbentuk, karena menunjukkan *cluster* yang lebih terpisah dan lebih padat. Berikut disajikan visualisasi *silhouette coefficient* pada Gambar 4 dari setiap produk dalam *cluster*.



Gambar 4. Visualisasi Persebaran Cluster

Sebagian besar data dalam *cluster* pertama memiliki nilai *silhouette* yang positif, tetapi terdapat beberapa titik dengan nilai mendekati nol atau bahkan negatif yang menunjukkan bahwa beberapa data kurang terdefinisi dengan jelas dalam kelompoknya. Di sisi lain, *cluster* kedua memiliki ukuran yang lebih kecil dan sebagian besar titiknya memiliki *silhouette score* yang lebih tinggi, menandakan bahwa *cluster* ini lebih kompak dan lebih terpisah dengan baik dibandingkan *cluster* pertama. Adapun produk yang memiliki nilai *silhouette* minus sebanyak 35 produk, seperti Tora Susu Kopi Susu Full Cream, Good Day Funtastic Mocacino, TOP Cappuccino, Tora Moka Kopi Susu & Cokelat, dan seterusnya. Adapun perbandingan hasil evaluasi *clustering* disajikan dalam **Tabel 7** berikut.

Tabel 7. Nilai Evaluasi *Cluster*

<i>Silhouette Coefficient</i>	<i>Dunn Index</i>
0.39553	0.137

Pada **Tabel 7** nilai *silhouette coefficient* sebesar 0.39553 (positif namun tergolong sedang) mengindikasikan sebagian besar data berada pada *cluster* yang sesuai, meskipun jarak antar *cluster* belum terlalu jelas atau terdapat beberapa titik yang berdekatan antar *cluster*. Nilai *dunn index* sebesar 0.137 menunjukkan rasio antara jarak minimum antar *cluster* dan jarak maksimum dalam *cluster* masih relatif rendah, yang berarti terdapat kedekatan antar *cluster* dan sebaran data dalam *cluster* yang cukup bervariasi.

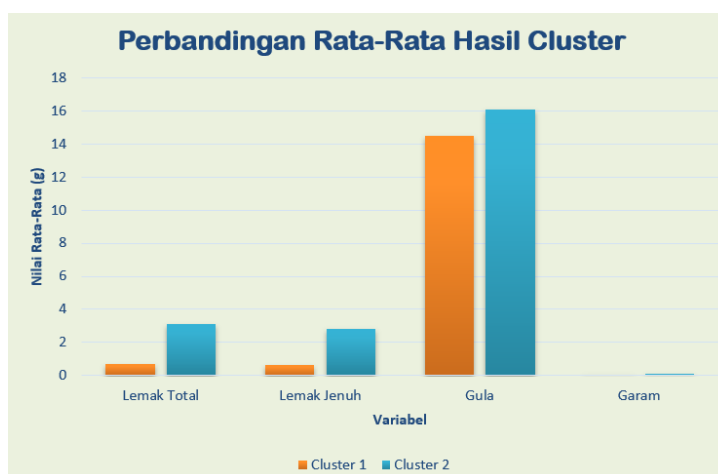
3.5 Profilisasi Hasil *Clustering*

Berikutnya, melakukan profilisasi dari hasil *clustering* analisis DBSCAN menggunakan bantuan *software Microsoft Excel* dengan hasil pada **Tabel 8** berikut.

Tabel 8. Perbandingan Rata-Rata Antar *Cluster*

<i>Cluster</i>	Lemak Total (g)	Lemak Jenuh (g)	Gula (g)	Garam (g)	Kategori
1	0.68502	0.60627	14.5226	0.03439	Rendah
2	3.096154	2.807692	16.07692	0.127308	Tinggi

Berdasarkan **Tabel 8**, hasil profilisasi dari masing-masing *cluster* menunjukkan bahwa *cluster* 1 termasuk dalam kategori rendah, sedangkan *cluster* 2 dikategorikan tinggi. Hal ini disebabkan oleh rata-rata kandungan gula, garam, dan lemak (GGL) dalam *cluster* 2 lebih tinggi dibanding *cluster* 1. Oleh karena itu, sebanyak 287 produk dalam *cluster* 1 memiliki rata-rata kandungan GGL yang relatif rendah, sehingga kemungkinan menyebabkan risiko penyakit diabetes lebih kecil dibandingkan dengan produk yang terdapat dalam *cluster* 2. Meskipun risiko pada *cluster* 1 tergolong rendah, konsumsi produk secara berlebihan tetap dapat meningkatkan potensi terjadinya gejala diabetes. Berikut merupakan visualisasi perbandingan rata-rata kandungan GGL antar *cluster* disajikan pada **Gambar 5**.



Gambar 5. Visualisasi Hasil Profilisasi *Cluster*

Perbandingan rata-rata kandungan GGL antar *cluster* yang disajikan dalam **Gambar 5** menunjukkan bahwa *cluster* 2 memiliki rata-rata kandungan yang jauh lebih tinggi dibandingkan *cluster* 1, meskipun kandungan gula pada *cluster* 1 juga tergolong tinggi.

3.6 Uji t

Berikutnya dilakukan uji *t* setelah proses profilisasi *clustering* untuk menguji signifikansi perbedaan rata-rata antar kelompok hasil *clustering*. Uji *t* digunakan untuk menentukan apakah terdapat perbedaan yang signifikan dalam kandungan gula, garam, dan lemak (GGL) antara dua *cluster*. Oleh karena itu, dilakukan pengujian hipotesis guna membuktikan perbedaan antar *cluster*.

a. Hipotesis

$H_0: \mu_1 = \mu_2$ (Tidak terdapat perbedaan signifikan dalam kandungan GGL antara dua *cluster*)

$H_1: \mu_1 \neq \mu_2$ (Terdapat perbedaan signifikan dalam kandungan GGL antara dua *cluster*)

b. Tingkat Signifikansi

$\alpha = 0.05$

c. Daerah Kritis

Tolak H_0 jika $|t_{hitung}| > t_{\alpha/2}$ atau $p\text{-value} < \alpha$

d. Statistik Uji

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

(8)

e. Keputusan

Berikut keputusan dari hasil uji t pada Tabel 9.

Tabel 9. Keputusan Uji t

Variabel	μ_1	μ_2	$p\text{-value}$	Tanda	α	Keputusan
Lemak Total	-0.15866	1.75136	$< 2.2 \times 10^{-16}$	<	0.05	Tolak H_0
Lemak Jenuh	-0.15557	1.71726	$< 2.2 \times 10^{-16}$	<		
Gula	-0.02031	0.22422	0.02213	<		
Garam	-0.20302	2.24101	$< 2.2 \times 10^{-16}$	<		

f. Kesimpulan

Dengan menggunakan tingkat kepercayaan 95% data yang ada dikatakan tolak H_0 yang artinya terdapat perbedaan signifikan dalam kandungan GGL antar *cluster* 1 dengan *cluster* 2.

4. KESIMPULAN

Penelitian ini berhasil mencapai tujuan utamanya yaitu melakukan segmentasi produk minuman tidak termasuk susu berdasarkan kandungan gizi menggunakan algoritma DBSCAN. Sebagian besar produk minuman memiliki kandungan lemak total dan lemak jenuh yang rendah, sesuai dengan batas konsumsi harian yang disarankan, sementara kandungan gula cenderung lebih tinggi dengan sebagian besar produk mengandung kadar gula mendekati 15 gram, dan kandungan garam relatif konsisten dan rendah. Visualisasi *boxplot* menunjukkan adanya variasi nilai gizi dengan beberapa *outlier*, terutama pada kandungan gula dan garam, serta banyaknya nilai nol pada lemak dan garam, mencerminkan rendahnya nilai gizi tersebut dalam sebagian besar produk. Proses *clustering* dimulai dengan standarisasi data menggunakan transformasi *z-score* guna mengatasi potensi bias akibat perbedaan skala, data produk minuman berhasil melakukan *clustering* DBSCAN yang terbagi menjadi dua *cluster* dan satu *cluster noise - cluster* 0 sebagai *noise* sebanyak 42 produk, *cluster* 1 dengan 287 produk merupakan kelompok terbanyak, dan *cluster* 2 terdiri dari 26 produk. Evaluasi *clustering* menunjukkan hasil terbaik dengan *silhouette coefficient* sebesar 0.396 dan *dunn index* sebesar 0.137, dimana *silhouette coefficient* positif mengindikasikan sebagian besar data telah terkelompok dengan baik meskipun jarak antar *cluster* belum sepenuhnya jelas, sedangkan *dunn index* yang rendah menunjukkan kedekatan antar *cluster* dan sebaran data yang bervariasi, namun secara keseluruhan performa *clustering* dinilai cukup baik. Profilisasi *cluster* menunjukkan bahwa *cluster* 1 dengan 287 produk memiliki kandungan GGL rendah sehingga risiko diabetes lebih kecil dibandingkan *cluster* 2 yang memiliki kandungan GGL lebih tinggi, dengan hasil uji hipotesis mengonfirmasi perbedaan signifikan kandungan GGL antara kedua *cluster*, dimana *cluster* 2 memiliki kadar gula, garam, dan lemak yang jauh lebih tinggi. Hasil tersebut dapat menjadi acuan masyarakat dalam memilih produk yang akan dikonsumsi untuk mencegah meningkatnya penyakit diabetes di Indonesia. Pemerintah juga dapat menjadikan hasil penelitian ini menjadi acuan dalam rencana pembuatan *Nutri-Level* seperti *NutriGrade* yang dimiliki oleh Singapura. Namun, penelitian ini masih terbatas dalam data karena tidak mewakili produk yang tersebar di seluruh Indonesia. Selain itu, untuk meningkatkan hasil *clustering*, perlu dilakukan eksplorasi metode *clustering* lain yang lebih tepat. Dengan demikian, diharapkan hasil penelitian dapat lebih komprehensif dan menghasilkan segmentasi yang lebih akurat, yang kemudian dapat digunakan untuk penerapan klasifikasi yang lebih efektif.

REFERENCES

- [1] World Health Organization, "Diabetes." Accessed: Nov. 07, 2024. [Online]. Available: https://www.who.int/health-topics/diabetes#tab=tab_1
- [2] IDF Diabetes Atlas, "IDF Diabetes Atlas 10th Edition." Accessed: Nov. 05, 2024. [Online]. Available: <https://diabetesatlas.org/>
- [3] Administrator, "Cegah Dini Ancaman Diabetes." Accessed: Nov. 05, 2024. [Online]. Available: <https://indonesia.go.id/kategori/editorial/8401/cegah-dini-ancaman-diabetes?lang=1>
- [4] Purwowitz, "Menakar Pembatasan Minuman Berpemanis Dalam Kemasan." Accessed: Nov. 05, 2024. [Online]. Available: <https://mediakeuangan.kemenkeu.go.id/article/show/menakar-pembatasan-minuman-berpemanis-dalam-kemasan>



- [5] Rokom, “Konsumsi Gula Berlebih, Waspada! Risikonya.” Accessed: Nov. 15, 2024. [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20220927/2841159/konsumsi-gula-berlebih-waspada-risikonya/>
- [6] L. Qadrini *et al.*, “Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017,” *J Statistika*, vol. 13, no. 2, pp. 5–11, 2020, doi: <https://doi.org/10.36456/jstat.vol13.no2.a2886>.
- [7] S. Mutiah, Y. Hasnataeni, A. Fitrianto, L. Risman Dwi Jumansyah, and S. dan Sains, “Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat,” *Teorema: Teori dan Riset Matematika*, vol. 09, no. 02, pp. 247–260, Sep. 2024, doi: <http://dx.doi.org/10.25157/teorema.v9i2.16290>.
- [8] M. R. M. Hasan, “Analisis Dampak Konsumsi Makanan dan Minuman Berpemanis terhadap Tingkat Diabetes Pada Remaja Indonesia,” *MagnaSalus: Jurnal Keunggulan*, vol. 06, no. 4, Oct. 2024, [Online]. Available: <https://journalpedia.com/1/index.php/jkk>
- [9] A. Syafrianto and E. Riswanto, “Pengelompokkan Jumlah Kunjungan Mahasiswa ke Perpustakaan Kampus Menggunakan Algoritma DBSCAN,” *G-Tech: Jurnal Teknologi Terapan*, vol. 7, no. 1, pp. 75–81, Jan. 2023, doi: <https://doi.org/10.33379/gtech.v7i1.1925>.
- [10] N. Nasution and F. Rakhmawati, “Segmentasi Pengguna E-Wallet dengan Menggunakan Metode DBSCAN (Density Based Spatial Clustering Application with Noise) di Kota Medan,” *Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 4, no. 2, Aug. 2023, doi: <https://doi.org/10.46306/lb.v4i2>.
- [11] Badan Pengawas Obat dan Makanan Republik Indonesia, *Peraturan BPOM nomor 13 Tahun 2023*. 2023.
- [12] M. Bahrudin, “Cara Membaca Informasi Nilai Gizi pada Kemasan Makanan.” Accessed: Nov. 07, 2024. [Online]. Available: [https://perpustakaan.bsn.go.id/index.php?p=news&id=1486#:~:text=Informasi%20nilai%20gizi%20\(nutrition%20facts,konsumen%20untuk%20membeli%20suatu%20barang](https://perpustakaan.bsn.go.id/index.php?p=news&id=1486#:~:text=Informasi%20nilai%20gizi%20(nutrition%20facts,konsumen%20untuk%20membeli%20suatu%20barang).
- [13] H. Astuti, “Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Pengelompokkan Data Pelanggan (Studi Kasus: PT. Pinus Merah Abadi),” *Jurnal Web Informatika Teknologi*, vol. 6, no. 1, pp. 1–8, Jun. 2021, [Online]. Available: <https://ejurnal-wit.ac.id/index.php/J-WIT/article/view/51>
- [14] T. D. Harjanto, A. Vatesia, and R. Faurina, “Analisis Penetapan Skala Prioritas Penanganan Balita Stunting Menggunakan Metode DBSCAN Clustering,” *Jurnal Rekursif*, vol. 9, no. 1, Mar. 2021, doi: <https://doi.org/10.33369/rekursif.v9i1.14982>.
- [15] D. Fitriyah, W. Gunawan, and R. Algian Kurniaputra, “Implementasi Algoritma DBScan dalam Pemngambilan Data Menggunakan Scatterplot,” *chno Xplore: Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 6, no. 2, p. 91, Oct. 2021.
- [16] R. R. Muhima *et al.*, *Kupas Tuntas Algoritma Clustering: Konsep Perhitungan Manual dan Program Oleh: Rani Rotul Muhima*. Yogyakarta: Penerbit Andi, 2021.
- [17] D. A. Puspitasari, Y. Cahyana, and S. A. P. Lestari, “Penerapan Algoritma Density Based Spastial Clustering Algorithm With Noise Untuk Pengelompokkan Penyakit Pasien,” *Scientific Student Journal for Information, Technology and Science*, vol. 4, no. 1, pp. 102–106, Jan. 2023.
- [18] T. Rahmawati, Y. Wilandari, and P. Kartikasari, “Analisis Perbandingan Silhouette Coefficient dan Metode Elbow Pada Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator IPM dengan K-Medoids,” *Jurnal Gaussian*, vol. 13, no. 1, pp. 13–24, Aug. 2024, doi: <https://doi.org/10.14710/j.gauss.13.1.13-24>.
- [19] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, “Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *Jurnal Teknologi dan Informasi (JATI)*, vol. 12, no. 1, pp. 64–77, Mar. 2022, doi: <https://doi.org/10.34010/jati.v12i1.6674>.
- [20] H. Malikhatin, A. Rusgiyono, and D. A. I. Maruddani, “Penerapan k-Modes Clustering dengan Validasi Dunn Index Pada Pengelompokkan Karakteristik Calon TKI Menggunakan R-GUI,” *Jurnal Gaussian*, vol. 10, no. 3, pp. 359–366, 2021, doi: <https://doi.org/10.14710/j.gauss.10.3.359-366>.
- [21] A. Fitri *et al.*, *Dasar-Dasar Statistika Untuk Penelitian*. Medan: Yayasan Kita Menulis, 2023.
- [22] M. Marisyah, “Analisis Data Menggunakan Uji t: Menentukan Perbedaan yang Signifikan antara Dua Kelompok,” *OSF Preprints*, Jun. 2023, doi: <https://doi.org/10.31219/osf.io/t458y>.