MODEL SEGMENTASI KONSUMSI PANGAN BERBASIS *UNSUPERVISED LEARNING* UNTUK STRATEGI DASAR REKOMENDASI KETAHANAN PANGAN



Disusun Oleh:

N a m a : Syaharul Hafidz

NIM : 21523050

PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM INDONESIA

2025

HALAMAN PENGESAHAN DOSEN PEMBIMBING

MODEL SEGMENTASI KONSUMSI PANGAN BERBASIS *UNSUPERVISED LEARNING* UNTUK STRATEGI DASAR REKOMENDASI KETAHANAN PANGAN

TUGAS AKHIR

Disusun Oleh:

N a m a : Syaharul Hafidz

NIM : 21523050

Yogyakarta, Juli 2025

Pembimbing,

(Sri Mulyati, S.Kom., M.Kom.)

HALAMAN PENGESAHAN DOSEN PENGUJI

MODEL SEGMENTASI KONSUMSI PANGAN BERBASIS *UNSUPERVISED LEARNING* UNTUK STRATEGI DASAR REKOMENDASI KETAHANAN PANGAN

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 31 Juli 2025

Tim Penguji

Sri Mulyati, S.Kom., M.Kom.

Anggota 1

Erika Ramadhani, S.T., M.Eng.

Anggota 2

Nur Wijayaning Rahayu, S.Kom., M.Cs.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia

(Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama	:	Syaharul Hafidz
NIM	:	21523050

Tugas akhir dengan judul:

MODEL SEGMENTASI KONSUMSI PANGAN BERBASIS *UNSUPERVISED LEARNING* UNTUK STRATEGI DASAR REKOMENDASI KETAHANAN PANGAN

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apa pun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, Juli 2025



(Syaharul Hafidz)

HALAMAN PERSEMBAHAN

Pertama saya akan memanjatkan puji dan syukur kepada Allah SWT. yang telah memberikan rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan laporan tugas akhir ini. Kemudian shalawat serta salam saya limpahkan kepada Rasulullah SAW. yang telah membimbing ke jalan kebenaran dengan hadits dan tauladannya.

Tugas akhir ini akan saya persembahkan dengan rasa bangga dan hormat saya kepada kedua orang tua saya yang telah menuntun, mendukung, dan memberikan dukungan dari segi psikologis, moral, hingga finansial. Kemudian kepada adik saya dan seluruh saudara saya yang telah mendukung saya hingga dapat menyelesaikan laporan akhir ini. Saya berterimakasih juga kepada partner dan juga teman-teman saya yang sudah mendukung dalam pengerjaan laporan ini. Dan kepada seluruh responden yang sudah bersedia mengisi draft kuesioner yang saya sebarkan.

Saya juga mengucapkan terimakasih yang sebesar-besarnya kepada Ibu Sri mulyati, S.Kom., M.Kom., sebagai dosen pembimbing yang telah memberikan petunjuk, motivasi, dan ilmu yang bermanfaat. Kemudian saya juga berterimakasih kepada semua civitas akademik Universitas Islam Indonesia, mulai dari program studi hingga universitas karena telah memberikan saya kesempatan untuk menuntut ilmu di sini. Semoga semua kebaikan, ilmu, doa, harapan, dan motivasi yang diberikan oleh seluruh pihak yang terlibat dapat menjadi amal dan dibalas oleh Allah SWT. Aamiin ya rabbal alamin.

HALAMAN MOTO

“Jangan biarkan dunia yang mengubahmu. Kamulah yang seharusnya mengubah dunia“

KATA PENGANTAR

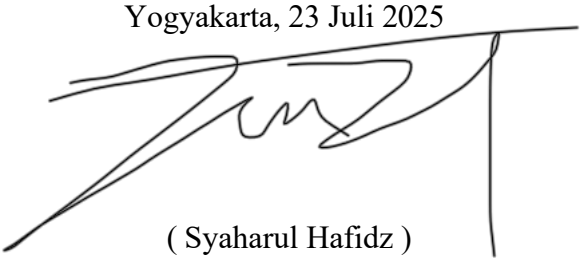
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Puji Syukur saya panjatkan kepada Allah SWT. yang telah memberikan rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan laporan tugas akhir ini. Kemudian shalawat serta salam saya limpahkan kepada Rasulullah SAW. yang telah membimbing ke jalan kebenaran dengan hadits dan tauladannya. Tugas akhir yang berjudul "MODEL SEGMENTASI KONSUMSI PANGAN BERBASIS *UNSUPERVISED LEARNING* UNTUK STRATEGI DASAR REKOMENDASI KETAHANAN PANGAN" dibuat sebagai pemenuhan syarat untuk memperoleh gelar sarjana di Program Studi Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia. Dalam penyusunan tugas akhir ini tidak terlepas dari dukungan, doa, serta motivasi yang diberikan kepada saya sebagai satu pondasi semangat yang diberikan oleh beberapa pihak terkait. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, karena telah memberikan karunia-Nya selama kegiatan Kuliah Kerja Nyata.
2. Rasulullah SAW, yang menjadi pedoman kita dalam menjalankan kehidupan sehari-hari dan menuntun kita ke jalan yang benar.
3. Kedua Orang Tua yang selalu memberikan kasih sayang dan doa kepada saya sehingga laporan akhir ini bisa selesai.
4. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku Ketua Jurusan Informatika Fakultas Teknologi Industri Universitas Islam Indonesia.
5. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku Kaprodi Informatika Program Sarjana Fakultas Teknologi Industri Universitas Islam Indonesia.
6. Ibu Sri Mulyati, S.Kom., M.Kom. selaku Dosen Pembimbing Tugas Akhir yang telah memberikan arahan dan motivasi selama penyelesaian tugas akhir.
7. Bapak Andhika Giri Persada, S.Kom., M.Eng., sebagai dosen pembimbing akademik yang telah memberikan informasi dalam hal akademik selama perkuliahan.
8. Seluruh Bapak dan Ibu dosen informatika yang telah memberikan ilmu yang bermanfaat kepada saya.
9. Pemilik NIM 21320041 dari Fakultas Psikologi Universitas Islam Indonesia yang telah memberikan dukungan secara moral maupun verbal.
10. Kepada teman-teman grup "P1 Mengguncang Dunia" yang saling mendukung satu sama lain.

11. Kepada sahabat-sahabat dan teman-teman saya yang telah mendukung saya.
12. Kepada seluruh pihak yang tidak bisa disebutkan satu persatu yang terlibat dan mendukung secara tidak langsung atas kelancaran pembuatan tugas akhir ini

Yogyakarta, 23 Juli 2025



(Syaharul Hafidz)

SARI

Pola konsumsi adalah sebuah kebiasaan untuk memilih jenis makanan yang akan dikonsumsi. Pola konsumsi di Indonesia masih sangat beragam, mulai dari memilih makanan yang sehat, lebih suka makanan praktis, hingga yang setia pada pangan local. Keragaman inilah yang akan dicari dalam penelitian ini agar mengetahui, bagaimana pola makan yang ada di tengah-tengah masyarakat saat ini. Untuk itu, penelitian ini membagi kelompok dengan menggunakan pendekatan statistik dan pengelompokkan data. Data yang akan digunaakan berasal dari kuesioner sebanyak 385 data, yang kemudian hanya 353 data yang berusia 17-45 Tahun akan diolah menggunakan *unsupervised learning* seperti *Hierarchicaal clustering*, *K-Means Clustering*, dan *Gaussian Mixture Model* (GMM). Hasilnya, penelitian ini dapat memetakan Masyarakat masuk ke dalam beberapa kelompok utama yang mempunyai ciri khas konsumsi masing-masing. Hasil menunjukkan Gaussian Mixture Model menghasilkan segmentasi dengan skor siluet sebesar 0,2801 diikuti K-Means dengan 0,3513 dan Hierarchical Clustering dengan 0,3398. Temuan ini diharapkan bisa menjadi bahan pengetahuan untuk mengenali pola konsumsi yang ada di Masyarakat dan para pemangku kepentingan bisa membuat kebijakan sesuai dengan pola konsumsi Masyarakat agar ketahanan pangan nasional bisa berjalan efektif dan berkelanjutan.

Kata kunci: Segmentasi konsumsi; Pola makan; *Unsupervised learning*; *Hierarchical clustering*, *K-Means clustering*, *Gaussian Mixture Model*

GLOSARIUM

Clustering	Proses pengelompokan data menjadi beberapa grup atau klister dengan berdasarkan dari kemiripan karakteristik tertentu.
Confirmatory Factor Analysis (CFA)	Metode statistik untuk melakukan pengujian kesesuaian antara model teori dan data empiris, terutama untuk validasi atribut..
Cronbach's Alpha	Satuan ukuran statistic yang berguna untuk menilai konsistensi internal atau reliabilitas dari suatu instrumen atau kuesioner.
Dendogram	Diagram bentuk pohon yang yang menggambarkan proses penggabungan klaster secara hierarki pada metode *hierarchical clustering.*
Gaussian Mixture Model (GMM)	Model statistik untuk mengelompokkan data yang memperbolehkan satu data dapat memiliki kemungkinan ada di lebih dari satu klaster (*soft clustering*).
Hierarchical Clustering	Salah satu metode pengelompokan data secara bertingkat yang membentuk struktur berjenjang dengan berdasarkan kepada kemiripan antar data.
K-Means Clustering	Algoritma pengelompokan data ke dalam jumlah klister tertentu dengan berdasar kepada kedekatan data ke pusat klaster.
Min-Max Scaling	Salah satu teknik normalisasi data yang mengubah skala nilai data ke rentang tertentu, biasanya antara nilai 0 dan 1.
Silhouette Score	Nilai atau skor yang digunakan untuk mengukur hasil evaluasi dari algoritma, seberapa baik hasil klasterisasi dalam memisahkan data ke dalam klaster yang berbeda.
Soft Clustering	Pendekatan pengelompokan di mana satu dat bisa memiliki probabilitas keanggotaan di beberapa klaster, tidak hanya di satu klaster saja.
Unsupervised Learning	Metode pembelajaran pada mesin di mana model dilatih tanpa label/kategori yang bertujuan untuk menemukan pola atau struktur dalam data.

DAFTAR ISI

HALAMAN PENGESAHAN DOSEN PEMBIMBING i
HALAMAN PENGESAHAN DOSEN PENGUJI ii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR iii
HALAMAN PERSEMBAHAN iv
HALAMAN MOTO v
KATA PENGANTAR vi
SARI viii
GLOSARIUM ix
DAFTAR ISI x
DAFTAR TABEL xii
DAFTAR GAMBAR xiii
BAB I PENDAHULUAN 1
1.1. Latar Belakang 1
1.2. Rumusan Masalah 4
1.3. Tujuan 4
1.4. Manfaat 4
1.5. Batasan Masalah 5
1.6. Sistematika Penulisan 5
BAB II KAJIAN PUSTAKA 8
2.1. Kajian Pustaka 8
2.2. Segmentasi Pola Konsumsi 34
2.3. Ketahanan Pangan 34
2.4. *Hierarchical Clustering* 35
2.5. *K-Means Clustering* 37
2.6. *Gaussian Mixture Model* 38
2.7. Uji Reliabilitas, Validitas, dan *Confirmatory Factor Analysis* 40
BAB III METODOLOGI PENELITIAN 43
3.1. *Hierarchical Clustering* 44
3.2. *K-means Clustering* 46
3.3. *Gaussian Mixture Model (GMM)* 48
3.4. Desain Penelitian 49
3.4.1. Pendekatan Penelitian 49
3.4.2. Metode Pengumpulan Data 50

3.5. Indikator dan Variabel Penelitian 50
BAB IV HASIL DAN PEMBAHASAN 56
4.1. Deskripsi Data Awal 56
4.2. *Pre-Processing* Data 57
4.2.1. Pembersihan Tahap Awal 57
4.2.2. Pembersihan dan Pemfilteran Data 57
4.2.3. Pra-Pemrosesan Tahap Akhir 59
4.3. Pemilihan Attribut 60
4.4. Penentuan Jumlah Klaster Optimal 64
4.5. *Hierarchical Clusteirng* 70
4.6. *K-Means Clustering* 73
4.7. *Gaussian Mixture Model* (GMM) 79
4.8. Interpretasi Klaster 83
BAB V KESIMPULAN DAN SARAN 92
5.1. Kesimpulan 92
5.2. Saran 93
Daftar Pustaka 94
Lampiran 98

DAFTAR TABEL

Tabel 2. 1. Tabel Kajian Pustaka.....9
Tabel 2. 2. Tabel Kajian Pustaka Metode 25
Tabel 3. 1. Tabel Indikator dan Pertanyaan Kuesioner 50
Tabel 4. 1. Tabel Uji Reliabilitas dan Validitas 60
Tabel 4. 2. Tabel Hasil *Confirmatory Factor Analysis* (CFA).....62
Tabel 4. 3. Tabel Uji Relabilitas.....63
Tabel 4. 4. Tabel Perbandingan Silhouette Score 66
Tabel 4. 5. Tabel Perbandingan Silhouette Score setelah pemilihan atribut 69
Tabel 4. 6. Tabel Karakteristik Klaster GMM 84
Tabel 4. 7. Tabel Profil Klaster 85
Tabel 4. 8. Tabel Hasil Ketiga Metode 89

DAFTAR GAMBAR

Gambar 2. 1. Badan Pusat Statistik Nasional (https://www.bps.go.id/id/statistics-table/1/MTA0MyMx/impor-beras-menurut-negara-asal-utama-2017-2023.html) 8
Gambar 2. 2. Alur Proses Algoritma *Hierarchical Clustering* 36
Gambar 2. 3. *Flowchart* step penggunaan algoritma *k-means* 38
Gambar 2. 4. *Flowchart* Algortima *Gaussian Mixture Model* 40
Gambar 3. 1. Alur Diagram Penyelesaian Masalah 43
Gambar 4. 1. *Code* Program Menampilkan Data Awal 57
Gambar 4. 2. *Code* Pembersihan dan Filter Data 58
Gambar 4. 3. *Output Code* Pembersihan dan Filter Data 59
Gambar 4. 4. Code *MinMax Scaler* 60
Gambar 4. 5. *Code* Pra-Pemrosessan Akhir 60
Gambar 4. 6. *Code* untuk mencari K optimal 65
Gambar 4. 7. Nilai K Optimal Menggunakan *Silhouette Score* 66
Gambar 4. 8. *Code* untuk mendefinisikan atribut 68
Gambar 4. 9. Grafik nilai k optimal setelah pemilihan atribut 69
Gambar 4. 10. Implementasi *Hierarchical Clustering* 71
Gambar 4. 11. Dendogram *Hasil Hierarchical Clustering* 72
Gambar 4. 12. Dendogram Hasil *Hierarchical Clustering* Setelah Pemilihan Atribut 73
Gambar 4. 13. Algoritma *K-Means Clustering* 74
Gambar 4. 14. Visualisasi Hasil *K-Means Clustering* 75
Gambar 4. 15. Code untuk Analisis Klaster *K-Means* 76
Gambar 4. 16. Visualisasi Hasil *K-Means Clustering* dengan Pemilihan Atribut 78
Gambar 4. 17. Code Untuk Menjalankan Algoritma GMM 80
Gambar 4. 18. Code Menghitung Probabilitas GMM 81
Gambar 4. 19. Visualisasi PCA 2D GMM Clustering 82
Gambar 4. 20. Visualisasi PCA 2D GMM Clustering Setelah Pemilihan Atribut 83
Gambar 4. 21. Diagram Gambaran Penerapan Ketiga Metode 91

BAB I
PENDAHULUAN

1.1. Latar Belakang

Pangan merupakan salah satu aspek penting yang menunjang kehidupan manusia. Pangan mencakup bahan olahan yang akan dijadikan konsumsi untuk semua orang. Pangan sendiri menjadi kebutuhan primer saat ini, terlebih lagi dalam beberapa tahun ke depan pangan dapat menjadi sesuatu yang menjadi sangat penting dan akan terjadi kelangkaan jika ketahanan pangan tidak dijaga oleh suatu negara. Ketahanan pangan sendiri adalah sebuah konsep yang mencakup kemampuan suatu negara, komunitas, atau individu untuk menyediakan pangan yang cukup, bernutrisi, dan aman bagi seluruh masyarakat yang ada di suatu negara, baik pada masa kini maupun masa depan.

Ketahanan pangan di Indonesia saat ini adalah aspek yang sangat krusial dalam hal pembangunan nasional. Aspek utama dari krusialnya ketahanan pangan di Indonesia terdapat pada besarnya jumlah penduduk dan terus meningkat, kemudian keberagaman kondisi geografis serta sosial ekonomi di berbagai wilayah Indonesia juga menjadi faktor keduanya. Indonesia termasuk ke dalam negara agraris, tetapi masih ketergantungan dalam importir bahan pangan, seperti gandum dan kedelai. Kegiatan impor bahan pangan menunjukkan bahwa ketersediaan bahan pangan domestik masih menjadi tantangan, ditambah dengan distribusi pangan di wilayah Indonesia masih terjadi ketimpangan, terutama di luar pulau Jawa yang ikut memperburuk situasi. Penelitian sesuai dengan program pemerintah saat ini yaitu, program nawacita tentang pangan.

Masyarakat di Indonesia masih cenderung bergantung pada beras yang menjadi sumber pangan utama, meskipun banyak sumber daya pangan tradisional seperti umbi-umbian dan sagu yang belum dimanfaatkan optimal. Faktor eksternal yang mempengaruhi ketahanan pangan adalah perbuahan iklim, di mana perubahan iklim mengancam stabilitas ketahanan pangan dengan mengganggu produksi pangan di berbagai wilayah. Kemudian bencana alam seperti banjir dan kekeringan juga merusak ekosistem pangan berkelanjutan yang mengakibatkan bahan pangan tidak tersedia atau mempengaruhi hasil pertanian. Faktor eksternal yang lainnya adalah pandemi, seperti pandemi *COVID-19* yang menambah buruk situasi dengan ketidakpastian pangan sehingga pasokan rantai pangan terganggu dan terjadi penurunan daya beli masyarakat.

Kebijakan pangan yang ada di Indonesia saat ini masih bersifat secara umum dan belum memanfaatkan teknologi sains data untuk membuat segmentasi pola makan, sehingga

menyebabkan strategi yang diterapkan sering kali tidak sesuai dengan kebutuhan konsumsi yang spesifik dari setiap kelompok masyarakat. Pengaruhnya adalah tidak semua kelompok masyarakat mendapatkan perhatian yang sama dalam strategi pangan. Letak geografis juga bisa menjadi salah satu tantangan dalam membuat kebijakan, karena setiap daerah mempunyai ciri khas pola makannya masing-masing dan kebijakan tidak bisa dibuat secara umum saja. Kemudian, ketersediaan pangan saja tidak cukup untuk menunjang ketahanan pangan, tetapi juga diperlukan pendekatan yang edukatif dengan menyesuaikan segmen pola makannya.

Pola konsumsi masyarakat bisa dipengaruhi oleh berbagai faktor, seperti kondisi ekonomi, kesadaran kesehatan, serta aksesibilitas terhadap pangan lokal. Konsumen dengan kesadaran kesehatan yang tinggi cenderung memilih makanan dengan kadar gula dan lemak rendah, sementara kelompok lain masih mengandalkan makanan olahan dengan nilai gizi yang kurang optimal. Preferensi terhadap pangan lokal juga bervariasi, bergantung pada ketersediaan dan kebiasaan konsumsi di suatu wilayah. Segmentasi pola konsumsi dapat membantu mengidentifikasi kelompok masyarakat berdasarkan karakteristik konsumsi mereka, sehingga kebijakan pangan dapat disesuaikan dengan kebutuhan dan preferensi tiap segmen. Faktor pola konsumsi juga dipengaruhi berdasarkan perilaku dan konsumen tentang ketahanan pangan.

Menurut O'Neill et al.(2023) perilaku konsumsi yang mempengaruhi ketahanan pangan terbagi dalam empat segmen berbeda yaitu *food dabblers* yang merupakan kelompok dengan pemikiran dan praktik keberlanjutan terbatas dan belum sepenuhnya mengintegrasikan aspek berkelanjutan dalam pola konsumsi. Yang kedua adalah *Food Appreciators* yang merupakan konsumen dengan menunjukkan ketertarikan pada produk pangan lokal, organik, atau bebas daging, tetapi belum konsisten dalam praktiknya. Yang ketiga adalah *Pro-Sus (Pro-Sustainability)* yang merupakan konsumen aktif dalam menerapkan perilaku berkelanjutan (meminimalkan limbah, memilih produk ramah lingkungan, dan kelola stok makanan secara efisien). Yang terakhir adalah *Pressured* yang merupakan konsumen yang terbatas dalam menerapkan praktik berkelanjutan karena faktor eksternal (waktu, biaya, dan akses).

Konsumsi pangan yang berkelanjutan saat sini semakin penting dan perlu mendapatkan perhatian utama dalam suatu negara, karena meliputi aspek keseimbangan antara kebutuhan manusia dan pelestarian lingkungan. Menurut Schäufele-Elbers dan Janssen (2023), terdapat tiga dimensi konsumsi pangan berkelanjutan yang dapat di segmentasi menjadi pola konsumsi, yaitu konsumsi daging, makanan organik, dan makanan manis yang diambil berdasarkan data panel rumah tangga di Jerman. Hal ini memberikan gambaran betapa pentingnya ketahanan pangan berkelanjutan.

Bogusz et al. (2023) melakukan kajian konsumsi berkelanjutan dengan pendekatan *zero waste*, di mana penelitian ini berfokus pada pengurangan limbah rumah tangga dan optimalisasi efisiensi sumber daya. Satu dimensi berkelanjutan yang perlu diperhatikan salah satunya ramah lingkungan (Schäufele-Elbers dan Janssen, 2023). Dalam hal ini memberikan petunjuk bahwa perubahan kecil pada pola konsumsi bisa memberikan kontribusi besar terhadap berkelanjutan pangan dan lingkungan. Penelitian yang dilakukan di kota Izmir, Turki menunjukkan bahwa preferensi dan kebiasaan konsumsi individu, seperti memilih produk lokal dan organik secara signifikan dapat mempengaruhi ketahanan pangan (Biresselioglu et al., 2023). Semakin tinggi kesadaran masyarakat terhadap isu keberlanjutan, maka akan semakin besar kecenderungan individu untuk berpartisipasi dalam upaya mendukung sistem pangan yang lebih ramah lingkungan dan berkelanjutan.

Pendekatan baru diperkenalkan oleh Burchi & De Muro (2016), yaitu analisis ketahanan pangan tidak hanya berfokus pada ketersediaan pangan, tetapi juga pada kemampuan masyarakat dalam mengakses dan memanfaatkan gizi yang terkandung dalam makanan. Pendekatan ini menekankan bahwa betapa pentingnya melihat ketahanan pangan secara holistik, dengan melihat dan mempertimbangkan beberapa faktor, seperti ekonomi, kesehatan, dan sosial yang mempengaruhi kemampuan individu untuk memperoleh makanan bergizi. Pendekatan ini memberikan aspek pendekatan lain yang memperluas pemahaman tentang bagaimana ketahanan pangan tidak hanya berkaitan dengan produksi pangan, melainkan juga perlu diperhatikan kesejahteraan masyarakat secara keseluruhan, termasuk dalam aspek kesehatan dan nutrisi. Kemudian pendekatan holistik juga dijelaskan oleh Abdella et al.(2020) di mana sebagai implikasi untuk ketahanan pangan yang mengintegrasikan aspek lingkungan, ekonomi, dan sosial dalam kebijakan pangan untuk mencapai sistem pangan berkelanjutan.

Untuk mendukung ketahanan pangan diperlukan transformasi sistem pangan ke yang lebih berkelanjutan. Menurut Ruben et al.(2021) tantangan global dalam sistem pangan harus diubah tetapi dengan mempertimbangkan kesehatan, nutrisi, keberlanjutan, lingkungan, mata pencaharian inklusif, dan ketahanan sistem pangan melalui lima pergeseran paradigma utama, yaitu:

1. Meningkatkan ambisi kebijakan. Kebijakan pangan perlu lebih progresif untuk menjawab tantangan multidimensi (kesehatan, ekonomi, dan lingkungan).
2. Selaraskan tujuan produksi dan konsumsi dengan cara mengurangi *trade-off* antara produksi pangan dan pola konsumsi berkelanjutan.
3. Memperkuat konektivitas sistem dengan cara meningkatkan koherensi antara rantai pasok, distribusi, dan perilaku konsumen untuk efisiensi sistem.

4. Optimalisasi kinerja sistem pangan dengan berfokus pada peningkatan produktivitas.
5. Tata kelola inklusif dan partisipatif dengan melibatkan pemangku kepentingan (petani, konsumen, industri) dalam pengambilan keputusan yang adil.

Penelitian sebelumnya sebagian besar membahas tentang perilaku konsumsi, pendekatan berkelanjutan, faktor yang mempengaruhi, preferensi konsumsi dan lainnya, namun tidak membahas secara khusus terkait segmentasi berdasarkan preferensi pangan sehat dan ketahanan pangan di negara Indonesia. Dari penelitian tersebut dapat dilihat keterbaruan yang diusulkan, yaitu:

1. Penelitian menggunakan *Hierarchical* Clustering, K-*Means Clustering* dan *Gaussian Mixture Model* untuk membangun model segmentasi pola konsumsi berdasarkan kombinasi variabel kondisi ekonomi, kesadaran kesehatan, serta aksesibilitas terhadap pangan lokal.
2. Studi ini tidak hanya membahas tentang pola konsumsi, tetapi juga memberikan interpretasi hasil dan rekomendasi kebijakan berbasis data bagi pemerintah untuk meningkatkan ketahanan pangan.

1.2. Rumusan Masalah

1. Bagaimana membangun model segmentasi pola konsumsi masyarakat menggunakan metode *hierarchical clustering, k-means clustering, dan gaussian mixture model* (GMM), guna mengelompokkan masyarakat berdasarkan pola konsumsi pangan mereka?
2. Bagaimana kualitas model dapat saling melengkapi untuk hasil segmentasi pola konsumsi masyarakat?

1.3. Tujuan

Tujuan dari penelitian ini adalah :

1. Membangun model segmentasi pola konsumsi masyarakat menggunakan metode *machine learning*, yaitu *hierarchical clustering*, *k-means clustering*, dan *gaussian mixture model* (GMM) untuk mengelompokkan masyarakat berdasarkan pola konsumsi pangan mereka.
2. Mengevaluasi dan menginterpretasikan hasil segmentasi yang diperoleh dari hasil klastering ketiga metode tersebut.

1.4. Manfaat

Secara umum, ada dua kontribusi utama dalam penelitian ini, yaitu:

1. Mengembangkan model segmentasi pola konsumsi masyarakat berbasis data.
2. Memberikan evaluasi dan interpretasi dari setiap klaster yang terbentuk untuk mengetahui pola konsumsi masyarakat

1.5. Batasan Masalah

Dalam penelitian ini tidak sepenuhnya mencakup semua daerah dan keterbatasan lainnya. Oleh karena itu perlu adanya batasan masalah dari penelitian ini yaitu adalah sebagai berikut:

1. Jangkauan Wilayah Terbatas

   Penelitian ini difokuskan pada wilayah tertentu dengan karakteristik demografis tertentu, sehingga hasilnya tidak merepresentasikan pola konsumsi masyarakat seluruh Indonesia.

2. Keterbatasan pada Responden

   Keterbatasan responden dalam segi usia yang hanya memiliki target rentang usia dari 17 hingga 45 tahun. Hal ini karena pada usia anak-anak dan lansia belum mencerminkan preferensi konsumsi.

3. Data Kuesioner yang subjektif

   Data yang diperoleh dari kuesioner bergantung pada persepsi responden, yang mungkin terpengaruh oleh interpretasi pribadi atau sosial, sehingga terdapat kemungkinan bias.

1.6. Sistematika Penulisan

Penulisan tugas akhir ini disusun dengan sistematika yang terstruktur agar memudahkan dalam pemahaman dan analisis terhadap topik yang dibahas. Berikut adalah penjelasan mengenai sistematika penulisan tugas akhir ini:

BAB I Pendahuluan

Pada bab ini hal yang dibahas adalah latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Latar belakang menjelaskan secara umum tentang pentingnya ketahanan pangan berkelanjutan dan peran dari segmentasi pola konsumsi yang ada di masyarakat dan pendukungnya. Rumusan masalah dalam bab 1 mengidentifikasi pertanyaan penelitian yang akan dijawab dalam penelitian ini, sedangkan tujuan dan manfaat penelitian menjelaskan hasil yang diharapkan dari penelitian ini dan kontribusinya. Batasan masalah menjelaskan keterbatasan yang ada dalam penelitian, dan sistematika penulisan memberikan gambaran struktur penulisan tugas akhir.

BAB II Kajian Pustaka dan Teori

Pada bab 2 ini berisi kajian pustaka atau penelitian terdahulu yang memberikan landasan teoritis dan relevan dengan topik tugas akhir ini. Bab ini juga menjelaskan tentang teori yang digunakan dalam penelitian, seperti segmentasi pada pola konsumsi, ketahanan pangan, dan metode yang akan digunakan. Kajian pustaka dan teori ini menjadi dasar untuk memahami konteks penelitian dan mengembangkan metodologi yang akan digunakan.

BAB III Metodologi Penelitian

Pada bab 3 ini dijelaskan metodologi yang digunakan dalam penelitian, termasuk pendekatan penelitian, desain penelitian, metode pengumpulan data, serta indikator dan variabel yang digunakan dalam pengambilan data. Metodologi yang dijelaskan dalam bab 3 ini adalah *Hierarchical* Clustering, K-*Means Clustering* dan *Gaussian Mixture Model* yang digunakan dalam proses pengolahan data di dalam penelitian ini untuk mengelompokkan data berdasarkan pola konsumsi masyarakat. Pada bab ini juga dijelaskan tahapan dalam pengolahan data, normalisasi, hingga interpretasi hasil klaster.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian yang diperoleh dari analisis data menggunakan metode *Hierarchical* Clustering, K-*Means Clustering* dan *Gaussian Mixture Model*. Hasil penelitian mencakup pengelompokan pola konsumsi masyarakat berdasarkan variabel-variabel yang telah ditentukan. Selain itu, bab ini juga membahas implikasi hasil penelitian terhadap ketahanan pangan berkelanjutan, serta memberikan interpretasi terhadap temuan-temuan yang diperoleh.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan, serta saran dan rekomendasi untuk penelitian selanjutnya. Kesimpulan merangkum temuan-temuan penting dari penelitian, sedangkan saran dan rekomendasi memberikan arahan untuk pengembangan kebijakan dan penelitian lanjutan terkait ketahanan pangan berkelanjutan.

DAFTAR PUSTAKA

Bagian ini berisi daftar referensi yang digunakan dalam penelitian, baik berupa buku, jurnal, artikel, maupun sumber-sumber lain yang relevan. Daftar pustaka disusun sesuai dengan standar penulisan akademik yang berlaku.

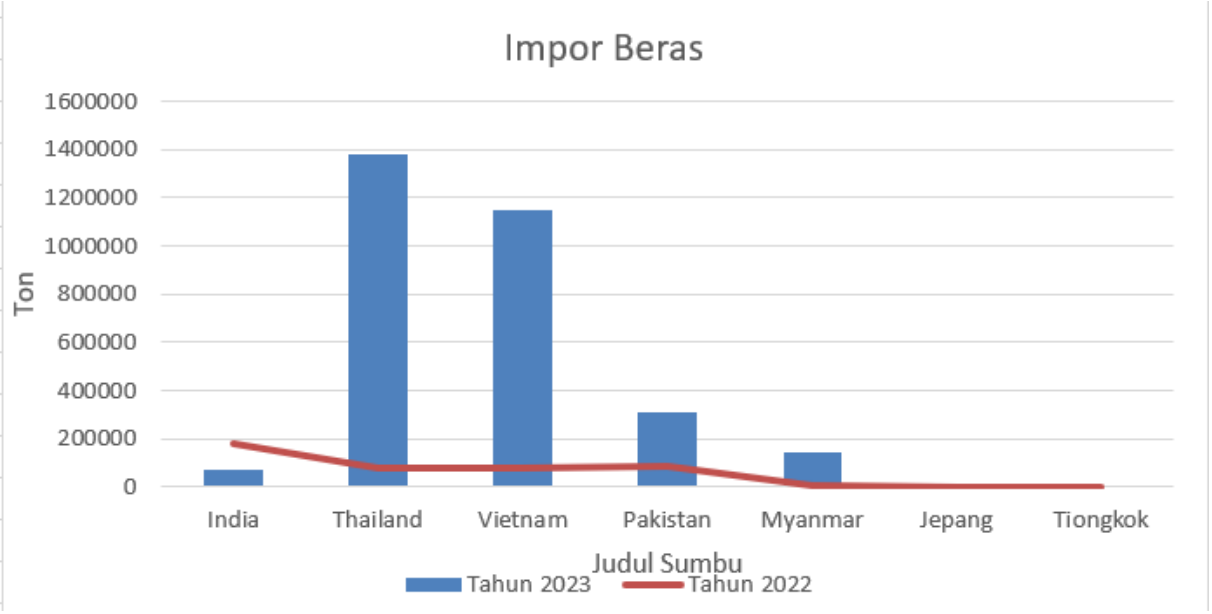
LAMPIRAN

Lampiran berisi dokumen-dokumen pendukung yang digunakan dalam penelitian, seperti kuesioner, data mentah, atau hasil analisis tambahan yang tidak dimasukkan dalam bab-bab utama.

BAB II
KAJIAN PUSTAKA

2.1. Kajian Pustaka

Ketahanan pangan merupakan salah satu aspek penting di dalam suatu negara, dimana ketahanan pangan merupakan salah satu upaya sebuah negara dalam menjaga perekonomian dalam konteks konsumsi masyarakat. Ketahanan pangan bisa dikenali mengenali pola konsumsi masyarakat di dalam suatu negara. Keadaan saat ini di Indonesia sangat memprihatinkan, masalahnya berada di ketersediaan pangan di Indonesia. Indonesia yang merupakan negara agraris, tetapi masih bergantung pada produk impor, khususnya pada bahan pokok seperti beras. Dalam dua tahun ke belakang 2022 sampai 2023 ditemukan bahwa konsumsi masyarakat Indonesia masih sangat tinggi, sehingga diperlukan impor bahan pokok yaitu beras.



Gambar 2. 1. Badan Pusat Statistik Nasional (https://www.bps.go.id/id/statistics-table/1/MTA0MyMx/impor-beras-menurut-negara-asal-utama-2017-2023.html)

Dilihat dari pusat Badan Pusat Statistik impor makanan pokok makanan pokok, yaitu beras masih sangat tinggi. Hal inilah yang harus dibenahi dari pola konsumsi masyarakat dengan mengenali pola konsumsi dari beberapa segmentasi untuk ditindaklanjuti lebih lanjut. Kemudian ada beberapa masalah tertentu yang mempengaruhi impor bahan pokok di Indonesia, antara lain yaitu :

1. Ketergantungan pada pasokan pangan luar negeri: Impor beras yang tinggi dapat membuat suatu negara lebih bergantung pada pasokan pangan negara lain, membuatnya lebih rentan terhadap kebijakan ekspor negara pemasok dan fluktuasi harga internasional.
2. Dampak terhadap petani lokal: Harga beras lokal sering turun ketika beras impor masuk ke pasar domestik, yang mengurangi pendapatan petani dan dapat menghambat keberlanjutan sektor pertanian lokal. Jika nilai ekspor tidak dapat menutupi biaya impor, pengeluaran impor yang besar dapat meningkatkan defisit neraca perdagangan.
3. Kurangnya daya saing produksi lokal: Ketergantungan pada impor beras sering dikaitkan dengan masalah infrastruktur, teknologi, atau kebijakan pertanian yang tidak efisien.
4. Ancaman terhadap ketahanan pangan nasional: Bergantung pada impor mengurangi kemandirian pangan dan meningkatkan risiko jika terjadi gangguan pada rantai pasokan global seperti perang, bencana alam, atau pandemi.
5. Transportasi impor beras berdampak pada lingkungan karena mengambil waktu yang lama, yang menyebabkan emisi gas rumah kaca dan efek negatif lainnya. Konsumen kadang-kadang memilih beras lokal karena perubahan pola konsumsi, yang mengakibatkan penurunan permintaan terhadap produk dalam negeri.
6. Subsidi yang tidak tepat sasaran: Ada saat-saat ketika kebijakan atau subsidi impor beras dimaksudkan untuk mengatasi kekurangan pangan, tetapi mereka malah mengganggu pasar dan membahayakan produsen kecil.

Untuk mengatasi hal di atas diperlukan adanya segmentasi pola konsumsi masyarakat untuk mengetahui pola konsumsi yang ada di Indonesia dan faktor yang mempengaruhi hal tersebut untuk meningkatkan ketahanan pangan.

Tabel 2. 1. Tabel Kajian Pustaka

No	Tema/Topik	Penulis	Judul Penelitian	Ringkasan Temuan	Relevansi
1.	Faktor Sosio-demografis dalam	(Almulhim & Abubakar, 2024)	A segmentation approach to understanding water	- Konsumsi air di Saudi Arabia sangat tinggi dihitung dari per kapita, karena	- Membantu memahami variabel-variabel penting dalam

	Pola Konsumsi		consumption behavioral patterns among households in Saudi Arabia for a sustainable future	ketergantungan pada desalinasi air laut yang intensif energi. - Mengidentifikasi enam segmen konsumen air berdasarkan perilaku dan faktor sosio-demografis mereka. Faktor-faktor yang mempengaruhi seperti pendidikan, ukuran rumah tangga, pendapatan, usia, dan kewarganegaraan yang berpengaruh pada pola konsumsi.	segmentasi konsumsi pangan. - konsep segmentasi berdasarkan perilaku dan faktor sosio-demografis yang sangat relevan.
2.	Faktor Sosio-demografis dalam Pola Konsumsi	(Biresseliog lu et al., 2023)	How to Exploit Sustainable Food Consumption Habits of Individuals: Evidence from	- Faktor sosio-demografik seperti usia, gender, pendidikan, pendapatan, dan jumlah anggota	Membantu memahami variabel penting dalam segmentasi konsumsi pangan.

			a Household Survey in Izmir, Türkiye	rumah tangga mempengaruhi perilaku konsumsi makanan berkelanjutan.	
3.	Nilai, Sikap & Motivasi Konsumen	(Piracci et al., 2023)	Identifying key attributes in sustainable food choices: An analysis using the food values framework	Tiga segmen konsumen berdasarkan nilai pangan yang konsumen anggap penting dalam memilah produk yang berkelanjutan: - *Private Benefit Seekers* (35%): Fokus pada kesehatan dan harga. - *Sustainability Focused* (32.3%): Menempatkan keberlanjutan sebagai prioritas utama. *Naturalness and Health Driven*	Relevansi dalam pembahasan yang sama, yaitu segmentasi pola konsumsi makanan dengan mengidentifikasi nilai yang dapat mempengaruhi pilihan makanan.

				(32.7%): Memprioritaskan kealamian dan manfaat kesehatan.	
4.	Nilai, Sikap & Motivasi Konsumen	(Gazdecki et al., 2021)	8 of Food Consumers Based on Their Sustainable Attitude	Temuan dari studi ini mengidentifikasi tiga segmen konsumen yang didasarkan pada sikap berkelanjutan: - *Doers*: memiliki perilaku berkelanjutan tetapi memiliki pemahaman yang terbatas - *Conscious*: memiliki pengetahuan, tetapi kurang dalam hal tindakan. - *Reluctant*: tidak menunjukkan sikap maupun dimensi lain	Studi ini menawarkan pendekatan dalam segmentasi berbasis sikap konsumen terhadap keberlanjutan.

				dalam hal berkelanjutan.	
5.	Nilai, Sikap & Motivasi Konsumen	(Maciejewski & Lesznik, 2022)	Consumers Towards the Goals of Sustainable Development: Attitudes and Typology	- Identifikasi tiga tipe konsumen berdasarkan hasil survei kepada 1.045 konsumen dewasa, yang terbagi dalam tiga segmen: - *Informed Apologists*: mengetahui sekaligus mendukung SDGs - *Moderate Supporters*:mendukung SDGs, tetapi memiliki pengetahuan yang minim. - *Uninformed Critics*: pengetahuan yang kurang disertai dengan dukungan yang kurang	Studi ini menawarkan pendekatan dalam segmentasi berbasis sikap konsumen terhadap keberlanjutan.

				terhadap SDGs	
6.	Lingkungan Sosial & Tempat Tinggal	(Piligrimiene et al., 2021)	Sustainable consumption patterns in different settings	Metode analisis dengan menggunakan *two-step cluster analysis* dengan mengidentifikasi empat kluster : - Di Rumah: - Convenience-Based Sustainability: Pola konsumsi yang mengutamakan kenyamanan. - True Followers: Konsumen yang berperilaku secara konsisten berkelanjutan. - Di Tempat Kerja: - Environment-Independent Sustainability: Pola konsumsi berkelanjutan	Membahas tentang segmentasi pola konsumsi masyarakat dengan memberikan wawasan perilaku konsumen di lingkungan yang berbeda.

				yang tidak tergantung pada lingkungan kerja. - Followers of Basic Principles: Konsumen yang mengikuti prinsip dasar keberlanjutan.	
7.	Perubahan Iklim & Adaptasi Konsumsi	(Ashari, 2023)	Dampak Perubahan Iklim terhadap Ketahanan Pangan dan Adaptasinya oleh Masyarakat Pesisir	Adaptasi masyarakat dalam menghadapi perubahan iklim relevan untuk ketahanan pangan	Menambah aspek lokal dan urgensi adaptasi pola konsumsi
8.	Dimensi & Preferensi Keberlanjutan	(Schäufele-Elbers & Janssen, 2023)	Consumer segmentation based on three dimensions of sustainable food consumption: a simultaneous analysis of meat, organic	Segmentasi berdasarkan dimensi konsumsi berkelanjutan.	Membantu mengidentifikasi pola konsumsi masyarakat.

			food, and sweet snack purchases based on household panel data in Germany		
9.	Dimensi & Preferensi Keberlanjutan	(O'Neill et al., 2023)	Food interests, preferences and behaviours : a profile of the sustainable food consumer	Diidentifikasikan dalam empat segmen perilaku konsumsi: - *Food Dabblers*: memiliki minat rendah terhadap keberlanjutan, namun sesekali mempraktikan perilaku keberlanjutan. - *Food Appreciators*: memiliki minat sedang terhadap keberlanjutan dan menghargai kualitas pangan lokal	Penelitian ini memiliki relevansi dengan menggambarkan segmentasi konsumen berdasarkan preferensi, minat, dan perilaku yang dapat memahami, mendorong, menjadi dasar segmentasi untuk mendukung ketahanan pangan berkelanjutan.

				dan juga organik. - *Pro-Sus*: konsumen yang tertarik dan berperan aktif dalam perilaku keberlanjutan - *Pressured*: konsumen yang terbatas oleh waktu dan sumber daya, padahal ingin berperilaku keberlanjutan.	
10.	Pentingnya Informasi & Edukasi	(Ran et al., 2022)	Information as an enabler of sustainable food choices: A behavioural approach to understanding consumer decision-making	- Konsumen memiliki niat untuk membeli pangan secara berkelanjutan, tetapi pembelian dipengaruhi oleh faktor-faktor yang sangat kompleks. - Keputusan pembelian makanan bisa dipengaruhi	Membantu dalam pemahaman bagaimana informasi dapat digunakan untuk mempengaruhi perubahan perilaku konsumen ke arah berkelanjutan.

				oleh informasi, tetapi mekanisme perubahannya belum dipahami. - Menggunakan model COM-B (Capability, Opportunity, Motivation, and Behavior) yang dipengaruhi faktor kunci untuk berbelanja makanan adalah kualitas, kesehatan, produksi lokal, kesejahteraan hewan dan kenyamanan. - Teknik perubahan perilaku didapatkan dari informasi yang efektif disesuaikan dengan seluruh belanja konsumen.	
11.	Peran Kebijakan &	(Mensah et al., 2024)	Sustainable food	- Konsumsi pangan	Membahas aspek berkelanjutan dari

	Indikator Ketahanan Pangan; Tantangan dan Solusi dalam Konsumsi Berkelanjutan		consumption and Sustainable Development Goal 12: Conceptual challenges for monitoring and implementation	berkelanjutan sering kali terabaikan oleh banyak kebijakan nasional yang berfokus pada produksi pangan berkelanjutan. - SDG 12 tidak cukup untuk menilai kemajuan menuju konsumsi pangan berkelanjutan karena lebih berfokus pada aspek ekonomi dan lingkungan. - Perlu ada indikator kebijakan yang lebih komprehensif untuk mencakup semua elemen konsumsi pangan berkelanjutan. Definisi yang jelas tentang	konsumsi pangan dengan mengidentifikasi tantangan dalam mengukur dan mendorong konsumsi makanan yang berkelanjutan, yang dapat membantu memberikan informasi tentang upaya untuk menyebarluaskan pola konsumsi masyarakat ke bagian-bagian tertentu.

				konsumsi pangan berkelanjutan sangat penting.	
12.	Peran Kebijakan & Indikator Ketahanan Pangan	(Drewnowski et al., 2020)	Toward healthy diets from sustainable food systems	- Terdapat empat dimensi utama keberlanjutan dalam sistem pangan yaitu, kesehatan, ekonomi, lingkungan, dan sosial. - Perlu adanya strategi yang diperlukan untuk mendorong konsumen memilih diet sehat dari sistem pangan berkelanjutan	Penelitian ini mencakup faktor kesehatan seperti diet untuk ketahanan pangan berkelanjutan dan kebijakan-kebijakan yang mendukung ketahanan pangan, dan aspek kesehatan, sosial, dan ekonomi.
13.	Peran Kebijakan & Indikator Ketahanan Pangan; Transformasi Sistem Pangan Global	(Ruben et al., 2021)	Towards food systems transformation —five paradigm shifts for healthy, inclusive and sustainable food systems	Penelitian ini mendapatkan lima pergeseran paradigma utama yang dibutuhkan untuk pembangunan sistem pangan yang sehat,	Membahas secara keseluruhan transformasi sistem pangan yang menjadikan pergeseran paradigma sebagai dasar dalam identifikasi segmen konsumen

				inklusif, dan berkelanjutan: - Mengembangkan ambisi kebijakan. - Menyelaraskan produksi dan konsumsi agar seimbang - Meningkatkan konektivitas antar semua komponen dalam sistem pangan - Meningkatkan kinerja sistem pangan Mendorong dan memperkuat tata kelola sistem pangan yang inkllusif dan pastisipasi	
14.	Peran Emosi dalam Pilihan Konsumsi	(Wongpraw mas et al., 2021)	Food Choice Determinants and Perceptions of a Healthy Diet among Italian Consumers	- Terdapat dua faktor utama yang merupakan determinan utama dalam pemilihan makanan di Italia.	Mengidentifikasi dan membahas segmen konsumen dan faktor-faktor yang mempengaruhi pilihan makanan

				- Menggunakan metode *k-means clustering* berhasil mengidentifikasi dua segmen utama yaitu, *emotional eating consumer*(konsumen yang dipengaruhi faktor emosional) dan *health-driven consumer*(konsumen yang mengutamakan aspek kesehatan) - Faktor terkuat adalah konsumen memiliki kesadaran kesehatan yang baik dan lingkungan yang sehat.	menggunakan *k-means clustering*.
15.	Inovasi Pengolahan Pangan untuk	(Knorr et al., 2020)	Advancing the Role of Food Processing for Improved	- Proses pengolahan pangan adalah aspek penting	Memperkaya perspektif tentang perbaikan pola konsumsi.

	Keberlanjutan		Integration in Sustainable Food Chains	dalam rantai nilai pangan. - Kurangnya perhatian pada pengolahan pangan untuk keberlanjutan pangan yang hanya terfokus pada produksi pertanian dan konsumsi. - Pada proses pengolahan makanan penting untuk mempertimbangkan dampak lingkungan dan sosial. Untuk meningkatkan efisiensi dan mengurangi limbah diperlukan inovasi dalam teknologi pengolahan pangan.	
16.	Hubungan Konsumsi &	(Delabre et al., 2021)	Actions on sustainable food	- Saat ini sektor produksi dan konsumsi tidak	Integrasi konsumsi dan ekologi.

	Keanekaragaman Hayati		production and consumption for the post-2020 global biodiversity framework	konsisten terhadap visi Konvensi Keanekaragaman Hayati 2050. - Kerangka kerja keanekaragaman hayati pasca-2020 dapat membantu transformasi sistem pangan melalui reformasi subsidi, penilaian nilai, standar keberlanjutan, pengurangan limbah makanan, dan diet berkelanjutan.	
17.	Adaptasi Generasi Baru terhadap Nilai Pangan	(Garai-Fodor, 2021)	Food Consumption Patterns, in a Values-based Approach, for Generation Z	- Memvalidasi dan mengkaji model hubungan antara sistem nilai dan perilaku konsumsi makanan pada Generasi Z	Dapat mengidentifikasi dan mengeksplorasi segmen terhadap nilai dan kelompok konsumen yang mendukung

				Nilai-nilai seperti perhatian teradap kesehatan dan motivasi pemilihan produk makanan berdasarkan nilai.	ketahanan pangan.

Berdasarkan penelitian sebelumnya menunjukkan bahwa segmentasi pola konsumsi masyarakat berperan penting dalam mendukung ketahanan pangan berkelanjutan. Faktor sosio-demografis seperti pendidikan, pendapatan, dan ukuran rumah tangga yang disampaikan Almulhim & Abubakar (2024) dan Biresselioglu et al. (2023) serta nilai pangan dan keselarasan berkelanjutan disampaikan oleh Piracci et al.(2023) dan Wongprawmas et al.(2021) mempengaruhi pola konsumsi.

Kemudian, dapat ditarik kesimpulan bahwa segmentasi pola konsumsi masyarakat adalah pendekatan yang strategis untuk memahami dan mengelola keragaman perilaku konsumsi. Segmentasi dapat dilakukan dengan memanfaatkan teknik statistik, seperti *hierarchical clustering*, *k-means clustering*, dan *gaussian mixture model* (GMM). Selain itu, segmentasi juga dapat memfasilitasi keberagaman faktor sosio-demografis dan psikografis yang dapat mempengaruhi perilaku konsumsi, seperti contohnya usia, pendapatan, motivasi, sampai nilai keberlanjutan ((Almulhim & Abubakar, 2024; Piracci et al., 2023). Kemudian, nilai, sikap, dan motivasi konsumen terhadap kesehatan dan keberlanjutan juga dapat membentuk segmen khusus dalam pola konsumsi masyarakat (Gazdecki et al., 2021; Piracci et al., 2023). Pemakaian model *machine learning* yang efektif akan dibahas studi literaturnya di bawah ini.

Tabel 2. 2. Tabel Kajian Pustaka Metode

No.	Penulis	Judul Penelitian	Metode	Temuan Utama	Relevansi
1.	(Li & Roe, 2024)	Segmenting U.S. consumers by food waste	*K-Means*	- Pola pemborosan pangan sekitar 50% limbah	Pendekatan segmentasi berbasis sikap

		attitudes and behaviors: Opportunities for targeting reduction interventions		makanan di AS terjadi di rumah tangga yang menjadikan perubahan perilaku pada konsumen. - Metode segmentasi penelitian dengan menggunakan *k-means clustering* dan pembobotan sampel dalam mamstikan representasi nasional pada 1.086 responden yang dilakukan secara daring. - Studi mengidentifikasi empat segmen konsumen makanan berdasarkan perilaku dan sikap mereka terhadap limbah makanan: - Conscientious Conservers**:** Konsumen yang sadar lingkungan dan berusaha mengurangi limbah makanan. - Harried Profligates: Konsumen yang sibuk dan	dan perilaku konsumsi masyarakat dapat diterapkan untuk mengidentifikasi kelompok masyarakat dengan potensi perubahan perilaku konsumsi yang signifikan. Dan relevan dengan metode segmentasi yang menggunakan *k-means clustering* dalam menentukan segmen tertentu.

				cenderung boros makanan. - Unrepentant Drink Wasters: Konsumen yang boros dalam mengkonsumsi minuman. - Guilty Carb Wasters: Konsumen yang merasa bersalah karena membuang makanan berkarbohidrat.	
2.	(Abdella et al., 2020)	Sustainability assessment and modeling based on supervised machine learning techniques: The case for food consumption	*K-Means*	- Kerangka kerja model dengan menggabungkan *k-means clustering* dan regresi logistik untuk membangan model berkelanjutan - Akurasi model logistik mencapai 91,67%. - Segmentasi konsumsi sebesar 13,7% dari sektor makanan dan minuman yang dapat dikategorikan	Metode segmentasi yang digunakan *k-means clustering* dapat diterapkan dalam identifikasi kelompok konsumen yang berperilaku dalam cara yang lebih berkelanjutan.

				sebagai tinggi, dengan industri roti dan kue sebagai sektor sentral	
3.	(Wongprawmas et al., 2021)	Food Choice Determinants and Perceptions of a Healthy Diet among Italian Consumers	*K-Means*	Menggunakan metode *k-means clustering* berhasil mengidentifikasi dua segmen utama yaitu, *emotional eating consumer*(konsumen yang dipengaruhi faktor emosional) dan *health-driven consumer*(konsumen yang mengutamakan aspek kesehatan	Mengidentifikasi dan membahas segmen konsumen dan faktor-faktor yang mempengaruhi pilihan makanan menggunakan *k-means clustering*.
4.	(Siemon et al., 2021)	Sequential transfer learning based on hierarchical clustering for improved performance in deep learning based food segmentation	*Hierarchical Clustering*	- Metode Sequential Transfer Learning (STL) berbasis Hierarchical Clustering (HC) meningkatkan akurasi segmentasi makanan hingga 6%. - Waktu pelatihan model berkurang hingga 50%. - Model lebih stabil dan kurang bias dalam	Relevan untuk penelitian menggunakan hierarchical *clustering* dalam analisis data.

				membedakan kategori makanan yang mirip secara visual.	
5.	(Zhang et al., 2021)	Grouping of dynamic electricity consumption behaviour: An f-divergence based hierarchical clustering model	*Hierarchical Clustering*	- Metode *clustering* berbasis *f-divergence* (Hellinger dan total variation)efektif mengelompokkan pola konsumsi listrik dinamis. - Hasil *clustering* berkorelasi dengan faktor eksternal seperti demografi dan suhu.	Relevan untuk penelitian yang memerlukan analisis pola dinamis dan *hierarchical clustering* berbasis probabilitas.
6.	(Alamtaha et al., 2023)	Implementasi Algoritma Hierarchical Clustering dan Non-Hierarchical Clustering untuk Pengelompokkan Pengguna Media Sosial	*K-Means & Hierarchical Clustering*	- Penelitian ini mengelompokkan tingkat penggunaan media sosial di Kabupaten Bone Bolango dengan menggunakan *Hierarchical Clustering* dan *Non-Hierarchical Clustering* (*K-Means & K-Medoids*)	Penelitian ini relevan karena mengimplementasikan *hierarchical clustering* dan *k-means clustering* yang bisa diadaptasi dengan beberapa penyesuaian. Kemudian teknik evaluasinya juga menggunakan *silhouette indeks*

				- Menemukan hasil klaster optimal, yaitu dua klaster, *cluster old* dan *cluster milenial*. - Hasil pengelompokkan dapat memetakan kelompok masyarakat dengan tingkat penggunaan media sosial yang berbeda.	yang dapat diaplikasikan untuk menilai kualitas segmentasi data.
7.	(Ummami & Winarno, 2023)	Gaussian Mixture Model dengan Algoritme Expectation Maximization untuk Pengelompokan Data Distribusi Air Bersih di Jawa Barat	*Gaussian Mixture Model* (GMM)	- Penelitan ini menggunakan metode *Gaussian Mixture Model* dengan menggunakan algoritma *Expectation Maximazation* menunjukkan 6 klaster terbaik berdasarkan nilai *Bayesian Information Criterion* (BIC). - Dari hasil klaster menunjukkan bahwa terdapat beberapa	Relevansi dengan penelitan ini adalah karena menggunakan *Gaussian Mixture Model* sebagai pola segmentasi membagi ke beberapa klaster yang dibentuk oleh algoritma.

				kota/kabupaten yang belum mendapatkan distribusi air bersih.	
8.	(Singhal et al., 2020)	Modeling and prediction of COVID-19 pandemic using Gaussian mixture model	*Gaussian Mixture Model* (GMM)	- Penelitan ini menggunakan GMM untuk memprediksi tren dan total kasus COVID-19 - GMM terbukti efektif digabungkan dengan *Fourier Decomposition Method* (FDM) sebagai pendekatan non-parametrik dalam pemahaman dan prediksi pola penyebaran pandemi.	Relevansi pada penelitian ini adalah penggunaan GMM yang digunakan untuk memodelkan dan memprediksi tren data pada pola konsumsi
9.	(An et al., 2022)	Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection	*Gaussian Mixture Model* (GMM)	- Penelitian ini mengusulkan model gabungan *ensemble unsupervised autoencoders* dengan GMM untuk mendeteksi anomali pada data,	Penelitian ini relevan karena GMM dapat mengelompokkan dan mendeteksi serta memahami pola konsumsi yang tidak dominan dan

				khusunya data serangan siber. - GMM yang dioptimalkan dengan *expectation maximization* (EM) terbukti sangat efektif dalam pengenalan data anomali pada data yang kompleks, termasuk pada data dengan dimensi yang tinggi.	tidak biasa atau segmen minoritas dalam data.

Metode k-means sering dipakai untuk mengategorikan konsumen berdasarkan sikap, perilaku, dan motivasi mereka dalam berbelanja, khususnya terkait dengan pengurangan limbah makanan dan pemilihan makanan yang lebih sehat. Penelitian yang dilakukan oleh Li & Roe, (2024) dan Abdella et al., (2020) mengindikasikan bahwa k-means dapat menemukan kelompok dengan kemungkinan perubahan perilaku yang besar, sehingga intervensi bisa lebih fokus dan tepat sasaran. Selain itu kebijakan dan akses terhadap informasi, turut berperan dalam membentuk kebiasaan konsumsi yang lebih berkelanjutan yang disampaikan oleh Mensah et al. (2024) dan Ruben et al.(2021). Dengan memahami perbedaan pola konsumsi masyarakat dan faktor yang mempengaruhinya akan dilakukan penelitian untuk segmentasi pola konsumsi masyarakat untuk mendukung ketahanan pangan sehingga dapat memberi wawasan strategis dalam pengembangan kebijakan dan intervensi yang lebih efektif untuk meningkatkan ketahanan pangan.

Sebelum memasuki metode *k-means clustering*, metode yang pertama kali akan digunakan adalah metode *hierarchical clustering*. Metode ini sudah terbukti sangat membantu dalam segmentasi pola konsumsi masyarakat. Metode *hierarchical clustering* dapat membentuk struktur hierarki yang dapat merepresentasikan kesamaan antar individu atau rumah tangga dengan berdasar kepada perilaku konsumsi maupun karakteristik demografi. Siemon et al., (2021) dalam jurnalnya dapat menunjukkan bahwa pendekatan *hierarchical*

clustering dapat meningkatkan akurasi segmentasi hingga 6% dalam konteks pengelompokkan makanan dan juga dapat mengurangi waktu pelatihan model secara signifikan yang mengakibatkan model menjadi lebih efisien dan stabil. Kemudian terdapat temuan serupa yang diungkapkan oleh Zhang et al., (2021) yang juga menggunakan *hierarchical clustering*, tetapi berbasis *f-divergence* untuk mengelompokkan data pola konsumsi listriknya. Hasilnya menunjukkan bahwa *clustering* mampu mengungkapkan hubungan antara perilaku konsumsi dan faktor eksternal. Hal ini menunjukkan bahwa metode ini dapat digunakan karena bisa memetakan hubungan antara perilaku konsumsi dengan faktor lainnya. Kombinasi antara *hierarchical clustering* dan *k-means clustering* juga dapat menghasilkan segmentasi yang optimal dan dapat dengan mudah dievaluasi menggunakan *silhouette indeks*, meskipun terdapat batasan pada jumlah variabel dan klaster yang dianalisis seperti yang ditemukan dalam penelitian yang dilakukan oleh Alamtaha et al., (2023).

Metode selanjutnya yang akan dipakai adalah *Gaussian Mixture Model* (GMM). GMM sendiri saat ini semakin banyak digunakan, khususnya pada segmentasi data konsumsi yang lebih kompleks dan dinamis karena GMM merupakan model yang sudah modern. GMM memiliki keunggulan melalui mekanismenya, yaitu *soft clustering* di mana setiap individu diberikan probabilitas keanggotaan dalam setiap klaster, sehingga hasilnya lebih dinamis. Penelitian yang dilakukan oleh Ummami & Winarno (2023) yang menggunakan GMM dengan algoritma *expectation maximization* (EM) pada pengelompokkan distribusi air di Jawa Barat yang hasilnya mampu membagi data menjadi 6 klaster optimal dengan berdasarkan nilai *Bayesian Information Criterion* atau biasa disebut BIC. Keunggulan GMM lainnya ditunjukkan dalam penelitian yang dilakukan oleh An et al., (2022) dalam mendeteksi anomali data dan mendeteksi segmen minoritas atau pola konsumsi yang tidak dominan. Secara garis besar, integrasi GMM dalam segmentasi pola konsumsi dapat memberikan feksibilitas dan akurasi yang lebih tinggi, tepatnya ketika diterapkan pada data dengan keberagaman karakteristik responden yang luas.

Mengenali pola konsumsi merupakan aspek yang sangat krusial dalam membangun ketahanan pangan berkelanjutan. Metode *hierarchical clustering* dapat membantu segmentasi dan menemukan jumlah klaster optimal. Kemudian, teknik seperti *k-means clutering* membantu kita membagi konsumen menjadi beberapa kelompok atau segmen yang berbeda dan *Gaussian Mixture Model* dapat membantu data yang bias mendapatkan klaster yang sesuai.

Dilihat dari sisi teknologi, inovasi dalam pengolahan pangan menjadi salah satu kunci untuk meningkatkan efisiensi sistem pangan dan mendukung program *zero waste concept*. Edukasi konsumen mengenai dampak lingkungan, diet sehat, dan konsumsi pangan lokal juga

terbukti bahwa edukasi mampu mendorong perubahan perilaku konsumsi ke arah yang lebih berkelanjutan.

Segmentasi pola konsumsi memberikan landasan strategis dalam mendukung ketahanan pangan berkelanjutan. Integrasi faktor-faktor yang mempengaruhi seperti sosio-demografis, perilaku, dan kebijakan menjadi elemen penting untuk memastikan bahwa strategi ini berhasil. Tetapi perlu adanya adaptasi lokal karena pengaruh perbedaan budaya, ekonomi, dan kebijakan di setiap negara agar pendekatan-pendekatan ini lebih relevan.

2.2. Segmentasi Pola Konsumsi

Segmentasi pola konsumsi merupakan salah satu metode yang strategis dalam membuat kategori konsumsi ke dalam kelompok atau segmen tertentu dengan berdasarkan perilaku yang atau karakteristik yang sama (Akande et al., 2024). Secara bahasa segmentasi adalah pembagian struktur sosial ke dalam unit-unit tertentu yang sama. Segmentasi membantu dalam proses analisis dalam bidang apa pun, termasuk dalam pengambilan keputusan. Karena pada dasarnya segmentasi membantu mengetahui setiap karakteristik dari konsumen dan dikelompokkan sehingga dapat memudahkan dalam proses analisis dan pengambilan data. Pendekatan dengan metode segmentasi dapat memungkinkan organisasi untuk terlibat secara efektif dengan setiap kelompok dengan mengoptimalkan upaya penjangkauan organisasi tersebut dengan memanfaatkan atribut klien dan menyesuaikannya dengan strategi pemasaran, analisis tren, mengembangkan rencana produk, merancang kampanye, dan menawarkann barang-barang yang relevan dalam proses bisnis (Akande et al., 2024).

Dalam penelitian ini segmentasi dapat membantu untuk mengelompokkan pola konsumsi masyarakat dengan berdasarkan apa yang dikonsumsi. Hal ini dapat membantu untuk menganalisis, mencari solusi, dan membantu dalam pengambilan langkah selanjutnya dalam konteks ketahanan pangan.

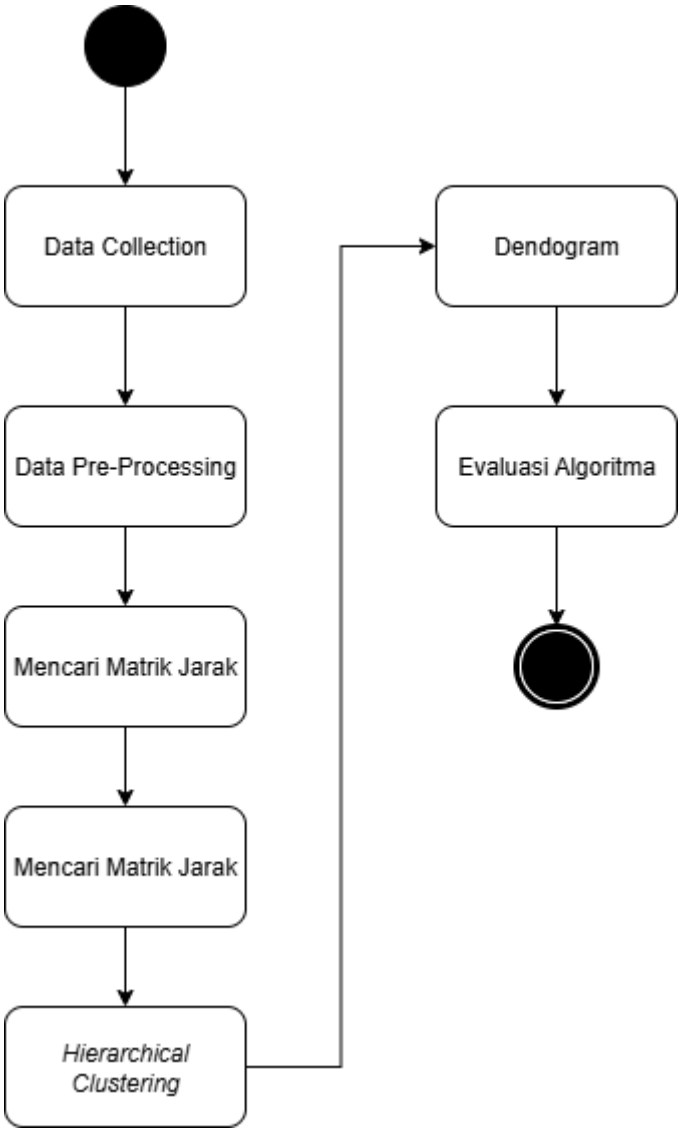
2.3. Ketahanan Pangan

Ketahanan pangan merupakan suatu kondisi di mana semua orang memiliki akses dan kontrol pangan yang sama, yang mencakup kemudahan mendapatkan makanan, aman, bergizi, dan sesuai dengan budaya. Dalam proses mencapai ketahanan pangan berkelanjutan membutuhkan kemampuan produsen dan konsumen untuk beradaptasi terhadap perubahan yang tidak terduga dalam lingkungan, termasuk alam dan kebijakan (Ruben et al., 2021). Adaptasi ini dapat dicapai melalui diversifikasi mata pencaharian, pola makan, dan pasar yang memungkinkan respon fleksibel dan tepat waktu terhadap perubahan global (Ruben et al., 2021).

Ketahanan pangan merupakan salah satu aspek penting dalam sebuah negara yang mengacu pada kesejahteraan penduduk suatu negara tersebut. Indonesia sendiri sudah banyak menerapkan kebijakan sistem ketahanan pangan berkelanjutan untuk masyarakat, tetapi tetap saja pemerataan konsumsi masih belum bisa teratasi. Maka dari itu perlu adanya segmentasi pola konsumsi masyarakat Indonesia untuk mengetahui pola konsumsi masyarakat di berbagai daerah, mengingat Indonesia sendiri merupakan negara kepulauan dengan budaya konsumsi yang berbeda-beda.

2.4. *Hierarchical Clustering*

Hierarchical clustering adalah salah satu metode yang kerap digunakan untuk mengelompokkan data ke dalam struktur berjenjang atau seperti pohon keluarga yang mengacu pada tingkat kemiripan setiap data. Dalam konteks segmentasi pola konsumsi masyarakat, metode ini dapat membantu memperlihatkan bagaimana kelompok-kelompok kecil dengan kemiripan yang hampir sama yang kemudian bersatu menjadi kelompok yang lebih besar secara bertahap. Metode ini memiliki keunggulannya sendiri, yaitu jumlah klaster tidak ditentukan di awal, sehingga proses pembentukan kelompok bisa lebih alami dan fleksibel. Pada penelitan yang dilakukan oleh Siemon et al. (2021) menunjukkan bukti bahwa penggunaan *hierarchical clustering* dapat meningkatkan akurasi segmentasi hingga 6%, yang dilakukan dalam pengelompokan makanan, sehingga dapat menghemat waktu pelatihan model. Pendekatan ini juga mampu mengidentifikasi hubungan anatara perilaku konsumsi dan faktor eksternal seperti demografi (Zhang et al., 2021). Adapun pada penelitian yang dilakukan oleh Alamtaha et al. (2023), *hierarchical clustering* bisa dikombinasikan dengan metode lain seperti *k-means clsutering* untuk dapat memperoleh segmentasi yang optimal dan mudah dievaluasi menggunakan *silhouette indeks*. Oleh karena itu, *hierarchical clustering* dapat dianggap relevan dan efektif untuk mengenali pola konsumsi yang beragam, khususnya di masyarakat yang memiliki karakteristik yang berbeda-beda. Alur diagram proses dari *hierarchical clustering* terdapat pada diagram yang dikemukakan oleh Zhang et al. (2021), tetapi disederhanakan sesuai dengan penelitian ini.



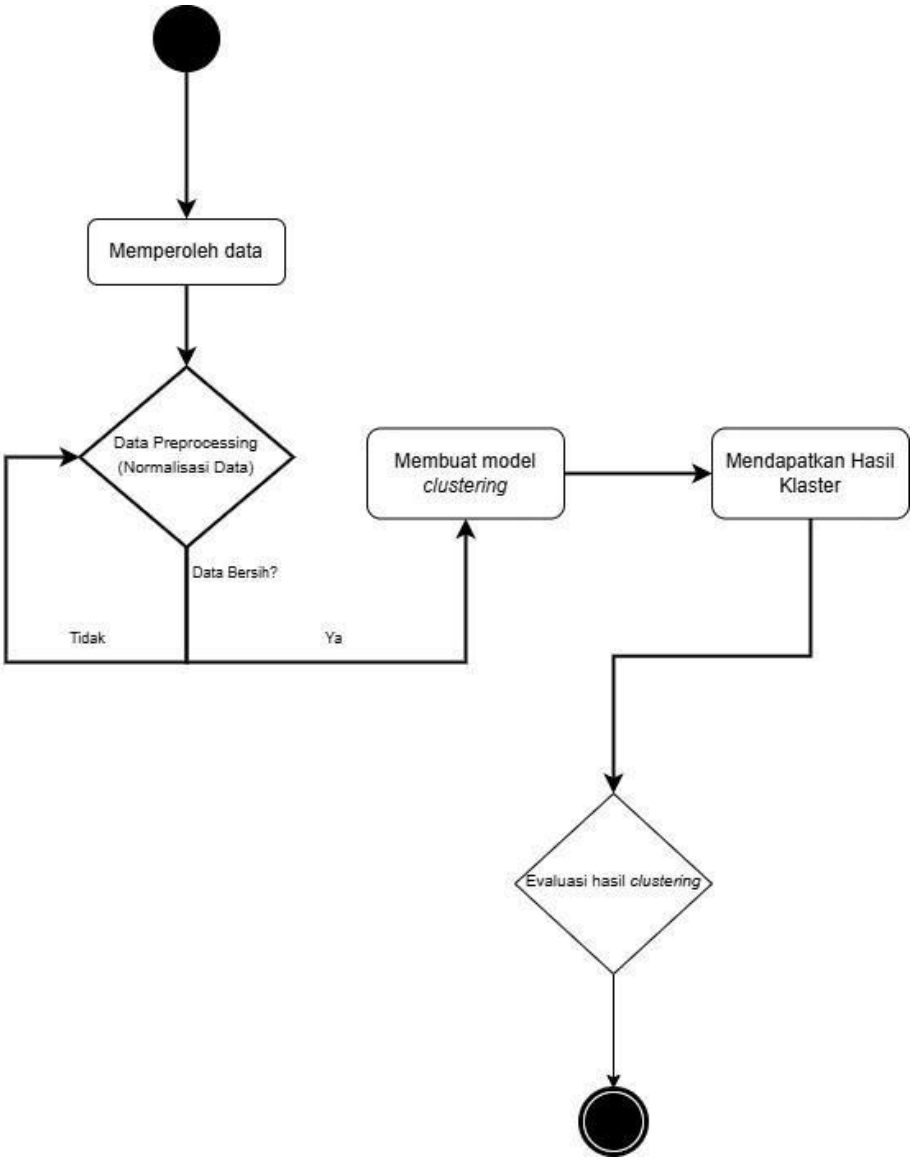
Gambar 2. 2. Alur Proses Algoritma *Hierarchical Clustering*

Gambar di atas menjelaskan alur proses algoritma, dimulai dari pengumpulan data pola makan. Kemudian data di bersihkan dan dinormalisasi sebagai persiapan untuk proses pembentukan klaster. Menghapus atribut yang tidak dipakai dan juga mengubah data non-numerik ke bentuk numerik agar bisa dijalankan oleh algoritma, biasanya disebut dengan *mapping*. Setelah data di prosess maka data masuk ke tahap proses algoritma. Di dalam algoritma terdapat fungsi untuk mencari matriks jarak yang berguna untuk memetakan data ke dalam klaster. Setelahnya baru algoritma *hierarchical clustering* bisa dijalankan. Kemudian hasil visualisasi dari *hierarchical clustering* disebut dengan dendogram. Dendogram terbentuk dari sumbu X dan sumbu Y dengan beberapa titik yang digabungkan sesuai dengan jarak terdekatnya. Bentuknya hampir sama dengan akar pohon. Setelah terbentuk klaster, algoritma perlu dievaluasi agar bisa diketahui seberapa baik algoritma dalam memisahkan data ke dalam klaster.

2.5. K-Means Clustering

K-means clustering merupakan salah satu algoritma *machine learning* yang berjalan tanpa pengawasan atau disebut juga *unsupervised learning* yang digunakan untuk membuat kelompok data ke dalam sejumlah *cluster* tertentu. Pengertian *unsupervised learning* sendiri adalah salah satu metode *machine learning* yang di mana menggunakan algoritma untuk menganalisis dan membuat kelompok dari suatu data yang tidak memiliki label atau kategori yang sudah ditentukan sebelumnya, Tujuan dari membuat segmentasi adalah untuk mengelompokkan pola konsumsi ke dalam segmen atau kelompok tertentu yang memiliki sifat dan karakteristik yang sama, sehingga dapat memahami kebutuhan dan preferensi masyarakat dengan lebih baik. *K-means clustering* bekerja dengan mengelompokkan *dataset* ke dalam jumlah *cluster* (k) yang telah ditentukan sebelumnya, di mana setiap *cluster* memiliki pusat (*centroid*) yang strategis (Tabianan et al., 2022). *K-means clustering* dapat membantu dalam pengelompokkan pola konsumsi masyarakat.

Pada algoritma *k-means clustering* terdapat tahapan yang harus di sesuaikan. Dalam penelitian yang dilakukan oleh Tabianan et al (2022) terdapat diagram yang menunjukkan tahapan dalam menjalankan algoritma *k-means* clustering, tetapi pada diagram ini terdapat beberapa modifikasi menyesuaikan dengan kebutuhan pada tahap normalisasi data. Berikut adalah alur dari tahapan penggunaan algoritma *k-means*.



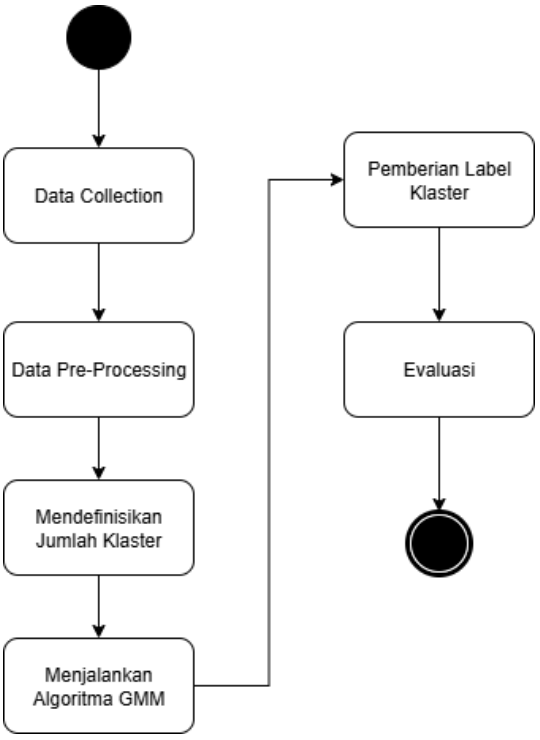
Gambar 2. 3. *Flowchart* step penggunaan algoritma *k-means*

Pada gambar di atas dapat dijelaskan bahwa sebelum menggunakan algoritma *k-means* harus memperoleh data terlebih dahulu. Kemudian data yang sudah didapatkan tidak bisa langsung digunakan begitu saja, karena kemungkinan ada beberapa data yang tidak berbentuk numerik yang tidak bisa digunakan dalam algoritma dan juga ada beberapa nilai *null* yang bisa menghambat jalannya algoritma, maka perlu dilakukannya normalisasi data atau pembersihan data dalam tahap persiapan pemrosesan data.

2.6. Gaussian Mixture Model

Gaussian mixture model (GMM) merupakan pendekatan yang populer dan modern dalam hal segmentasi pola konsumsi, karena kemampuannya dalam membagi data secara lebih

fleksibel. Tidak seperti dua metode sebelumnya yang hanya memasukkan data ke dalam satu kelompok secara tegas, GMM justru dapat memberikan peluang atau ruang untuk setiap individu "berada" di beberapa kelompok sekaligus dengan menyesuaikan probabilitas kemiripannya. Hal ini sangat sesuai untuk mencerminkan realitas pola konsumsi masyarakat yang dinamis. Pada penelitian yang dilakukan oleh Ummami & Winarno (2023), GMM mampu mengidentifikasi enam klaster distribusi air bersih dengan berdasarkan kriteria statistik yang ketat, sehingga setiap kota atau kabupaten dapat dipetakan sesuai dengan kebutuhannya. Penelitian lain yang dilakukan oleh An et al. (2022) menunjukkan keunggulan dari GMM dalam mendeteksi pola konsumsi yang tidak biasa atau minoritas melalui pendekatan *soft clustering* yang lebih adaptif. Dengan pendekatan *soft clustering* ini, GMM sangat efektif saat diterapkan pada data masyarakat yang heterogen, di mana satu data bisa saja memiliki preferensi terhadap pangan sehat, namun juga sensitif terhadap harga atau bahkan terkadang bisa terpengaruh oleh faktor aksesibilitas. Kemudian kesesuaian dengan realitas sosial, di mana dalam kenyataannya, pola konsumsi masyarakat tidak selalu bisa dikelompokkan secara kaku, karena satu orang bisa menunjukkan kecenderungan ke beberapa kelompok sekaligus. Misalnya satu orang peduli terhadap kesehatan, tetapi juga bisa terpengaruh oleh faktor harga. Dengan menggunakan GMM akan mampu menangkap hal bias ini. GMM juga digunakan untuk penyempurnaan struktur klaster dari *k-means*. Hasil awal dari klaster *k-means* digunakan sebagai dasar untuk mengetahui jumlah klaster optimal. Kemudian GMM diterapkan untuk memperhalus hasil tersebut dengan membuat pertimbangan distribusi data yang lebih kompleks. Integrasi GMM sebagai metode segmentasi dapat membantu menyempurnakan hasil dari dua metode klastering sebelumnya, sehingga segmentasi yang dihasilkan menjadi lebih mendalam dan dapat mencerminkan kompleksitas dalam data pola konsumsi.



Gambar 2. 4. *Flowchart* Algoritma *Gaussian Mixture Model*

Flowchart di atas terdapat dalam penelitian yang dilakukan oleh Ummami & Winarno (2023). Alur diagram di atas menunjukkan proses awal dari mengumpulkan data pola makan dan kemudian dibersihkan. Termasuk *mapping* data yang non-numerik ke dalam data numerik. Kemudian menghilangkan atribut yang tidak akan dipakai dalam proses klastering. Sebelum menjalankan algoritma GMM, klaster perlu didefinisikan di awal dengan melihat skor BIC atau *Bayesian Information Criterion*. Setelah data bersih dan klaster ditetntukan barulah algoritma bisa dijalankan. Dan akan hasilnya akan terbentuk klaster yang kita inginkan dan kita beri label klaster sesuai dengan atribut penyusunnya. Setelah itu kita perlu mengevaluasi algoritma agar mengetahui seberapa sesuai GMM memuat data ke dalam setiap klaster.

2.7. Uji Reliabilitas, Validitas, dan *Confirmatory Factor Analysis*

Dalam penelitian ini akan dilakukan uji reliabilitas dan validitas, serta melihat *confirmatory factor analysis* (CFA). Validitas untuk memastikan bahwa item merupakan item yang seharusnya digunakan dan reliabilitas untuk memastikan bahwa konsistensi hasil pengukuran (Sayili et al., 2024). Validitas merupakan ketepatan item atau variabel dalam menggambarkan fenomena yang sedang diukur. Salah satu metode penilaian validitas adalah dengan menggunakan *item-rest correlation*. Metode ini bertujuan mengevaluasi sejauh mana masing-masing item atau variabel sejalan dengan kebutuhan secara keseluruhan. Menurut

Patria et al. (2023), nilai *item-rest correlation* yang lebih tinggi menunjukkan bahwa konsistensi internal atribut yang lebih baik. *Item-rest correlation* diukur dengan menggunakan rumus sebagai berikut.

$$r_{it} = \frac{\sum(X_i - \bar{X}_i)(X_t - \bar{X}_t)}{\sqrt{\sum(X_i - \bar{X}_i)^2 \Sigma(X_t - \bar{X}_t)^2}} \quad (1)$$

r_{it} = *item-rest correlation*

X_i = skor item ke-i

X_t = skor total instrument (variable) tanpa item ke-i

$\bar{X}_i$ dan $\bar{X}_t$ = rata-rata masing-masing skor item dan skor total variable

Pada umumnya rata-rata nilai yang digunakan yaitu jika nilai *item-rest correlation* lebih dari 0,30, yang dianggap sebagai indicator atribut yang baik (Patria et al., 2023).

Reliabilitas mencerminkan konsistensi hasil pengukuran dengan menggunakan teknik yang umum digunakan, yaitu *Cronbach's Alpha*. Reliabilitas dengan *Cronbach's Alpha* merujuk pada tingkat konsistensi internal dalam instrument pengukuran (Sayili et al., 2024). *Cronbach's Alpha* dapat dirumuskan sebagai berikut.

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum\sigma_i^2}{\sigma_t^2}\right) \quad (2)$$

α = *Cronbach's Alpha*.

k = Jumlah item.

σ_i^2 = varians dari masing-masing item.

σ_t^2 = varians total skor instrument

Nilai *Cronbach's Alpha* ≥ 0,70 yang dianggap baik dan bisa menunjukkan bahwa instrumen memiliki konsistensi internal yang tinggi (Sayili et al., 2024). Kemudian terdapat validitas konstruk yang umumnya dapat dievaluasi dengan menggunakan *Confirmatory Factor Analysis* (CFA). CFA adalah analisis statistik yang sering digunakan dalam menguji kesesuaian antara model teori dan data empiris, dengan mengasumsikan bahwa instrumen hanya mengukur konstruk yang telah ditetapkan sebelumnya (Radde et al., 2021). Model CFA dapat dirumuskan secara umum sebagai berikut.

$$X = \Lambda\xi + \delta \quad (3)$$

X = Variabel Terukur (Indikator atau atribut dalam instrument),

Λ = *Factor Loading* (Muatan Faktor),

ξ = Variabel laten (konstruk teoritis yang ingin diukur),

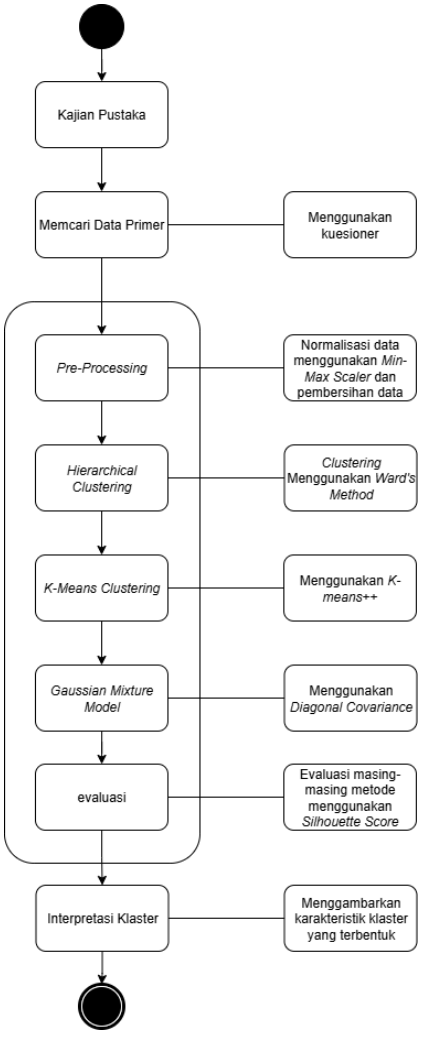
δ = Error atau residual dari pengukuran

.

Dengan melakukan uji validitas dan reliabilitas menggunakan *item-rest correlation* dan CFA, Instrumen atau variable dalam kuesioner penelitian ini dapat menjadi alat ukur yang tepat dalam menjalankan segmentasi pola makan Masyarakat.

BAB III
METODOLOGI PENELITIAN

Metode penelitian yang akan digunakan pada kasus segmentasi adalah *hierarchical clustering*, *k-means clustering*, dan *gaussian mixture model*. Ketiganya saling mendukung satu sama lain agar hasil klaster bisa lebih optimal dan saling melengkapi. Ketiga metode tersebut akan memiliki input data yang sama dari hasil *pre-processing* data yang kemudian akan dilakukan pembentukan klaster oleh ketiga model tersebut dan akan terbentuk klaster. Kemudian ketiga metode tersebut akan dievaluasi untuk melihat seberapa baik proses klastering tersebut dan setiap klaster akan diinterpretasikan. Berikut alur penyelesaian menggunakan ketiga metode di atas.



Gambar 3. 1. Alur Diagram Penyelesaian Masalah

Diagram 3.1 menunjukkan tahapan sistematis dari penelitian ini. Dimulai dari kajian pustaka dengan mengumpulkan, membaca, dan menganalisis literatur atau penelitian terdahulu. Tahapan ini bertujuan untuk memperkuat dasar teoritis, memahami konsep, serta memperjelas metodologi yang akan digunakan dalam penelitian. Selanjutnya, mencari data primer menggunakan metode survei melalui kuesioner. Setelah mendapatkan data primer, tahapan selanjutnya adalah *pre-processing* yang digunakan untuk membersihkan data dari nilai *null* dan *mapping* data yang non-numerik ke dalam data numerik, serta membersihkan atribut yang tidak dipakai. Setelah data siap maka langkah pertama adalah menggunakan *hierarchical clustering* dengan pendekatan *ward's method*. Metode ini digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik antar responden secara bertahap hingga menghasilkan klaster yang optimal. Metode kedua yang digunakan adalah *k-means clustering* dengan inisialisasi *centroid K-means++*. Pendekatan ini bertujuan untuk mengoptimalkan proses klasterisasi dari algoritma sebelumnya, sehingga pembentukan klaster menjadi lebih akurat, efisien, dan konsisten. Analisis *clustering* terakhir adalah dengan *Gaussian Mixture Model* (GMM) menggunakan *Diagonal Covariance*. Metode ini memungkinkan analisis yang lebih fleksibel karena mempertimbangkan adanya distibusi probabilitas data dalam setiap klaster, sehingga pengelompokkan menjadi lebih presisi. Dari tiga metode di atas diterapkan, akan dilakukan evaluasi dengan menghitung *silhouette score* dari masing-masing metode. Skor ini dapat membantuk untuk mengukur seberapa baik hasil klasterisasi dengan melihat tingkat keterpisahan data antar klaster. Tahap terakhir adalah menginterpretasikan klaster yang terbentuk untuk memberikan gambaran dan penjelasan yang jelas mengenai karakteristik setiap klaster yang terbentuk. Interpretasi yang dilakukan ini bertujuan untuk memudahkan pemahaman megenai profil setiap kelompok responden berdasarkan pola konsumsi.

3.1. Hierarchical Clustering

Pada kasus ini metode yang pertama kali digunakan adalah *hierarchical clustering* sebagai tahap awal untuk menganalisis data. Kelompok-kelompok yang dianggap sebagai satu klaster ini secara iteratif akan digabung berdasarkan dengan kemiripan pola konsumsinya, yang diukur dengan menggunakan metrik jarak tertentu sampai seluruh individu tergabung dalam satu klaster besar. Kemudian, visualisasi hasil dari metode *hierarchical clustering* adalah sebuah dendogram. Berikut tahapan utama dalam proses *hierarchical clustering*.

1. Normalisasi Data

   Normalisasi data yang mana merupakan salah satu metode pembersihan dan pemilihan data sebelum data diolah menggunakan *hierarchical clustering*. Salah

satu teknik untuk melakukan normalisasi adalah *Min-Max Scaler*, di mana teknik ini membuat agar variabel yang memiliki satuan berbeda dapat diolah secara konsisten oleh algoritma. Rumus dari *Min-Max Scaling* adalah sebagai berikut.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4}$$

x = nilai asli

x' = nilai yang sudah dinormalisasi

x_{min} = nilai minimum pada data

x_{max} = nilai maximum pada data

2. Menghitung matriks jarak

Pada algoritma *hierarchical clustering* terdapat perhitungan untuk menentukan jarak matriks tiap objek. Tujuannya adalah untuk mengetahui seberapa kemiripan satu objek dengan objek lainnya. Jika dijabarkan dengan rumus matriks, maka rumus untuk menghitung secara manual, adalah sebagai berikut.

$$d(i,j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \tag{5}$$

Keterangan:

$d(i,j)$ = jarak antara objek i dan j.

x_{ik}, x_{jk} = nilai fitur ke-k dari objek i dan j.

n = jumlah fitur

3. Memilih metode *linkage*

Metode *linkage* adalah metode yang akan menentukan bagaimana jarak antar klaster dihitung. Terdapat banyak metode *linkage* dalam *machine learning hierarchical clustering*, tetapi dalam penelitian ini metode *linkage* yang akan digunakan adalah *Ward's Method* yang berguna untuk meminimalkan variasi intra-klaster. Metode ini dipilih karena ketepatan dalam menggabungkan dua klaster yang akan menambah variansi terkecil dan cocok untuk data numerik dan juga

memiliki kemampuan untuk menghasilkan kluster yang seimbang. *Ward's method* sudah tersedia di *library* dalam bahasa pemrograman *python*, tetapi terdapat juga rumus yang digunakan di dalamnya yaitu:

$$D(A,B) = \frac{|A||B|}{|A| + |B|} \|\bar{x}_A - \bar{x}_B\|^2 \tag{6}$$

- $|A|,\ |B|$ = Jumlah anggota dalam klaster A dan B
- $\bar{x}_A,\ \bar{x}_B$ = Centroid masing-masing klister

4. Membuat Dendogram dan Menentukan Klaster Optimal

   Setelah hasil didapatkan dari *linkage* dan menghitung jarak, maka langkah selanjutnya yang akan dilakukan oleh algoritma adalah membuat diagram. Diagram tersebut berbentuk pohon yang merepresentasikan proses penggabungan dua klaster. Diagram inilah yang akan menentukan kita dalam menentukan jumlah klaster yang optimal untuk data tersebut.

   Kemudian langkah selanjutnya adalah menentukan jumlah klaster yang optimal untuk data tersebut. Cara yang paling umum digunakan untuk menentukan klaster adalah dengan memotong dendogram pada jarak vertikal yang terbesar, maka dengan cara tersebut akan membentuk jumlah klaster yang "alami". Kemudian memasukkan setiap objek ke dalam klaster yang telah ditentukan berdasarkan struktur dendogram.

3.2. *K-means Clustering*

Setelah selesai menentukan klaster dan hasil analisis menggunakan *hierarchical* clustering, maka selanjutnya data akan diolah menggunakan k-*means clustering*. Di dalam *unsupervised learning* algoritma bekerja untuk menemukan pola, hubungan, atau struktur dalam data tanpa pengaruh atau dorongan manusia. Dalam kasus kali ini kita perlu menentukan segmen dari data yang sudah didapatkan, maka dari itu *k-means clustering* adalah salah satu cara yang tepat untuk menentukan segmen dari sebuah data. Berikut adalah tahapan penelitian dengan menggunakan metode *k-means clustering*:

1. Normalisasi data

   Normalisasi data yang mana merupakan salah satu metode pembersihan dan pemilihan data sebelum data diolah menggunakan *k-means clustering*. Salah satu teknik untuk melakukan standarisasi adalah *Min-Max Scaling*, di mana teknik ini

membuat agar variabel yang memiliki satuan berbeda dapat diolah secara konsisten oleh algoritma. Rumus dari *Min-Max Scaling* adalah sebagai berikut.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (7)$$

x = nilai asli

x' = nilai yang sudah dinormalisasi

x_{min} = nilai minimum pada data

x_{max} = nilai maximum pada data

2. Menentukan Jumlah Klaster (K)

Penentuan jumlah klaster memiliki metode sendiri agar algoritma berjalan dengan baik. Penentuan jumlah klaster yang optimal biasanya menggunakan *Elbow Method* atau *Silhouette Analysis* untuk melihat titik optimal di mana variasi dalam klaster diminimalkan. Dalam metode *Elbow*, grafik dibuat dengan nilai K yang berbeda dan melihat di mana letak rerjadinya "elbow" atau penurunan yang signifikan pada *within-cluster sum of squares* (WCSS).

Dalam penentuan jumlah klaster (k) metode yang dipakai adalah *Silhouette Analysis* karena metode ini dapat memberikan ukuran kualitas klaster yang lebih objektif dan jelas dengan cara menghitung seberapa sesuai setiap objek cocok dengan klaster sendiri dibandingkan dengan klaster lain. Dan metode *Silhouette Analysis* lebih akurat dalam menentukan jumlah klaster yang optimal, terutama ketika data tidak memiliki struktur klaster yang jelas atau tidak ada "siku" yang terlihat pada *Elbow Method*. Kemudian hasil dari metode ini juga dapat langsung digunakan dalam memastikan bahwa klaster yang terbentuk memiliki pemisah yang baik dan homogen, sehingga penentuan kebijakan yang diberikan akan lebih tepat sasaran.

3. Menjalankan Algoritma dan Evaluasi Hasil Klastering

Setelah menentukan jumlah klaster dengan jumlah K tertentu, maka algoritma *k-means* bisa dijalankan. Algoritma akan mengelompokkan responden ke dalam klaster berdasarkan kemiripan karakteristik yang sudah ditentukan sebelumnya.

Setelah menjalankan algoritma, kita perlu melihat hasil dan perlu melakukan evaluasi terhadap hasil tersebut. Nilai dihasilkan klastering dengan menghitung skor inersia dengan skor di atas 0,5 yang menunjukkan *clustering* yang baik,

sementara skor di bawah 0,5 dapat menunjukkan klaster yang saling bertumpuk satu sama lain

4. Interpretasi Klaster

   Setelah hasil klaster keluar dari algoritma, maka tahapan selanjutnya adalah melakukan analisis karakteristik masing-masing klaster, seperti profil demografi, pola konsumsi, preferensi produk sehat, dan kepedulian terhadap pangan lokal. Kemudian, selanjutnya adalah memberi label pada klaster, seperti "Segmen Kepedulian Kesehatan", "Segmen Pengonsumsi Konvensional", atau "Segmen Pangan Lokal", untuk memudahkan dalam interpretasi hasil.

3.3. Gaussian Mixture Model (GMM)

Metode tambahan yang terakhir untuk menyempurnakan *machine learning* sebelumnya adalah *gaussian mixture model* atau biasa disingkat dengan GMM. GMM merupakan salah satu bagian *model-based clustering* dalam *unsupervised learning*. GMM dipilih sebagai metode lanjutan untuk melengkapi hasil segmentasi yang sebelumnya dilakukan oleh *hierarchical clustering* dan *k-means clustering*. Berbeda dengan algoritma sebelumnya, GMM menggunakan pendekatan *soft clsutering*, di mana setiap individu akan diberikan probabilitas keanggotaan terhadap masing-masing klaster. Pendekatan ini lebih realistis dalam menggambarkan perilaku konsumsi yang kompleks dan dinamis. Berikut tahapan implementasi GMM.

1. Persiapan dan Pra proses Data : Data yang digunakan merupakan hasil transformasi dari kuesioner yang telah melalui tahapan pembersihan, normalisasi (*Min-Max Scaler*).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{8}$$

x = nilai asli

x' = nilai yang sudah dinormalisasi

x_{min} = nilai minimum pada data

x_{max} = nilai maximum pada data

2. Penentuan jumlah klaster: jumlah klaster ditentukan berdasarkan hasil terbaik dari *K-Means*, dan juga divalidasi menggunakan metrik seperti *Bayesian Information Criterion* (BIC) dan *Akaike Information Criterion* (AIC). Nilai BIC/AIC terendah

dipilih untuk memastikan adanya keseimbangan antara kecocokan model dan kompleksitas jumlah klaster.

3. Pelatihan Model GMM : Model ini dilatih menggunakan pustaka *scikit-learn* di *Python*, dengan pengaturan *diag covariance* untuk menangkap bentuk kluster yang tidak simetris. Data yang dianalisis sudah dinormalisasi untuk memastikan kesesuaian skala tiap fitur.
4. Evaluasi dan Interpretasi Hasil : Setiap responden diberikan nilai probabilitas keanggotaan pada semua klaster. Klaster utama diidentifikasi berdasarkan nilai probabilitas tertinggi, sementara nilai-nilai lainnya tetap dianalisis untuk memahami afiliasi sekunder yang relevan. Visualisasi dilakukan dengan grafik distribusi dan reduksi dimensi (PCA) untuk menunjukkan pemisahan antar klaster.
5. Pemberian Label pada Klaster : Setelah kluster terbentuk, setiap kluster dianalisis berdasarkan variabel yang dominan, seperti tingkat konsumsi pangan lokal, perhatian terhadap kandungan nutrisi, dan kesadaran lingkungan. Label diberikan untuk mempermudah pemahaman mengenai profil tiap segmen.

3.4. Desain Penelitian

3.4.1. Pendekatan Penelitian

Pada penelitian ini pendekatan yang digunakan adalah pendekatan kuantitatif dan kualitatif. Pada penelitian ini memungkinkan kita mendapatkan data numerik yang akan diolah secara statistik menggunakan metode *hierarchical clustering*, *k-means clustering*, dan *Gaussian Mixture Model* untuk membagi data dalam beberapa segmen. Kemudian data hasil dari segmen tersebut dipadukan dengan data non-numerik yang tersedia di kuesioner sehingga dapat menghasilkan analisis dan hasil yang akurat. Kedua pendekatan ini diambil untuk mencapai pemahaman yang komprehensif tentang preferensi konsumsi masyarakat dan implikasinya terhadap ketahanan pangan.

Penelitian ini bersifat eksploratif dan deskriptif yang bertujuan untuk identifikasi dan pemetaan pola konsumsi masyarakat di wilayah tertentu. Pada penelitian ini juga memiliki tujuan untuk menggambarkan karakteristik dari segmen-segmen tertentu pada masyarakat dengan berdasar kepada pola konsumsi dan preferensi masyarakat terhadap pangan lokal. Dari desain ini kemungkinan dapat diidentifikasi faktor-faktor utama yang dapat mempengaruhi pola konsumsi masyarakat serta aspek-aspek yang dapat mendukung kebijakan ketahanan pangan berkelanjutan.

3.4.2. Metode Pengumpulan Data

1. Kriteria Partisipan

1) Penduduk usia 17-45 tahun dengan kemampuan pengambilan keputusan atas pola konsumsi di rumah tangga.
2) Keragaman status sosial, seperti tingkat pendapatan untuk melihat faktor ekonomi terhadap pemilihan konsumsi.
3) Jenis konsumsi organik atau konvensional.
4) Aksesibilitas terhadap pangan lokal atau wilayah tempat tinggal, untuk menganalisis perbedaan preferensi berdasarkan kemudahan mendapatkan pangan lokal.
5) Jenis produk konsumsi seperti contoh (kandungan gula dan kandungan lemak).
6) Tingkat kepedulian terhadap pangan lokal, untuk menganalisis perbedaan preferensi berdasarkan kepedulian terhadap pangan lokal.

3.5. Indikator dan Variabel Penelitian

Indikator merupakan salah satu alat ukur untuk mencapai sebuah proses demi mendapatkan petunjuk. Indikator yang digunakan pada penelitian ini disesuaikan dengan kebutuhan penelitian dan tujuan penelitian ini. Kemudian, dari indikator tersebut kita bisa menyusun beberapa pertanyaan yang dibutuhkan untuk diproses menggunakan model. Pada model ini dibutuhkan data numerik, maka variabel yang digunakan juga harus berbentuk numerik. Maka, berikut ini adalah daftar indikator yang akan digunakan dalam penelitian ini.

Tabel 3. 1. Tabel Indikator dan Pertanyaan Kuesioner

Indikator	Pertanyaan	Variabel	Skala Pengukuran
	1. Berapa kali dalam seminggu Anda mengonsumsi makanan pokok (nasi, umbi, sagu)?	konsumsi_pangan_pokok	Rasio (frekuensi)

Pola Konsumsi Pangan	2. Berapa persen pengeluaran Anda untuk pangan dibandingkan non-pangan?	persentase_pengeluaran_pangan	Rasio (persentase)
	3. Sejauh mana Anda lebih memilih pangan lokal daripada impor? (1–5)	preferensi_lokal	Ordinal
	4. Berapa kali Anda mengonsumsi protein hewani dalam seminggu?	konsumsi_protein_hewani	Rasio (frekuensi)
	5. Berapa kali Anda mengonsumsi protein hewani dalam seminggu?	konsumsi_protein_nabati	Rasio (frekuensi)
	6. Berapa kali Anda membeli makanan instan/olahan dalam seminggu?	konsumsi_makanan_instan	Rasio (frekuensi)
Preferensi Pangan Berkelanjutan	7. Apakah Anda sadar tentang dampak lingkungan dari kebiasaan pola konsumsi Anda?	kesadaran_lingkungan	Skala Likert (1-5)
	8. Berapa persen produk organik/lokal yang Anda beli?	persentase_organik	Rasio (persentase)

	9. Seberapa sering Anda menghindari pemborosan makanan (food waste)?	pengurangan_food_waste	Skala Likert (1-5)
	10. Apakah Anda pernah ikut serta dalam kegiatan pertanian di komunitas/kampung?	partisipasi_pertanian	Nominal (Biner: 0 = tidak dan 1 = ya)
Aksesibilitas Terhadap Pangan	11. Berapa jarak rumah Anda ke pasar tradisional?	jarak_pasar_tradisonal	Rasio (numerik)
	12. Berapa jarak rumah Anda ke supermarket/minimarket	jarak_pasar_modern	Rasio (numerik)
	13. Seberapa mudah Anda untuk mendapatkan makanan bergizi?	Akses_pangan_bergizi	Skala Likert (1-5)
	14. Apakah Anda ketergantungan terhadap pangan impor?	ketergantungan_impor	Skala Likert (1-5)
Faktor Ekonomi	15. Apakah Anda sensitif terhadap harga bahan pangan organik (jika lebih mahal)?	sensitivitas_harga_organik	Skala (non-numerik)
	16. Berapa pendapatan rumah tangga Anda perbulan?	pendapatan	Rasio (numerik)

	17. Berapa persen dari pengeluaran yang dialokasikan untuk pangan berkelanjutan (organik/lokal)?	alokasi_pangan_berkelanjutan	Rasio (persentase)
Kepedulian Terhadap Kesehatan	18. Seberapa sering Anda membaca label nutrisi pada kemasan makanan sebelum membeli?	perhatian_nutrisi	Skala Likert (1-5)
	19. Apakah Anda memiliki kebiasaan untuk menghindari makanan tinggi gula?	menghindari_tinggi_gula	Skala (non-numerik)
	20. Apakah Anda memiliki kebiasaan untuk menghindari makanan tinggi lemak?	menghindari_tinggi_lemak	Skala (non-numerik)
	21. Seberapa penting menurut Anda menjaga pola makan sehat untuk mencegah penyakit?	pola_makan_sehat	Sakal Likert (1-5)
	22. Berapa kali dalam seminggu Anda mengonsumsi buah dan sayur?	konsumsi_buah_sayur	Rasio (frekuensi)
	23. Apakah Anda secara rutin melakukan pemeriksaan kesehatan (*check-up*) terkait pola makan?	pemeriksaan_kesehatan	Nominal (Biner: 0 = tidak dan 1 = ya)

	24. Apakah Anda mengikuti program penyuluhan atau sosialisasi terkait gizi/kesehatan?	partisipasi_sosialisasi	Nominal (Biner: 0 = tidak dan 1 = ya)
Karakteristik Sosio-Demografi	25. Siapa nama Anda? (boleh diisi inisial)	Nama	Isian singkat
	26. Berapa usia Anda?	usia	Isian singkat
	27. Jenis kelamin	jenis_kelamin	Nominal (Biner: 0 = perempuan dan 1 = laki-laki)
	28. Tingkat pendidikan terakhir	pendidikan	Pilihan ganda
	29. Pekerjaan utama	Pekerjaan	Pilihan ganda
	30. Berapa jumlah anggota keluarga Anda?	jumlah_keluarga	Isian singkat
	31. Lokasi tempat tinggal	tempat_tinggal	Pilihan ganda
	32. Apakah Anda mengelola/memiliki lahan pertanian?	lahan_pertanian	Nominal (Biner: 0 = tidak dan 1 = ya)

Indikator yang terdapat pada tabel di atas disesuaikan dengan penelitian terdahulu. Pertama indikator pola konsumsi pangan didapatkan berdasarkan penelitian yang dilakukan oleh Li & Roe (2024), Schäufele-Elbers & Janssen (2023), Wongprawmas et al. (2021) yang menganalisis pentingnya memahami kecenderungan masyarakat dalam memilih jenis makanan

dan juga frekuensi konsumsinya sebagai dasar dalam membentuk segmen konsumsi pangan yang representatif. Indikator preferensi pangan berkelanjutan yang mempertimbangkan dimensi nilai keberlanjutan dan perilaku konsumsi ramah lingkungan serta praktik dalam pengolahan limbah makanan (*food* waste). Disesuaikan penelitian terdahulu dari Gazdecki et al. (2021), Piracci et al. (2023), dan Ran et al. (2022) yang menemukan bahwa keberlanjutan telah menjadi salah satu faktor utama dalam pengambilan keputusan konsumsi, sehingga penting untuk dimasukkan dalam analisis segmentasi. Indikator aksesibilitas terhadap pangan ditentukan dengan aspek akses terhadap pangan dan juga distribusi pangan yang menyesuaikan hasil penelitian dari Ashari (2023), Drewnowski et al. (2020), dan Mensah et al. (2024). Akses terhadap pangan, khususnya di daerah rentan atau jauh dari perkotaan merupakan determinan penting dalam mewujudkan ketahanan pangan. Kemudian terdapat indikator dari faktor ekonomi dengan fokus pada pengaruh ekonomi dan pendapatan pada segmentasi dan konsumsi berkelanjutan dengan didasarkan dari penelitian Abdella et al. (2020), Biresselioglu et al. (2023), dan Piracci et al. (2023) yang memberikan dasar bahwa kemampuan ekonomi sangat memengaruhi akses dan juga preferensi dari konsumsi terhadap makanan sehat dan juga keberlanjutan. Selanjutnya adalah indikator kepedulian terhadap kesehatan di mana segmentasi didasarkan pada motivasi kesehatan dan juga kesadaran akan pangan sehat yang diambil dari penelitian Piracci et al. (2023) dan Wongprawmas et al. (2021), yang menunjukkan bahwa konsumen dengan tingkat kesadaran akan kesehatan yang tinggi dapat memiliki preferensi pangan yang berbeda secara signifikan. Dan indikator terakhir adalah karakteristik sosio-demografi, yaitu variabel sosio-demografi, seperti usia, jenis kelamin, tingkat pendidikan, dan ukuran rumah tangga sebagai dasar pembentukan segmen konsumsi pangan dengan didasarkan oleh penelitian yang dilakukan oleh Almulhim & Abubakar (2024) dan Biresselioglu et al. (2023). Pada penelitiannya menekankan bahwa faktor demografis dapat memiliki peran dalam pembentukan pola konsumsi pangan masyarakat.

BAB IV
HASIL DAN PEMBAHASAN

4.1. Deskripsi Data Awal

Pada penelitian ini digunakan data primer yang langsung diambil di masyarakat. Data ini melibatkan seluruh kalangan masyarakat agar data dapat bervariatif dan bisa merepresentasikan sedikit atau sebagian dari pola makanan masyarakat. Responden melibatkan masyarakat yang tersebar di wilayah pedesaan hingga perkotaan dengan rentang usia bervariatif, tetapi data yang akan diolah hanya pada responden yang berusia produktif, yaitu pada rentang usia 17 sampai 45 tahun saja. Karakteristik dari responden adalah mencakup latar belakang responden, jenis kelamin, pekerjaan, dan jumlah anggota keluarga responden. Mayoritas responden adalah berusia produktif dengan latar belakang pendidikan SMA sampai Pascasarjana serta banyak terdapat mahasiswa, serta bekerja sebagai pekerja pada sektor formal dan informal. Jika dilihat dari sisi tempat tinggal, mayoritas responden berasal dari wilayah semi-urban. Sebagian kecil dari responden memiliki lahan pertanian yang dikelola secara pribadi.

Data ini memiliki batasan yaitu tidak bisa mencakup semua wilayah di Indonesia, tetapi data isi bisa sedikit merepresentasikan pola makan masyarakat yang ada di Indonesia dengan berbagai latar belakang tempat tinggal. Data ini dapat mendukung pengamatan terhadap preferensi dan aksesibilitas pangan lokal.

Untuk melihat gambaran awal tentang struktur dan tipe data yang digunakan, berikut adalah kode dalam *machine learning* yang digunakan untuk menampilkan data awal.

```
#TAHAP 2: Upload file manual
uploaded = files.upload()

# Load Data
file_path = list(uploaded.keys())[0]
df = pd.read_csv(file_path)

# Menampilkan 5 baris pertama dari DataFrame untuk
mendapatkan gambaran umum tentang data
print("Lima baris pertama dari dataset:")
print(df.head())

# Menampilkan informasi ringkas tentang DataFrame,
termasuk tipe data dan nilai non-null
print("\nInformasi dataset:")
df.info()
```

Gambar 4. 1. *Code* Program Menampilkan Data Awal

Dari kode di atas kita bisa melihat data kita, tetapi hanya 5 baris awal yang ditunjukkan. *Output* awal ini menunjukkan bahwa dataset di atas terdiri dari 33 kolom dengan tipe data *int64* dan *object*. Akan ada beberapa kolom yang akan dihapus dalam tahap pra-pemrosesan karena tidak relevan dalam klastering, seperti kolom "Timestamp" dan kolom yang berisi informasi identitas individu.

4.2. *Pre-Processing* Data

4.2.1. Pembersihan Tahap Awal

Pada pembersihan data tahap awal ini akan difokuskan untuk mengubah data yang non-numerik ke dalam data numerik atau biasa disebut pemetaan. Contoh data yang diubah misalnya terdapat frekuensi konsumsi yang berisi "1-2 kali", "3-4 kali", "5-6 kali", dan "Setiap Hari", maka akan diubah menjadi 1 sampai 4, agar data bisa diproses oleh algoritma. Begitu pun seterusnya terhadap data yang non-numerik. Dan bila terdapat *binary option* seperti "ya" atau "tidak" dan "laki-laki" dan "perempuan", makan akan diubah menjadi 0 dan 1.

4.2.2. Pembersihan dan Pemfilteran Data

Pada tahapan ini, pembersihan data akan dilakukan. Pembersihan data dibersihkan dan akan difilter pada data usia. Pada penelitian ini, ranah usia akan dibatasi, yaitu hanya pada usia produktif saja yang akan dilakukan klastering. Hal ini dilakukan karena pada usia produktif preferensi makanan cenderung ditentukan oleh diri sendiri tanpa pengaruh orang

lain. Dan juga hal ini dilakukan untuk memastikan bahwa analisis klaster hanya berfokus pada populasi yang relevan dengan pengambilan keputusan pola konsumsi rumah tangga.

```
# TAHAP 3: PEMUATAN, PEMBERSIHAN, DAN FILTER DATA
try:
    df_ori = pd.read_csv('Kuesioner Pola Makan Masyarakat.csv')
    print(f"Dataset berhasil dimuat dengan {len(df_ori)} baris.")
except FileNotFoundError:
    raise SystemExit("Error: Pastikan file 'Kuesioner Pola Makan
Masyarakat.csv' sudah diunggah.")

# Cari dan bersihkan kolom usia
age_col_name = next((col for col in df_ori.columns if 'usia' in
col.lower()), None)
if age_col_name:
    df_ori['Usia_bersih'] = clean_age_data(df_ori[age_col_name])
else:
    raise SystemExit("Error: Kolom usia tidak ditemukan dalam
dataset.")

# Filter data
responden_awal = len(df_ori)
df_filtered = df_ori[df_ori['Usia_bersih'].notna() &
df_ori['Usia_bersih'].between(17, 45)].copy()
print(f"Jumlah responden awal: {responden_awal}")
print(f"Jumlah responden setelah difilter (usia 17-45 & valid):
{len(df_filtered)}")
```

Gambar 4. 2. *Code* Pembersihan dan Filter Data

Dari kode di atas sebelum kita melakukan pembersihan data akan dilakukan terlebih dahulu *try error*, dimana fungsi adalah untuk penanganan eror pada bahasa pemrograman *python* dengan memanggil fungsi “try...except”. Kemudian, pada data original masih terdapat banyak usia yang beragam penulisannya, seperti “26 Tahun”, “45thn”, dan sebagainya. Hal ini menunjukkan bahwa data masih berbentuk *string* dan tidak dapat difilter atau diproses pada program. Oleh karena itu, diperlukan pembersihan data agar semua data usia seragam, karena tujuan utama program adalah memfilter usia dari 17 hingga 45 tahun saja yang akan diperoses.

Fungsi yang digunakan adalah *clean_age_data(...)*. Fungsi ini akan menerima satu kolom, yaitu kolom usia yang asli sebagai input, kemudian akan membersihkan data, contohnya akan mengubah "25 tahun" menjadi 25, "empat puluh" menjadi 40, atau mengubah teks yang tidak valid menjadi nilai kosong/null, kemudian akan mengembalikan data usia yang sudah bersih dan dalam format angka yang akan disimpan dalam kolom baru bernama Usia_bersih dengan memanggil fungsi *df_ori['Usia_bersih'] = clean_age_data(df_ori[age_col_name])*. Secara dinamis tahapan ini adalah mencari kolom yang berkaitan dengan usia, lalu membersihkannya dan menyimpan di kolom baru. Maka hasil *output* dari kode di atas adalah sebagai berikut.

Setelah data dibersihkan dari nilai *null* ataupun data yang tidak sesuai, kemudian kita filter data yang akan digunakan, yaitu hanya usia 17-45 tahun. Maka akan kita gunakan fungsi *df_ori['Usia_bersih'].between(17, 45)* yang akan memilih baris di kolom baru dengan memfilter usia dari 17 hingga 45 tahun saja.

```
Dataset berhasil dimuat dengan 385 baris.
384 data usia berhasil dibersihkan dan divalidasi.
Jumlah responden awal: 385
Jumlah responden setelah difilter (usia 17-45 & valid): 353
```

Gambar 4. 3. *Output Code* Pembersihan dan Filter Data

Melihat hasil keluaran dari kode di atas maka bisa kita simpulkan bahwa data awal sebanyak 385 data yang kemudian terdapat satu usia dengan nilai *null* yang tidak dapat diproses. Dan data total adalah 384 yang kemudian difilter lagi melalui usia 17-45 tahun yang menghasilkan data siap proses sebanyak 353. Pemilihan rentang usia ini bertujuan agar memfokuskan analisis pada kelompok usia produktif yang memiliki kemandirian dalam mengambil keputusan konsumsi.

4.2.3. Pra-Pemrosesan Tahap Akhir

Setelah data difilter, maka kita masuk ke dalam tahap pembersihan data terakhir yang menggunakan *min-max scaling*. Sebelumnya, kita harus melakukan definisi fungsi *min-max scaling* yang terdapat di *library* dengan memanggil fungsi *MinMaxScaler()*, Seperti pada kode di bawah ini.

```
def preprocess_data(df):
       scaler = MinMaxScaler()
       df_norm = scaler.fit_transform(df_proc)
        print("Data telah dinormalisasi menggunakan MinMaxScaler.")
```

Gambar 4. 4. Code MinMax Scaler

Setelah kita mendefinisikan bahwa kita akan menggunakan *min-max scaling* maka step selanjutnya kita bisa langsung memanggil fungsi yang akan kita gunakan, yaitu *preprocess_data(df)*. Kolom yang tidak diperlukan juga akan dihapus di sini. Kolom yang tidak diperlukan seperti *Timestamp*, nama dan usia asli. Karena usia sudah dibuatkan kolom baru sebelumnya. Kode yang kita gunakan dalam tahap pra-pemrosesan akhir adalah sebagai berikut.

```
# TAHAP 4: PRA-PEMROSESAN AKHIR
print(f"Jumlah fitur sebelum pra-pemrosesan: {df_filtered.shape[1]}")
df_normalized = preprocess_data(df_filtered)
print(f"\nPra-pemrosesan selesai. Data siap untuk clustering dengan
{df_normalized.shape[1]} fitur.")
df_normalized.head()
```

Gambar 4. 5. *Code* Pra-Pemrosessan Akhir

Kemudian terdapat juga pada proses ini untuk penanganan *missing values*, yaitu nilai kosong atau hilang diisi menggunakan nilai median untuk kolom numerik dan modus untuk kolom kategorikal.

4.3. Pemilihan Attribut

Pemilihan atribut yang akan digunakan dalam klastering sangat penting, karena akan mempengaruhi hasil evaluasi setiap metode. Pemilihan atribut akan dilakukan menggunakan uji reliabilitas dan validitas untuk menentukan atribut mana yang cocok digunakan. Berikut hasil dari uji reliabilitas dan validitas dari setiap atribut dari dataset.

Tabel 4. 1. Tabel Uji Reliabilitas dan Validitas

Atribut	SD	*Item-Rest Correlation*
1.	1.007	0.2256
2.	0.611	0.2966

3.	0.841	0.2325
4.	0.751	0.2965
5.	0.754	0.5020
6.*	0.880	0.5231
7.	0.836	0.0919
8.	0.829	0.3273
9.*	1.101	0.2017
10.*	0.499	0.2768
11	0.723	0.1043
12.*	0.758	0.1135
13.	0.745	0.2597
14.*	1.255	0.5390
15.*	0.578	0.1796
16.	0.955	0.2031
17.	0.683	0.3947
18.*	0.821	-0.0127
19.*	0.548	0.0233
20.*	0.503	0.0774
21.	0.746	0.2234
22.	0.764	0.4237
23.*	0.500	0.1488
24.*	0.500	0.1537
26.	10.506	0.3972
27.*	0.481	0.3175
30.	1.282	0.1938
31.	0.654	-0.0715
32.*	0.483	0.2923

Dari hasil uji reliabilitas dan validitas di atas, atribut berbentuk angka mewakili nommor pertanyaan yang tertera pada data. Kemudian pemilihan fitur yang sudah diiuji akan dipilih berdasarkan nilai yang lebih dari 0.3. Untuk nilai diantara 0.2 sampai 0.3 akan dipertimbangkan juga melalui tabel hasil dari analisis faktor yang disebut *confirmatory factor analysis* (CFA) pada table di bawah ini.

Tabel 4. 2. Tabel Hasil Confirmatory Factor Analysis (CFA)

Indikator	Atribut	Estimate
Pola Konsumsi	1	0.35630
	2	0.17278
	3	0.19665
	4	0.35162
	5	0.42276
	6	-0.56269
Preferensi Pangan	7	0.12425
Berkelanjutan	8	-0.11142
	9	0.31908
	10	0.33929
Aksesibiltas Terhadap	11	-0.11746
Pangan	12	0.15554
	13	-0.10294
	14	0.67303
	15	0.16142
	16	-0.41268
	17	-0.4686
Kepedulian Terhadap	18	0.43636
Kesehatan	19	0.21360
	20	0.22952
	21	0.05931
	22	-0.01430
	23	0.37126
	24	0.38383
Karakteristik Sosio-	26	4.39293
Demografi	27	-0.22476
	30	0.26033
	31	-0.00453
	32	-0.22545

Dari hasil kedua tabel di atas, maka dapat disimpulkan bahwa beberapa atribut yang akan digunakan dalam proses segmentasi adalah atribut nomor 4, 5, 6, 10, 14, dan 17. Hal ini dapat dibuktikan dengan nilai dari kedua tabel di atas. Pada atribut 5, 6, 14, dan 17 nilai *item-rest correlation* menunjukkan angka lebih dari 0,3 yang berarti sudah bisa dipakai dalam proses *clustering*. Kemudian untuk atribut 4 dan 10 yang masing-masing memiliki nilai 0.2965 dan 0.2768 masih harus dipertimbangkan dengan menggunakan CFA. Hasil CFA dari atribut nomor 4 dan 10 adalah 0.35162 dan 0.33929 yang mengindikasikan bahwa kedua atribut tersebut layak dipakai. Sisa dari atribut yang ada di atas mengindikasikan bahwa, mungkin nilai *item-rest correlation* memadai dengan lebih dari 0.3 atau diantara 0.2 dan 0.3, tetapi setelah melihat nilai faktor *loading* atau *estimate* yang menunjukkan di bawah 0.3, maka dipastikan atribut terseebut tidak terpakai.. Nomor item didasarkan pada nomor pertanyaan di dalam tabel 3.1, yaitu atribut nomor 4 adalah frekuensi konsumsi protein hewani, nomor 5 adalah frekuensi konsumsi nabati, nomor 6 adalah frekuensi konsumsi makanan instan, nomor 10 adalah partisipasi kegiatan pertanian, nomor 14 adalah ketergantungan terhadap pangan impor, dan nomor 17 adalah persentase alokasi pangan berkelanjutan.

Tabel 4. 3. Tabel Uji Reliabilitas

	SD	**Cronbach's α**
scale	0.284	0.739

Berdasarkan hasil analisis reliabilitas instrumen menggunakan *Cronbach's Alpha*, diperoleh nilai sebesar 0.739 yang menunjukkan bahwa instrumen memiliki konsistensi internal yang baik. Seluruh item yang dipilih untuk analisis segmentasi memiliki nilai korelasi item-total terkoreksi di atas 0.3, sehingga layak dimasukkan ke dalam proses klastering. Selanjutnya, hasil analisis validitas faktor menggunakan CFA menunjukkan bahwa mayoritas atribut memiliki faktor *loading* yang signifikan dan memadai (>0.3), terutama pada atribut yang merefleksikan pola konsumsi pangan dan aksesibilitas pangan. Hal ini menegaskan bahwa indikator yang digunakan dalam segmentasi memang mengukur bentuk segmen yang dimaksud. Oleh karena itu, variabel-variabel ini dapat dinilai layak dan relevan untuk digunakan dalam analisis segmentasi dengan menggunakan metode klastering. Sebelum

melakukan klastering, perlu diperhatikan pada beberapa atribut yang bernilai terbalik, sehingga perlu dilakukan *reverse scoring* dahulu agar interpretasi hasil lebih akurat.

4.4. Penentuan Jumlah Klaster Optimal

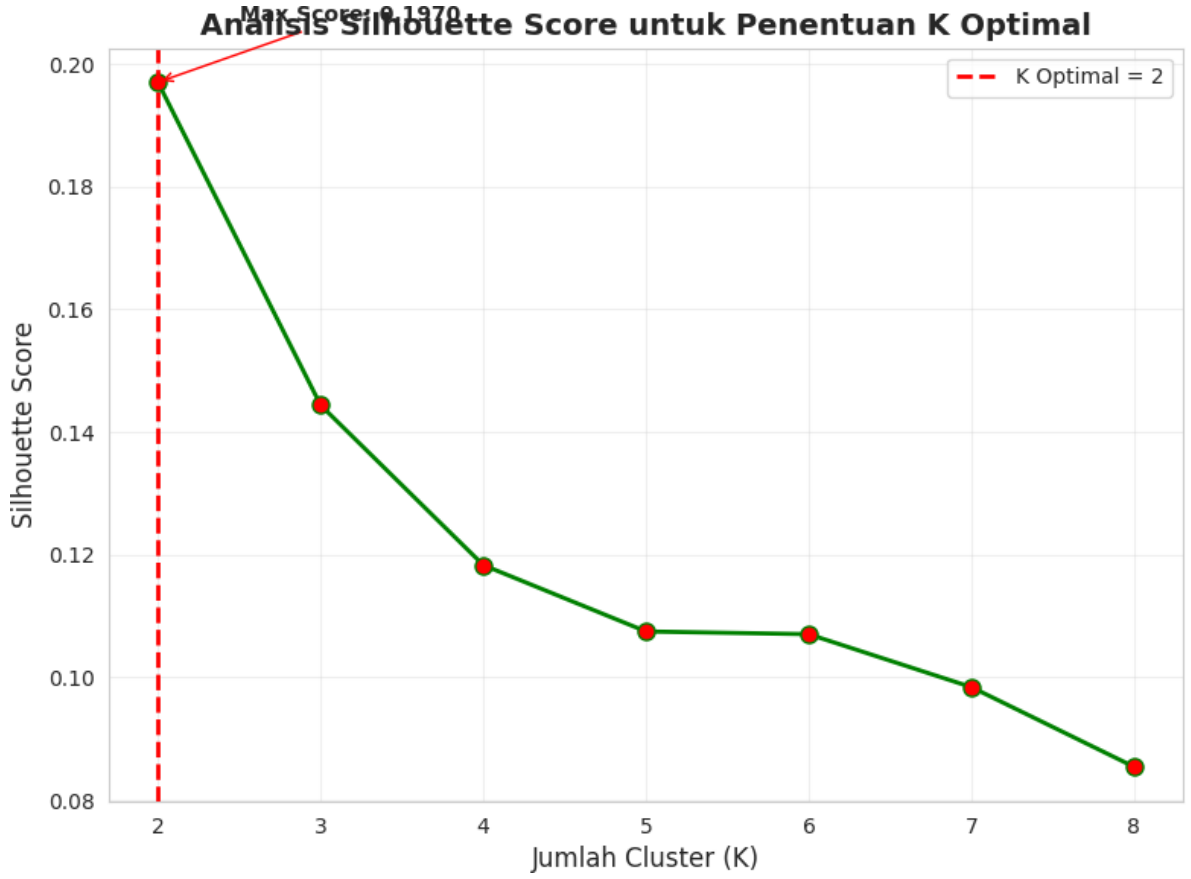
Langkah selanjutnya adalah penentuan jumlah klaster optimal yang dianggap krusial. Penentuan ini agar segmentasi yang dihasilkan bisa benar-benar merepresentasikan variasi pola konsumsi di masyarakat. Dalam penelitian ini akan digunakan metode *silhouette score* untuk memberikan evaluasi beragam kemungkinan jumlah klaster yang objektif. *Silhouette score* berkisar antara -1 sampai 1, yang menunjukkan bahwa nilai mendekati 1 berarti objek sangat cocok dengan klasternya dan terpisah dengan baik dengan klaster lain. Berikut akan ditunjukkan *code* yang akan menentukan hasil optimal klaster.

```
def plot_optimal_clusters(k_range, silhouette_scores, optimal_k):
    plt.figure(figsize=(8, 6))
    plt.plot(k_range, silhouette_scores, 'go-', markerfacecolor='red',
markersize=8, linewidth=2)
    plt.title('Analisis Silhouette Score untuk Penentuan K Optimal',
fontsize=14, fontweight='bold')
    plt.xlabel('Jumlah Cluster (K)', fontsize=12)
    plt.ylabel('Silhouette Score', fontsize=12)
    plt.axvline(x=optimal_k, color='red', linestyle='--', linewidth=2,
label=f'K Optimal = {optimal_k}')
    max_score = max(silhouette_scores)
    plt.annotate(f'Max Score: {max_score:.4f}',
                 xy=(optimal_k, max_score),
                 xytext=(optimal_k+0.5, max_score+0.01),
                 arrowprops=dict(arrowstyle='->', color='red'),
                 fontsize=10, fontweight='bold')
    plt.legend()
    plt.grid(True, alpha=0.3)
    plt.tight_layout()
    plt.savefig('silhouette_analysis.png', dpi=300, bbox_inches='tight')
    plt.show()
def find_optimal_clusters(data, max_k=8):
    print(f"\n=== MENCARI JUMLAH CLUSTER OPTIMAL ===")
    print(f"Data yang digunakan: {data.shape}")
    print(f"Fitur: {list(data.columns) if hasattr(data, 'columns') else
'Array NumPy'}")

    k_range = range(2, max_k + 1)
    silhouette_scores = []
    for k in k_range:
        kmeans = KMeans(n_clusters=k, init='k-means++', random_state=42,
n_init=10, max_iter=300)
        labels = kmeans.fit_predict(data)
        score = silhouette_score(data, labels)
        silhouette_scores.append(score)
        print(f"  K = {k}: Silhouette Score = {score:.6f}")
    optimal_k_math = k_range[np.argmax(silhouette_scores)]
    max_score = max(silhouette_scores)
    print(f"K Optimal (matematis): {optimal_k_math}")
    print(f"Silhouette Score Terbaik: {max_score:.6f}")
    print(f"Namun untuk tujuan penelitian, akan menggunakan K = 3")
    plot_optimal_clusters(k_range, silhouette_scores, optimal_k_math)
    return optimal_k_math, silhouette_scores
optimal_k_saran, all_silhouette_scores=find_optimal_clusters(df_normalized)
```

Gambar 4. 6. *Code* untuk mencari K optimal

Dari algoritma di atas terdapat dua fungsi. Fungsi pertama untuk membuat visualisasi hasil analisi pada jumlah klaster (K) optimal pada proses klastering. Membuat plot garis dari *range K* atau jumlah klaster yang diuji terhadap nilai *silhouette score* untuk tiap K dan membuat tanda garis vertikal untuk menandai klaster optimal dan memberi label, judul dan legenda. Berikut adalah hasil keluaran dari mencari nilai optimal K.



Gambar 4. 7. Nilai K Optimal Menggunakan *Silhouette Score*

Hasil dari menentukan klaster optimal sebagaimana ditunjukkan pada gambar di atas. Maka hasilnya menunjukkan klaster optimal adalah 2 atau K=2. Skornya menunjukkan hampir menyentuh 0,2. Skor ini adalah yang tertinggi dibandingkan dengan jumlah klaster lain. Tetapi dalam penggunaannya untuk segmentasi pola, K=2 atau 2 klaster tidak cukup relevan, karena data pola konsumsi yang sangat beragam, dinamis, dan banyak data yang tidak secara tegas masuk ke dalam satu klaster saja.

Tabel 4. 4. Tabel Perbandingan *Silhouette Score*

Jumlah Klaster (K)	*Silhouette Score*

K=2	0.197014
K=3	0.144363
K=4	0.118282
K=5	0.107465
K=6	0.107017
K=7	0.098383
K=8	0.085387

Dari tabel perbandingan di atas K=2 menunjukkan klaster optimal. Tetapi karena untuk tujuan segmentasi pola makan dengan alasan yang sudah disebutkan sebelumnya, maka akan dipakai K=3. Skor jumlah klaster 3 tidak jauh berbeda dengan jumlah klaster 2, maka akan dipakai 3 klaster untuk segmentasi pola makan ini.

Selanjutnya kita akan mencari nilai k optimal dengan pemilihan atribut. Sama seperti sebelumnya yang akan menggunakan *score silhouette* sebagai pembanding dari setiap nilai k. Kode yang digunakan juga berbeda dengan sebelumnya, maka digunakan kode seperti di bawah ini.

```
print("\n=== DEFINISI FITUR UNTUK CLUSTERING ===")
features_for_clustering = [
    '4. Berapa kali dalam seminggu Anda mengonsumsi protein hewani (daging,
ikan, telur)?',
    '5. Berapa kali dalam seminggu Anda mengonsumsi protein nabati(sayur-
sayuran)?',
    '6. Seberapa sering Anda membeli makanan olahan/instan (fast food)?',
    '10. Apakah Anda pernah ikut serta dalam berbagai kegiatan pertanian di
komunitas/kampung?',
    '14. Apakah Anda ketergantungan terhadap pangan impor?',  #
Correlation: 0.529
    '17. Berapa persen dari pendapatan yang dialokasikan untuk pangan
berkelanjutan (organik/lokal)?'
]
reverse_items = [
    '6. Seberapa sering Anda membeli makanan olahan/instan (fast food)?',
    '14. Apakah Anda ketergantungan terhadap pangan impor?',
]

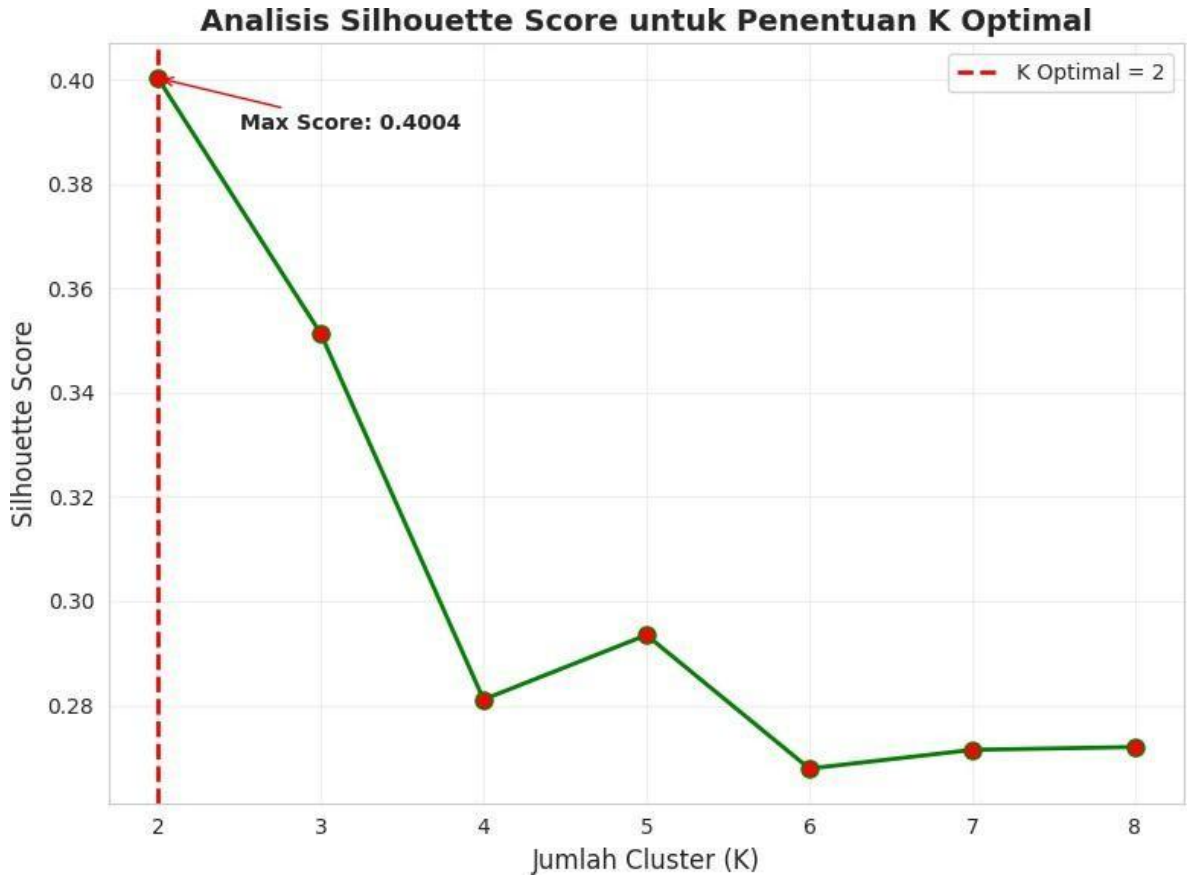
print("\n=== MEMULAI PREPROCESSING DENGAN FITUR TERPILIH ===")

df_normalized, scaler_used, final_features = preprocess_data(df_filtered,
features_for_clustering, reverse_items=reverse_items)

print(f"\n=== HASIL PREPROCESSING ===")
print(f"Fitur akhir yang digunakan: {len(final_features)}")
print(f"Sampel yang akan di-cluster: {len(df_normalized)}")
print(f"Shape data untuk clustering: {df_normalized.shape}")
```

Gambar 4. 8. *Code* untuk mendefinisikan atribut

Pertama kita perlu mendefinisikan atribut yang akan digunakan terlebih dahulu agar model dapat membaca beberapa atribut saja yang akan digunakan. Pendefinisian atribut di atas didasarkan pada hasil dari uji reliabilitas, validitas, dan CFA sebelumnya. Kemudian baru kita akan menghitung skor *silhouette* dengan kode yang masih sama pada gambar 4.6. Maka akan didapatkan hasil seperti pada gambar di bawah ini.



Gambar 4. 9. Grafik nilai k optimal setelah pemilihan atribut

Hasilnya dari grafik tersebut masih sama seperti sebelum memilih attribut, yaitu dengan nilai K optimal adalah 2. Tetapi terdapat perbedaan di skor *silhouette* yang dihasilkan. Dengan pemilihan attribut, skor yang dihasilkan lebih tinggi dari sebelumnya. Hal ini menandakan bahwa pemilihan atribut merupakan salah satu cara mengoptimalkan hasil segmentasi, berikut skor yang muncul dari hasil pemilihan atribut.

Tabel 4. 5. Tabel Perbandingan Silhouette Score setelah pemilihan atribut

Jumlah Klaster (K)	*Silhouette Score*
K=2	0.400360
K=3	0.351329
K=4	0.281143
K=5	0.293515
K=6	0.267898
K=7	0.271485
K=8	0.272036

Dari tabel di atas terlihat skor *silhouette* meningkat, yang menandakan bahwa segmentasi terdistribusi dengan lebih baik dari hasil sebelumnya. K==2 tetap menjadi yang paling optimal, tetapi untuk tujuan segmentasi kita akan memakai K=3 agar bisa terlihat beberapa bentuk pola konsumsi yang ada pada data. Skor K=2 dan K=3 tidak jauh berbeda, hanya berbeda 0.05 saja.

4.5. Hierarchical Clusteirng

Metode *hierarhical clusteirng* akan digunakan pertama untuk memberikan gambaran awal mengenai struktur kelompok data atau untuk melihat struktur hierarki data. Proses ini akan memetakan data ke dalam struktur pohon atau biasa disebut *dendogram* dan akan memudahkan dalam melihat struktur mengenai seberapa besar jarak antar klaster. *Hierarchical clusteirng* sebenarnya tidak membutuhkan penentuan jumlah klaster di awal dan kemudian hasilnya akan merepresentasikan bagaimana setiap data digabungkan ke dalam klaster yang lebih besar.

```
K_FINAL = 3

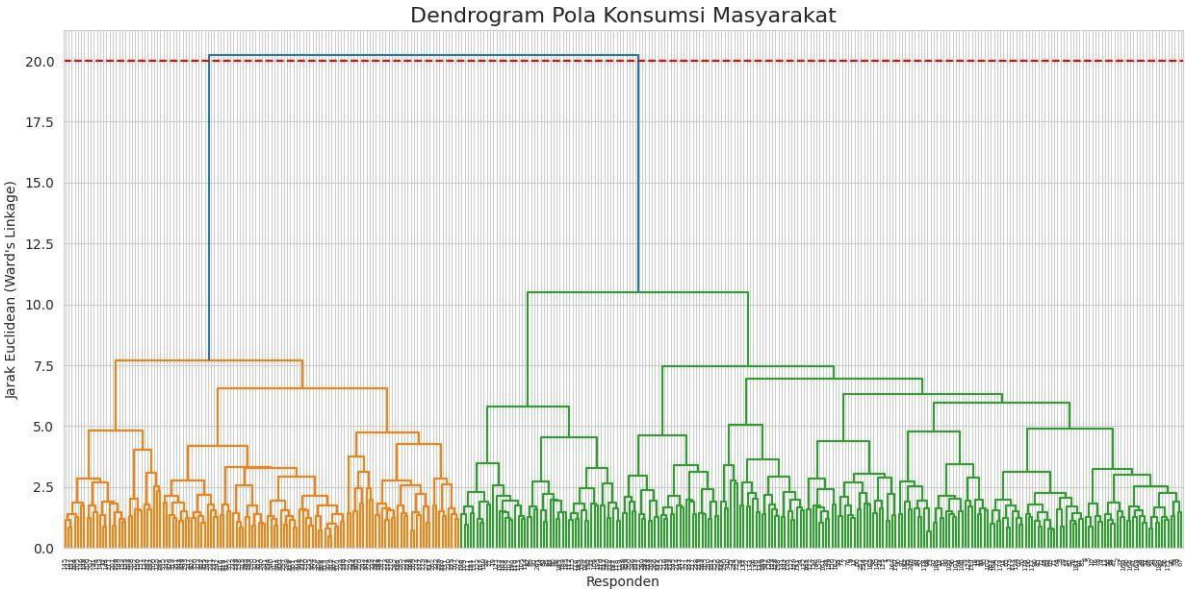
clustering_data = df_normalized.values

# 6.1 Hierarchical Clustering
print(f"\n--- Hierarchical Clustering ---")
hc_model = AgglomerativeClustering(n_clusters=K_FINAL,
linkage='ward')
hc_labels = hc_model.fit_predict(clustering_data)
df_filtered['Cluster_HC'] = hc_labels
hc_silhouette = silhouette_score(clustering_data, hc_labels)
print(f"Hierarchical Clustering - Silhouette Score:
{hc_silhouette:.6f}")

#Dendogram untuk pemetaan data
plt.figure(figsize=(15, 7))
plt.title("Dendrogram Pola Konsumsi Masyarakat", fontsize=16)
plt.xlabel("Responden")
plt.ylabel("Jarak Euclidean (Ward's Linkage)")
dendrogram = sch.dendrogram(sch.linkage(df_normalized,
method='ward'))
plt.axhline(y=20, color='r', linestyle='--') # Contoh garis
potong
plt.savefig('dendrogram_skripsi.png')
plt.show()
```

Gambar 4. 10. Implementasi Hierarchical Clustering

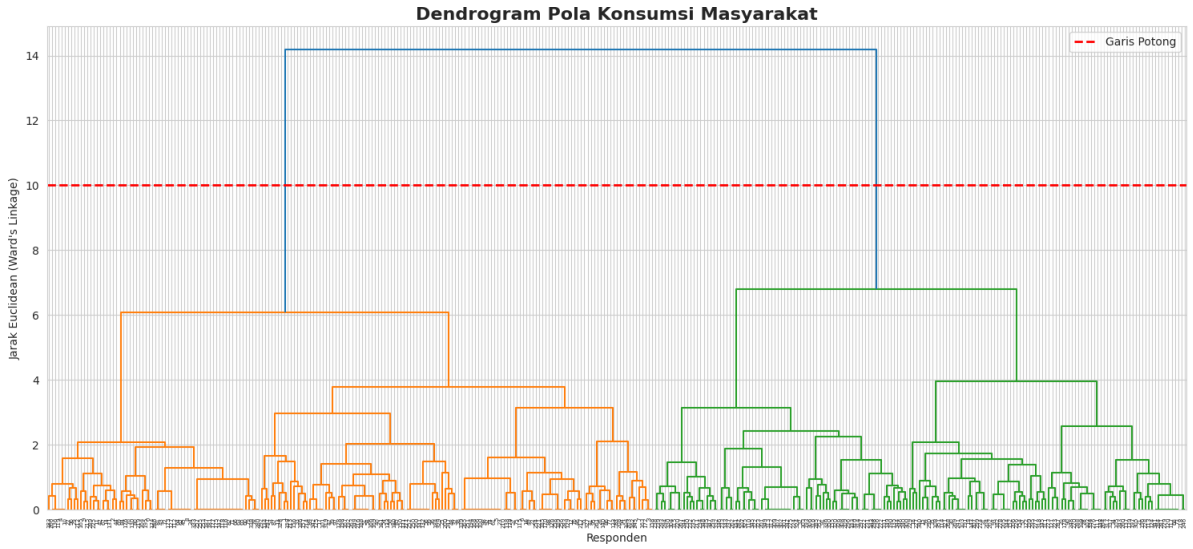
Mendifinisikan K=3, karena kita akan membentuk 3 klaster sebagai segmentasi pola konsumsi seperti alasan yang sudah dijelaskan sebelumnya. Kemudian kita masuk implementasi *hierarchical clustering* dengan memanggil metode *AgglomerativeClustering*, yang secara bertahap menggabungkan responden yang paling mirip hingga terbentuk sejumlah klaster yang diinginkan. Metode *linkage* yang digunakan dalam *clustering* ini adalah *Ward's Method*. Kemudian hasil klastering akan disimpan ke dalam kolom baru bernama *Cluster_HC* dengan memanggil fungsi *df_filtered['Cluster_HC']*. Dihitung juga nilai *silhouette score* untuk mengukur seberapa baik pemisah antar klaster yang terbentuk.



Gambar 4. 11. Dendogram Hasil Hierarchical Clustering

Dendogram di atas menggunakan jarak yang disebut *Euclidean*. Terdapat beberapa point penting dari grafik di atas, yang pertama terdapat sumbu Y yang menunjukkan jarak penggabungan antar kelompok. Jika semakin tinggi garis penggabungan, maka semakin jauh atau tidak mirip kelompok yang digabungkan. Kemudian garis putus berwarna merah dalam derajat horizontal y=20 menunjukkan batas pemotongan pohon untuk membentuk 3 klaster, di mana semua penggabungan yang berada di bawah garis akan membentuk klaster yang kita inginkan dan sumbu X menunjukkan seluruh responden yang diteliti. Dari hasil pemetaan klaster menggunakan *hieararchical clustering* didapatkan skor *silhouette* sebesar 0.1194. Skor ini cukup baik untuk segmentasi dan visualisasi dari *dendogram* dapat memperjelas pemetaan segmentasi.

Selanjutnya kita akan menggunakan algoritma pada data yang sudah dipilih atributnya sebelumnya sebagai gambaran jika terdapat perbedaan hasil skor *silhouette*. Algoritma yang dipakai masih sama seperti pada gambar 4.10, tetapi dengan data yang sudah dipilih atributnya. Hasilnya pada data yang diproses sebanyak 353 dan menggunakan 6 variabel dengan skor tertinggi. *Dendogram* yang dihasilkan seperti gambar di bawah ini.



Gambar 4. 12. Dendogram Hasil Hierarchical Clustering Setelah Pemilihan Atribut

Dari gambar 4.12 terlihat klaster-klaster yang terbentuk tidak sekompleks pada gambar 4.11, karena sebelumnya sudah memfilter atribut yang digunakan. Derajat pada garis Y juga hanya setengahnya yaitu y = 10, yang masih menunjukan klaster optimal sebanyak 2. Dari hasil klaster di atas, skor evaluasinya lebih tinggi dari sebelumnya. Skor *silhouette* menunjukan angka 0.339873, yang menunjukkan hasil distribusinya lebih baik dari sebelumnya.

4.6. K-Means Clustering

Pada tahap ini, proses pengelompokan data akan dilakukan menggunakan metode *k-means clustering* dengan jumlah klaster (K) yang sudah ditetapkan sebelumnya, yaitu 3 kelompok (K=3). Pemilihan 3 klaster mengacu pada kebutuhan data penelitian untuk memetakan pola konsumsi masyarakat yang kompleks dan dinamis ke dalam segmen yang berbeda. Proses pengelompokan ini dilakukan pada data yang sudah dinormalisasi, sehingga setiap variabel berada dalam skala yang sama. Kemudian, setiap responden akan diberi label klaster sesuai dengan hasil pengelompokan *k-means*.

```
# 6.2 K-Means Clustering
print("\nMenjalankan K-Means Clustering...")
kmeans_model = KMeans(n_clusters=K_FINAL, init='k-means++', random_state=42,
n_init=10)
df_filtered['Cluster_KMeans'] = kmeans_model.fit_predict(df_normalized)
kmeans_labels = kmeans_model.fit_predict(df_normalized)
kmeans_silhouette = silhouette_score(df_normalized,
df_filtered['Cluster_KMeans'])
print(f"K-Means Clustering - Silhouette Score: {kmeans_silhouette:.6f}")

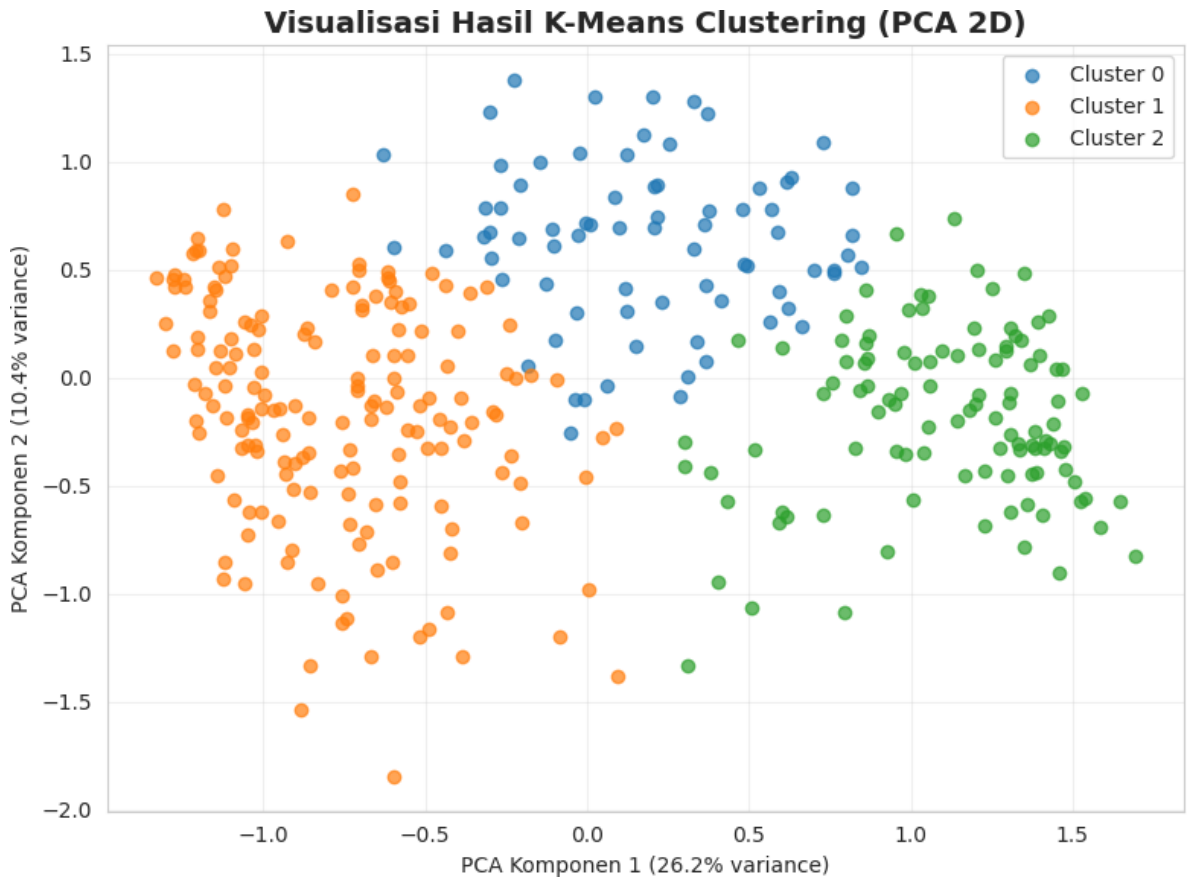
pca = PCA(n_components=2)
components = pca.fit_transform(df_normalized)

plt.figure(figsize=(8,6))
for cluster in range(K_FINAL):
    idx = df_filtered['Cluster_KMeans'] == cluster
    plt.scatter(
        components[idx, 0], components[idx, 1],
        label=f'Cluster {cluster+1}', alpha=0.7)

plt.title('Visualisasi Hasil K-Means Clustering (PCA 2D)', fontsize=14)
plt.xlabel('PCA Komponen 1')
plt.ylabel('PCA Komponen 2')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.savefig('kmeans_pca_visual.png')
plt.show()
```

Gambar 4. 13. Algoritma K-Means Clustering

Kode di atas adalah algoritma *k-means clustering* yang membagi data menjadi tiga klaster utama dan label klaster disimpan untuk setiap responden. Hasil klaster dari metode ini bisa divisualisasikan menggunakan *Principal Component Analysis* (PCA) ke dalam dua dimensi (2D) agar distribusi klaster dapat lebih mudah dipahami secara visual. Hasil transformasi PCA kemudian divisualisasikan dalam bentuk *scatter plot*, di mana setiap warna mewakili satu klaster yang terbentuk dari hasil *k-means*.



Gambar 4. 14. Visualisasi Hasil K-Means Clustering.

Gambar di atas memperlihatkan penyebaran data berdasarkan hasil *k-means clsutering*. Masing-masing titik pada grafik mewakili satu responden, sedangkan warna yang berbeda menunjukkan klaster yang berbeda. Terlihat persebaran klaster yang cukup jelas. Hal ini menunjukkan bahwa proses pengelompokan telah memisahkan kelompok data berdasarkan kemiripan pola konsumsi. Pemisahan antar klaster terlihat cukup tegas, meskipun ada beberapa area di mana titik dari klaster berbeda saling berdekatan. Hal inilah yang menunjukkan bahwa terdapat karakteristik yang mirip pada sebagian data, namun secara umum pembagian kelompok sudah cukup representatif. Nilai *silhouette score* pada metode ini menunjukkan angka 0.144363. Skor yang terbilang cukup tinggi dibandingkan skor pada metode sebelumnya. Hal ini dapat menunjukkan bahwa distribusi klaster lebih baik dari hasil sebelumnya. Analisis karakteristik atau interpretasi klaster memakai algoritma berikut.

```
def analisis_klaster_kmeans(df_original, df_processed, cluster_labels,
feature_names):
    # Buat dataframe untuk analisis
    df_analysis = df_processed.copy()
    df_analysis['Cluster'] = cluster_labels
    print(f"Data untuk analisis: {df_analysis.shape}")
    print(f"Distribusi klaster:")
    cluster_counts = pd.Series(cluster_labels).value_counts().sort_index()
    for cluster_id, count in cluster_counts.items():
        percentage = (count / len(cluster_labels)) * 100
        print(f"  Cluster {cluster_id}: {count} anggota ({percentage:.1f}%)")
    print(f"\nKarakteristik Rata-rata per Cluster (Data Normalized):")
    cluster_profile = df_analysis.groupby('Cluster')[feature_names].mean()
    # Tampilkan profil cluster
    print("\nProfil Cluster (5 fitur teratas per cluster):")
    for cluster_id in sorted(cluster_profile.index):
        print(f"\n--- Cluster {cluster_id} ---")
        cluster_data =
cluster_profile.loc[cluster_id].sort_values(ascending=False)
        top_features = cluster_data.head(5)
        for feature, value in top_features.items():
            print(f"  {feature[:50]}...: {value:.4f}")
    # Analisis pola untuk penamaan
    cluster_profiles = {}
    for cluster_id in sorted(cluster_profile.index):
        profile = cluster_profile.loc[cluster_id]
        top_features =
profile.sort_values(ascending=False).head(3).index.tolist()
        cluster_profiles[cluster_id] = top_features
    cluster_names = {
        0: "Peduli Kesehatan & Lingkungan",
        1: "Pro Pangan Lokal",
        2: "Konvensional & Pragmatis"}
    print(f"\n=== PENAMAAN KLASTER ===")
    for cluster_id, name in cluster_names.items():
        count = cluster_counts.get(cluster_id, 0)
        percentage = (count / len(cluster_labels)) * 100
        print(f"Cluster {cluster_id} -> '{name}'")
        print(f"  Anggota: {count} responden ({percentage:.1f}%)")
        print(f"  Karakteristik dominan: {cluster_profiles.get(cluster_id,
[])}")
    return cluster_names, cluster_profile
cluster_names_kmeans, cluster_profile_kmeans = analisis_klaster_kmeans(
    df_filtered, df_normalized, kmeans_labels, df_normalized.columns)
```

Gambar 4. 15. Code untuk Analisis Klaster K-Means

Hasil algoritma di atas akan menunjukkan interpretasi klaster yang terbentuk dari *k-means clustering*. Algoritma di atas memiliki fungsi *value_counts*() yang dapat menghitung jumlah anggota per klaster. Kemudian menggunakan fungsi *groupby()* dan *mean()* untuk menghitung rata-rata atribut per klaster. Dan fungsi *sort_values()* yang digunakan untuk mengurutkan atribut paling dominan dari setiap klaster yang terbentuk.

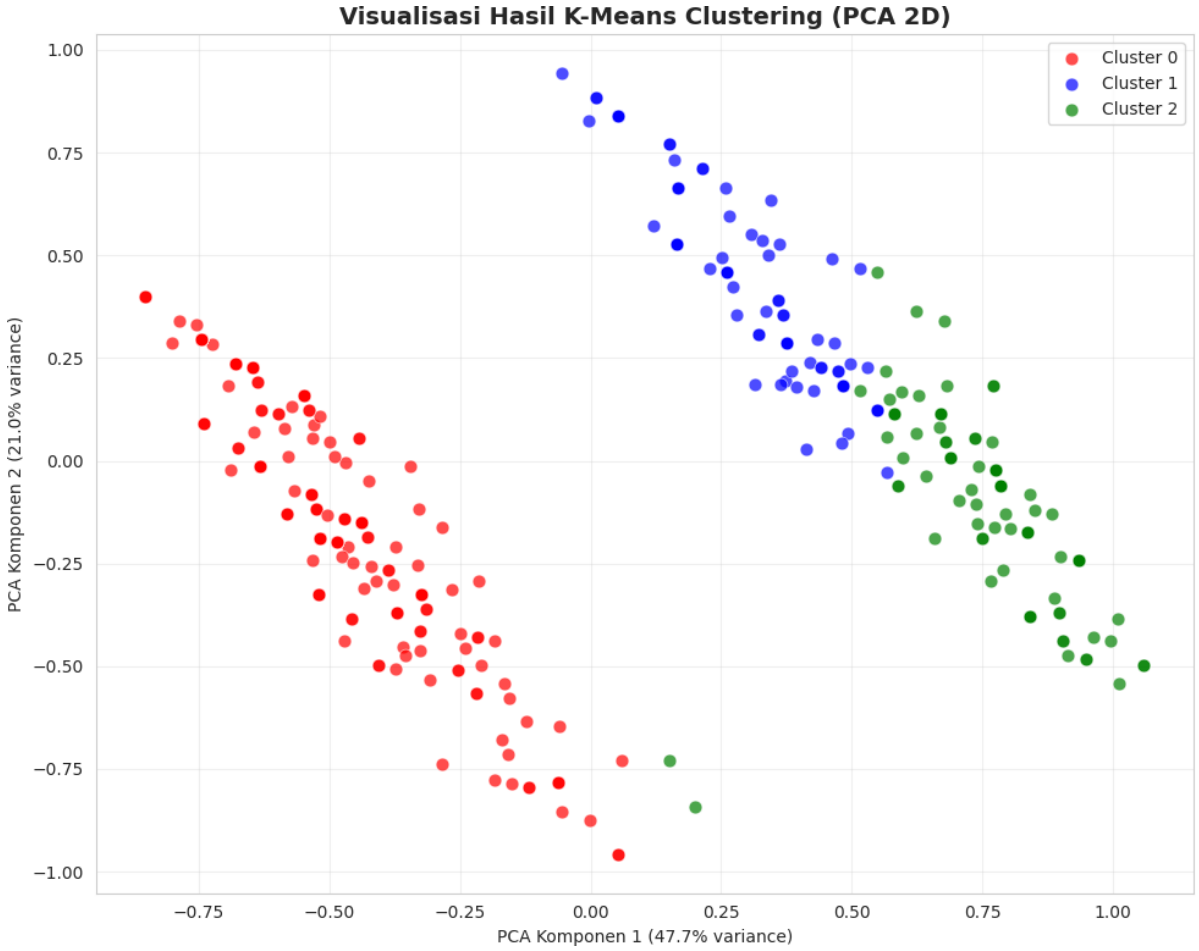
Klaster yang terbentuk ada 3, karena kita mendefinisikan K=3, dimulai dengan klaster 0. Hasil dari klaster 0 dengan nama "Konvensional & Pragmatis" menunjukkan jumlah anggota sebanyak 75 data atau sebanyak 21.2% dari seluruh data. Kemudian terdapat karakteristik klaster 0 yang terbentuk dengan menunjukkan 5 atribut dominan dari profilisasi klaster, yaitu atribut 24 dengan skor 0.9333, atribut 21 dengan skor 0.9233, atribut 13 dengan skor 0.9100, atribut 5 dengan skor 0.9067, dan atribut nomor 3 dengan skor 0.8833. Nomor atribut mengacu pada nomor pertanyaan pada tabel 3.1.

Kemudian terbentuk klaster 1 dengan nama klaster "Peduli Kesehatan & Lingkungan" dengan jumlah anggota sebanyak 171 data atau 48.4% dari total keseluruhan data. Kemudian dari profil klaster terdapat 5 atribut teratas yang mempengaruhi klaster tersebut, yaitu atribut nomor 21 dengan skor 0.8670, atribut 13 dengan skor 0.8538, atribut nomor 4 dengan skor 0.8363, atribut nomor 1 dengan skor 0.8324, dan atribut nomor 3 dengan skor 0.8260. Skor yang dihasilkan relatif tinggi, yang berarti menandakan ke 5 atribut tersebut yang berpengaruh terhadap terbentuknya klaster 1. Klaster ini berpotensi menjadi model bagi segmentasi lain karena menunjukkan keterlibatan aktif dalam gaya hidup sehat. Oleh karena itu, responden dalam klaster ini dapat dijadikan target utama sebagai agen perubahan atau duta pangan sehat di komunitasnya.

Klaster terakhir yang terbentuk adalah klaster 2 yang memiliki nama "Pro Pangan Lokal". Anggota dari klaster 2 ini sebanyak 107 data atau 30.3% dari total data yang diproses. Karakteristik dominan yang terbentuk terdapat 5 atribut yang paling berpengaruh, yaitu atribut nomor 23 dengan skor 0.9533, atribut nomor 10 dengan skor 0.9346, atribut nomor 24 dengan skor 0.9252, atribut nomor 27 dengan skor 0.8598, dan atribut nomor 7 dengan skor 0.8505. Nilainya cukup baik dari setiap atribut untuk membentuk klaster 2 tersebut. Klaster ini memiliki posisi strategis sebagai penghubung antara konsumen konvensional dan konsumen peduli kesehatan. Dengan intervensi edukatif yang tepat, kelompok ini berpotensi diarahkan untuk mengadopsi pola makan yang lebih sehat dan berkelanjutan tanpa meninggalkan kecintaan terhadap produk lokal.

Klaster di atas terbentuk dengan memakai seluruh atribut yang bisa diproses. Selanjutnya adalah memproses algoritma yang sama dengan data yang sudah dipilih atributnya

pada tahap *pre-processing*. Hasilnya juga akan divisualisasikan dalam bentuk PCA 2 dimensi di bawah ini.



Gambar 4. 16. Visualisasi Hasil K-Means Clustering dengan Pemilihan Atribut

Hasil visualisasi di atas menunjukkan perbedaan dari hasil sebelumnya. Pemilihan atribut ini membantu algoritma dalam memisahkan segmen secara lebih tegas. Jika sebelumnya klaster tersebar dan masih samar, maka dalam visualisasi setelah pemilihan atribut, klaster secara lebih tertata terbagi ke dalam 3 klaster, walaupun belum sempurna karena masih terdapat beberapa data yang masih samar terhadap segmennya. Setelah mendapatkan visualisasi klaster, selanjutnya akan kita interpretasikan klaster yang terbentuk dengan algoritma yang masih sama pada gambar 4.15.

Klaster pertama yang terbentuk adalah klaster 0 yang memiliki nama “Konvensional & Pragmatis“ dengan jumlah anggota sebanyak 186 data atau 52.7% dari total 353 data yang diproses. Karakteristik yang berpengaruh pada klaster ini adalah atribut nomor 4 dengan skor 0.8423, atribut nomor 14 dengan skor 0.8038, atribut nomor 5 dengan skor 0.7921, atribut

nomor 17 dengan skor 0.5995, atribut nomor 6 dengan skor 0.5681, dan atribut nomor 10 tidak berpengaruh apapun dengan menghasilkan skor 0.0000. Dari karakteristik pengaruh atribut dii atas dapat disimpulkan beberapa atribut yang sangat berpengaruh dalam pembentuka klaster 0 ini. Nomor atribut didasarkan pada nomor pertanyaan dalam table 3.1.

Klaster selanjutnya yang terbentuk adalah klaster 1 dengan nama "Peduli Kesehatan & Lingkungan" dengan jumlah anggota sebanyak 79 data atau sebanyak 22.4% dari 353 data yang diproses. Karakteristik pada klaster 1 ini dibentuk dari beberapa klaster yang berpengaruh, diantaranya atribut nomor 10 dengan nilai sempurna 1.0000, atribut nomor 5 dengan skor 0.8312, atribut nomor 14 dengan skor 0.8038, atribut nomor 4 dengan skor 0.7679, atribut nomor 17 dengan skor 0.6519, dan atribut nomor 6 dengan skor 0.5949. Atribut dengan skor sempurna berarti sangat berpengaruh dalam pembentukan klaster 1 ini dan atribut lainnya sebagai pendukung dari terbentuknya klaster 1. Nomor atribut didasarkan pada nomor pertanyaan dalam table 3.1.

Klaster 2 adalah klaster terakhir yang terbentuk dengan label nama "Pro Pangan Lokal". Anggota dari klaster 2 ini sebanyak 88 data atau 24.9% dari total 353. Klaster 2 ini dipengaruhi oleh beberapa atribut yang berperan dalam pembentukan kalster ini, diantaranya adalah klaster nomor 10 dengan skor 0.9773, atribut nomor 5 dengan skor 0.6364, atribut nomor 4 dengan skor 0.6250, atribut nomor 17 dengan skor 0.3864, atribut nomor 14 dengan skor 0.2642, dan atribut nomor 6 dengan skor 0.1364. atribut yang mendekati 1 atau bahkan satu berarti semakinn berpengaruh juga atribut tersebut terhadap pembentukan klaster. Nomor atribut didasarkan pada nomor pertanyaan dalam table 3.1.

Hasil evaluasi algoritma *k-means* sebelum pemilihan atribut adalah 0.144363, yang menunjukkan algoritma *k-means* belum mampu membagi data ke dalam segmen secara tegas. Artinya masih banyak data yang samar atau terdapat probabilitas dari setiap data untuk masuk ke dalam setiap klaster. Hasil evaluasi algoritma setelah pemilihan atribut adalah 0.351329, yang menunjukan algoritma sudah membagi data ke dalam klaster secara tegas karena atribut yang diambil adalah yang berpengaruh terhadap pembentukan klaster.

4.7. *Gaussian Mixture Model* (GMM)

Gaussian Mixture Model atau GMM digunakan sebagai metode terakhir untuk klastering. Jumlah klaster yang sebelumnya sudah didefinisikan dengan K=3 akan digunakan juga di dalam GMM sebagai pembentukan 3 klaster. Berikut adalah kode untuk menjalankan GMM.

```
K_FINAL = 3
gmm_model = GaussianMixture(n_components=K_FINAL, covariance_type='full',
random_state=42, max_iter=200, n_init=5)
gmm_labels = gmm_model.fit_predict(df_normalized)
gmm_probs = gmm_model.predict_proba(df_normalized)

# Tambahkan hasil clustering GMM ke dataframe utama yang difilter
df_filtered['Cluster_GMM'] = gmm_labels

# Tambahkan probabilitas keanggotaan untuk setiap klaster
for i in range(K_FINAL):
    df_filtered[f'GMM_Prob_{i}'] = gmm_probs[:, i]

# Hitung silhouette score untuk GMM
gmm_silhouette = silhouette_score(df_normalized, gmm_labels)
print(f"GMM Clustering - Silhouette Score: {gmm_silhouette:.4f}")
print("Implementasi GMM dengan semua fitur selesai.")
```

Gambar 4. 17. *Code* Untuk Menjalankan Algoritma GMM

Pada kode di atas, klaster didefinisikan dengan K=3. Terdapat parameter *covariance_type='diag'* yang memungkinkan model menangkap variabilitas data secara optimal. Terdapat pelatihan model terhadap data yang sudah di normalisasi dengan *fit_predict*. Proses ini akan secara otomatis memberikan label klaster (angka 0, 1, atau 2) kepada setiap data responden, yang kemudian akan disimpan di dalam kolom baru bernama *Cluster_GMM*.

Salah satu keunggulan GMM adalah kemampuannya untuk melakukan *soft-clustering*, untuk memperoleh nilai probabilitas keanggotaan setiap responden terhadap masing-masing klaster dengan menggunakan perintah *gmm_model.predict_proba()*. Hal ini penting karena GMM tidak hanya memberikan label klaster, tetapi juga dapat memperlihatkan seberapa besar kemungkinan setiap data masuk ke dalam klaster tertentu. Data yang sudah dibersihkan dan dinormalisasi kemudian dimasukkan ke dalam model dengan proses *gmm_labels = gmm_model.fit_predict(df_normalized*. Model menggunakan *fit* untuk menemukan tiga kelompok elips atau distribusi Gaussian yang paling cocok untuk menggambarkan persebaran data dan menggunakan *predict* untuk memberikan label klaster berupa angka 0, 1, dan 2 kepada setiap responden dengan berdasarkan kedekatan probabilitas responden tersebut terhadap pusat setiap klaster.

```
gmm_probs = gmm_model.predict_proba(df_normalized)
for i in range(K_FINAL):
    df_filtered[f'GMM_Prob_{i}'] = gmm_probs[:, i]
```

Gambar 4. 18. *Code* Menghitung Probabilitas GMM

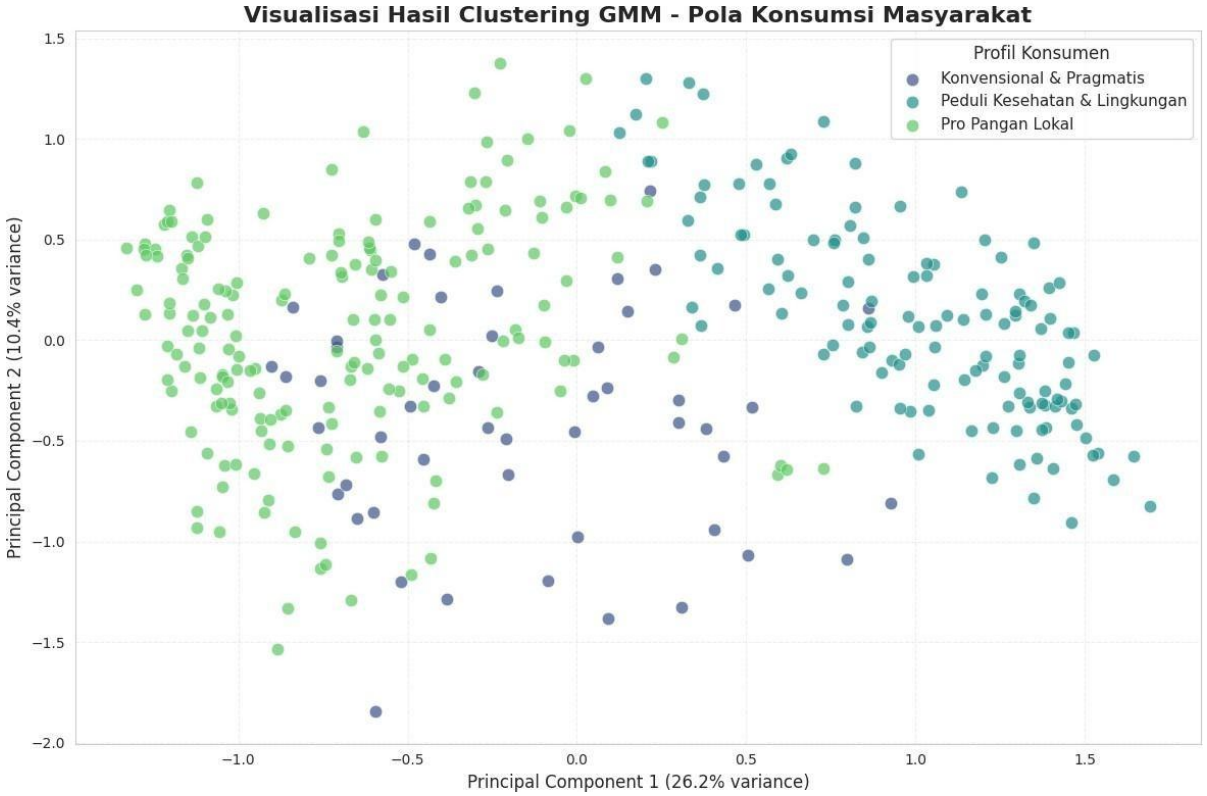
Sesuai dengan keunggulannya melakukan *soft clustering* GMM akan menggunakan kode di atas ini untuk menghitung tingkat keyakinan atau probabilitas untuk setiap responden terhadap semua klaster. Dalam data pola makan yang dinamis, seorang responden tidak hanya masuk di dalam satu klaster, misalnya klaster 0, tetapi model bisa memberitahu probabilitas bahwa responden tersebut memiliki probabilitas 85% masuk ke dalam klaster 0, 10% ke klaster 1, dan 5% ke dalam klaster 2. Informasi ini akan menunjukkan bahwa data tersebut lebih kuat ke dalam satu klaster dan memiliki probabilitas terhadap klaster lain yang kecil. Kolom probabilitas ini akan disimpan di dalam kolom baru bernama *GMM_Prob_0*, *GMM_Prob_1*, dan *GMM_Prob_2*.

Hasil dari algoritma GMM menghasilkan 3 klaster. Klaster pertama adalah klaster 0 dengan label nama "Konvensional & Pragmatis" dengan jumlah anggota yang sedikit, yaitu 51 responden atau 14.4% dari total data yang diproses sebanyak 353. Klaster 0 terbentuk dengan pengaruh atribut nomor 27 dengan skor 1.0000, atribut nomor 21 dengan skor 0.8480, atribut nomor 1 dengan skor 0.235, atribut nomor 3 dengan skor 0.8137, dan atribut nomor 13 dengan skor 0.8137. Atribut yang disebutkan adalah yang teratas dalam mempengaruhi pembentukan klaster 0.

Kemudian terdapat klaster 1 dengan label nama "Peduli Kesehatan & Lingkungan" yang memiliki anggota sebanyak 125 responden atau 35.4% dari 353 data yang diproses. Karakteristik dari klaster 1 dipengaruhi oleh 5 atribut teratas, yaitu atribut nomor 10 dengan skor 1.0000, atribut nomor 23 dengan skor 1.0000, atribut nomor 24 dengan skor 1.0000, atribut nomor 7 dengan skor 0.8640, dan atribut nomor 13 dengan skor 0.8600. Terdapat atribut dengan skor sempurna yang menandakan kesesuaian atribut dalam pembentukan klaster 1.

Klaster terakhir yang terbentuk adalah klaster 2 dengan label nama "Pro Pangan Lokal" yang memiliki anggota sebanyak 177 responden atau sebanyak 50.1% dari total 353 data yang diproses. Pembentukan klaster 2 ini dipengaruhi oleh beberapa atribut, berikut 5 teratas yang mempengaruhi pembentukan klaster ini. Yang pertama terdapat atribut nomor 21 dengan skor 0.8927, atribut nomor 13 dengan skor 0.8785, atribut nomor 4 dengan skor 0.8475, atribut nomor 1 dengan skor 0.8380, dan atribut nomor 3 dengan skor 0.8376. Skor yang terlihat masih dianggap bagus untuk pengaruh terhadap pembentukan klaster karena nilainya hampir

mendekati 1. Semua nomor atribut yang disebutkan di atas mengacu pada nomor pertanyaan yang tertera pada tabel 3.1.



Gambar 4. 19. Visualisasi PCA 2D GMM *Clustering*

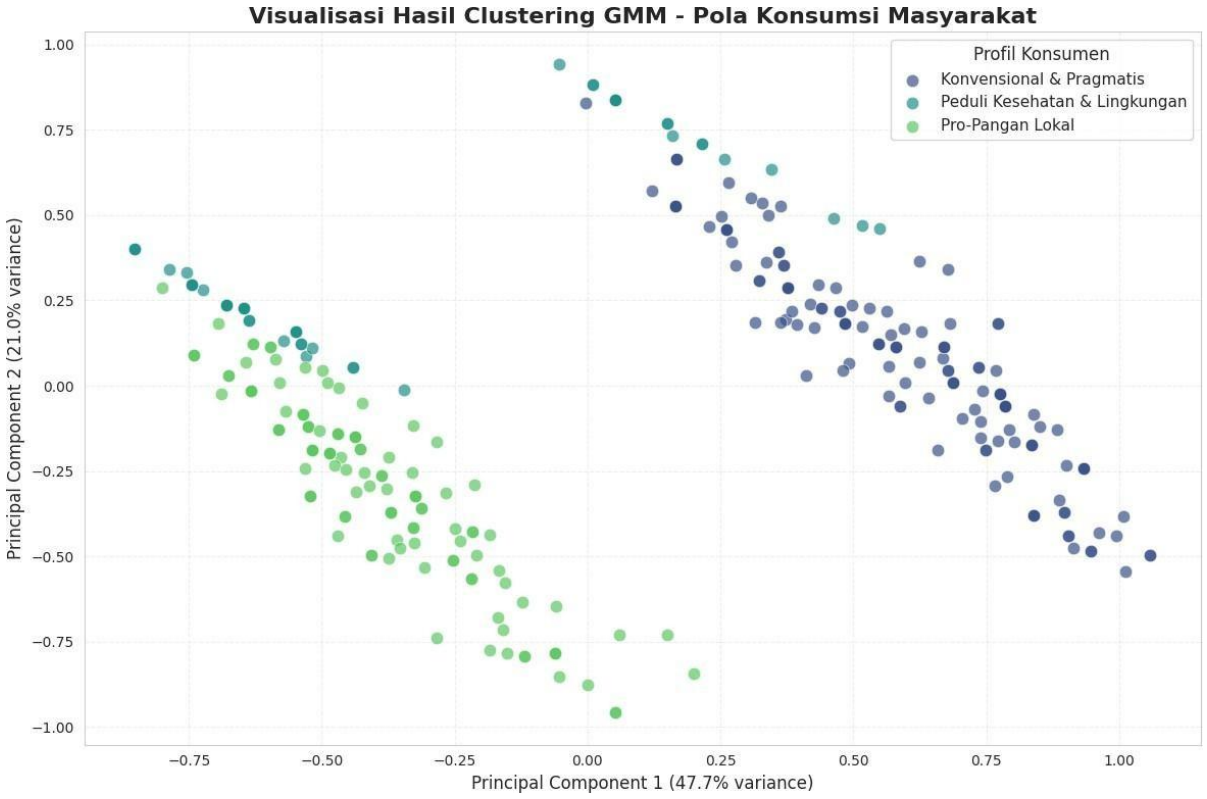
Hasil dari visualisasi dari GMM menunjukkan persebaran data yang acak. Karena data pola konsumsi yang kompleks memungkinkan data bisa masuk ke dalam beberapa klaster dengan persentase tersendiri atau dalam GMM bisa disebut probabilitas. Hasil dari probabilitas dari setiap data. Dan juga setiap atribut dapat probabilitas untuk membentuk setiap klaster. Hasil probabilitas dari atribut untuk membentuk klaster akan disajikan dalam lamprian.

Kemudian terdapat hasil *clustering* GMM dengan pemilihan atribut yang sudah ditetapkan dengan menggunakan algoritma yang sama. Maka dihasilkan klaster yang pertama, yaitu klaster 0 yang diberi label "Konvensional & Pragmatis" dengan jumlah anggota 144 klaster atau 40.8% dari 353 data yang diproses. Pembentukan klaster dipengaruhi oleh atribut nomor 10 dengan skor 1.0000, atribut nomor 5 dengan skor 0.6921, dan atribut nomor 14 dengan skor 0.0.6644. Ketiga atribut tersebut memiliki nilai yang cukup tinggi, yang berarti berpengaruh dalam pembentukan klaster 0.

Klaster 1 yang selanjutnya diberi label "Peduli Kesehatan & Lingkungan" dengan jumlah anggota 70 responden atau setara 19.8% dari 353 data yang diproses. Karakteristik atribut dominan dari klaster 1 adalah atribut nomor 5 dengan skor 1.0000, atribut nomor 17

dengan skor 1.0000, atribut nomor 4 dengan skor 0.9238, dan atribut nomor 14 dengan skor 0.8429. Berdasarkan skor di atas, keempat atribut ini sangat berpengaruh dalam pembentukan klaster 1.

Klaster 2 diberi label “Pro Pangan Lokal” dengan jumlah anggota 139 responden atau 39.4% dari 353 data yang diproses. Atribut yang berpengaruh pada klaster 2 ini antara lain, atribut nomor 4 dengan skor 0.8058, atribut nomor 14 dengan skor 0.7716, atribut nomor 5 dengan skor 0.7146. Ketiga atribut dengan nilai tertinggi yang berarti berpengaruh terhadap pembentukan klaster 2.



Gambar 4. 20. Visualisasi PCA 2D GMM Clustering Setelah Pemilihan Atribut

Dari hasil visualisasi di atas menunjukkan bahwa ketiga klaster cukup terpisah dengan baik, berbeda dari sebelum pemilihan atribut. Artinya setiap kelompok memiliki karakteristik konsumsi yang berbeda dan dapat diidentifikasikan dengan cukup jelas. Sumbu horizontal dan vertikal mewakili kombinasi variabel utama yang paling membedakan pola konsumsi antar responden.

4.8. Interpretasi Klaster

Klaster yang terbentuk dari *K-Means* dan GMM adalah 3 klaster yang semuanya merepresentasikan setiap pola makan dari responden dan karakteristik. Pada interpretasi klaster

ini difokuskan setelah pemilihan atribut. Dan hasil akhir dari GMM yang merupakan penyempurnaan dari hasil klaster sebelumnya, karena GMM mampu melakukan *soft clustering* dengan memunculkan probabilitas keanggotaan. GMM dapat memberikan gambaran yang realistis terkait pola konsumsi masyarakat yang fleksibel. Maka akan diperoleh probabilitas setiap pembentukan klaster pada tabel di bawah ini.

Tabel 4. 6. Tabel Karakteristik Klaster GMM

Nomor Atribut	Klaster		
	0	1	2
4	0.66	0.92	0.81
5	0.69	1.00	0.71
6	0.32	0.63	0.53
10	1.00	0.30	0.00
14	0.49	0.84	0.77
17	0.44	1.00	0.46

Dari tabel di atas dengan pemilihan atribut pada *pre-processing* menunjukkan bahwa pembentukan klaster 0 pada GMM dibentuk dengan 3 atribut teratas, yaitu atribut nomor 10, 5, dan 4. Berbeda dengan klaster *k-means* yang dibentuk dengan atribut 4, 14, dan 5. Pada klaster 0 *k-means*, atribut nomor 10 menunjukkan nilai 0.00 yang berarti tidak berpengaruh sama sekali, tetapi dalam klaster 0 GMM nilai atribut 10 mencapai nilai 1.00 yang berarti klaster tersebut terbentuk dari nomor atribut tersebut.

Pada klaster 1 GMM terbentuk berdasarkan probabilitas atribut nomor 5, 17, 4, dan 14. Dengan atribut nomor 5 dan 17 mencapai nilai 1.00 yang berarti dasar klaster tersebut terbentuk dari nomor atribut tersebut. Berbeda dengan klaster 1 *k-means* yang terbentuk berdasarkan atribut nomor 10, 5, 14, dan 4. Dengan atribut nomor 10 bernilai 1.00, yang menunjukkan bahwa dasar dari terbentuknya klaster tersebut dari atribut nomor 10. Hal ini memiliki arti bahwa GMM menyempurnakan klaster dari sebelumnya yang menunjukkan atribut nomor 10 lebih baik masuk ke dalam klaster 0.

Klaster selanjutnya adalah klaster 2 yang terbentuk pada GMM. Dibentuk berdasarkan dengan atribut nomor 4, 14, dan 5 sebagai nilai tertinggi probabilitas. Berbeda dengan *k-means*, klaster 2 algoritma ini adalah atribut nomor 10, 5, dan 4 sebagai nilai tertingginya. Atribut nomor 14 pada *k-means* tidak berdampak signifikan dengan nilai 0.2642, tetapi pada algoritma

GMM diperbaiki dengan klaster 14 berdampak signifikan pada klaster 2 dengan nilai probabilitas 0.77.

Pemberian label klaster didasarkan dari hasil penelitian sebelumnya, yang disesuaikan dengan kebutuhan segmentasi pada penelitian ini. Label pertama adalah "Pro-Pangan Lokal", yang menunjukkan kebiasaan responden terhadap pangan lokal. Penamaan klaster ini berdasarkan hasil penelitian yang dilakukan oleh Piracci et al. (2023), yang menemukan segmen "*Sustainability Focused*". Segmen yang ditemukan ini lebih menyukai pangan yang konsisten terhadap keberlanjutan. Relevansinya adalah dengan mendukung pangan lokal, artinya responden mendukung keberlanjutan dari dalam negeri, karena biasanya pemerintah memiliki kebijakan yang mendukung pangan lokal sebagai pendukung keberlanjutan. Kemudian, nama ini juga diambil dari hasil penelitian O'Neill et al. (2023), yang menemukan segmen "*Food Appreciators*". Segmen yang ditemukan ini mengarah kepada pemilihan dan penggunaan sumber daya pangan yang cermat dan menghindari pemborosan.

Label klaster yang kedua adalah "Konvensional & Pragmatis". Konvensional berarti mengacu pada pola makan sesuai dengan kebiasaan, tradisi, atau norma masyarakat pada umumnya. Dan pragmatis berarti pendekatan yang berfokus pada hasil praktis dan solusi yang paling efektif atau efisien. Penamaan label ini didasarkan dengan hasil penelitian Gazdecki et al. (2021), yang menemukan segmen "*Reluctant*". Segmen yang ditemukan ini mengacu pada kebiasaan pola makan dengan sikap negatif terhadap gagasan konsumsi keberlanjutan, sehingga pola makan terus mengikuti alur kebiasaan masyarakat pada umumnya dan mencari jalan praktis terhadap pangan.

Label klaster yang terakhir adalah "Peduli Kesehatan & Lingkungan". Label ini didapatkan berdasarkan hasil penelitian Piracci et al. (2023), yang menemukan segmen "*Naturalness and Health Driven*". Segmen yang ditemukan ini mementingkan kealamian dan manfaat positif produk makanan terhadap kesehatan pribadi. Jadi segmen ini mengutamakan makanan organik yang bebas dari pengaruh pupuk buatan demi menjaga kesehatan tubuh dan dampak luas lingkungan.

Tabel 4. 7. Tabel Profil Klaster

Atribut	Klaster GMM		
	0	1	2
Usia	31.29	38.35	31.84

Konsumsi Makanan Pokok	3,22	3,84	3,31
Pengeluaran Untuk Pangan	2.09	2.42	2.15
Pangan Lokal Dibanding Pangan Impor	4.32	4.68	4.20
Konsumsi Protein Hewan	1.99	2.77	2.41
Konsumsi Protein Nabati	2.07	3	2.14
Membeli Makanan Instan	3.04	2.1	2.39
Kesadaran Dampak Lingkungan dari Pola Konsumsi	4.39	4.51	4.10
Persentase Produk Organik/Lokal yang Dibeli (per 1 bulan)	2.49	3.3	2.58
Menghindari *Food Waste*	3.99	3.57	3.50
Partisipasi Kegiatan Pertanian	1	0.3	0
Akses Makanan Bergizi	4.38	4.82	4.33
Ketergantungan Pangan Impor	3.05	1.62	1.91
Sensitivitas terhadap Harga Bahan Pangan	0.93	0.75	0.81
Alokasi Pangan Berkelanjutan	1.88	3	1.91

Membaca Label Nutrisi Sebelum Membeli	3.36	3.31	2.70
Konsumsi Buah dan Sayuran	1.82	2.72	1.96
Check-Up Terkait Pola Makan atau Gizi	0.83	0.38	0.25
Partisipasi Program Penyuluhan Terkait Gizi/Kesehatan	0.86	0.42	0.22

Klaster pertama yaitu "Konvensional & Pragmatis" didominasi semua usia mulai dari 17-45 tahun yang memiliki kecenderungan memilih konsumsi yang praktis dan sederhana. Responden pada klaster ini memiliki frekuensi konsumsi makanan pokok dan protein yang cukup, namun konsumsi makanan cepat saji dan makanan instan masih relatif tinggi. Kemudian dari segi pengeluaran untuk pangan sehat atau organik masih relatif rendah, dan responden pada klaster ini tidak terlalu sering dalam membeli produk pangan lokal. Klaster ini memiliki pengetahuan terhadap isu lingkungan yang cukup, tetapi belum menjadikannya prioritas utama dalam pola makan sehari-hari. Pada klaster ini juga cenderung tidak membaca label gizi dalam membeli makanan, jarang untuk mengikuti penyuluhan gizi, dan tidak terlalu rutin dalam melakukan *check-up* kesehatan. Klaster ini relatif lebih sensitif terhadap harga pangan sehat dan organik, sehingga lebih memilih makanan yang terjangkau dan mudah diakses.

Dari karakteristik yang terbentuk, rekomendasi kebijakan yang bisa diambil yaitu pemangku kepentingan perlu meningkatkan edukasi terkait pentingnya konsumsi pangan sehat dan berkelanjutan melalui strategi yang dapat mudah dipahami dan diaplikasikan. Kemudian pemangku kepentingan dapat melakukan subsidi atau insentif untuk pangan sehat/organik yang terjangkau. Dan juga menerapkan program edukasi berbasis komunitas dan pendekatan digital bisa memperluas jangkauan informasi ke klaster ini. Hal di atas mungkin bisa membantu klaster ini berpindah dari pola konsumsi yang praktis atau instan ke pola makan yang lebih baik.

Klaster selanjutnya adalah "Peduli Kesehatan dan Lingkungan" yang berisi responden dengan dominasi di usia 31 – 45 Tahun dengan menerapkan pola makan sehat. Klaster ini

paling sering mengonsumsi protein baik hewani/nabati, buah, dan sayur segar. Pada klaster ini yang paling tinggi pengeluaran bulanan untuk pangan sehat, organik, dan lokal dibandingkan klaster lain. Dan juga sering membeli produk pangan lokal dan aktif membaca label gizi sebelum membeli makanan, serta sangat memperhatikan komposisi makanan sehari-hari. Klaster ini juga sadar terhadap isu lingkungan dan kesehatan yang sangat tinggi. Responden pada klaster ini juga paling jarang dalam membeli makanan instan. Selain itu, pada klaster ini juga rutin melakukan pemeriksaan kesehatan dan mengikuti penyuluhan gizi atau program edukasi terkait pola makan sehat, klaster ini memiliki kecenderungan yang tidak terlalu sensitif terhadap pangan sehat. Kualitas dan keberlanjutan pangan menjadi prioritas dalam mengonsumsi makanan sehari-hari.

Klaster ini adalah panutan dalam mengonsumsi makanan sehat dan ramah lingkungan, kebijakan yang tepat untuk anggota klaster ini adalah dengan mempertahankan dan memperluas akses terhadap produk pangan sehat/lokal, dan juga menjadikan mereka agen perubahan. Pemangku kepentingan juga bisa melakukan kebijakan memfasilitasi akses pada produk pangan organik/lokal, kemudian bisa juga membuat komunitas "duta pangan sehat" dan melibatkan klaster ini sebagai bagian dari komunitas untuk mensosialisasikan gaya hidup sehat ke masyarakat luas.

Klaster yang terakhir adalah "Pro-Pangan Lokal" yang didominasi dengan usia muda pada 17-35 Tahun yang sangat mendukung pangan lokal. Dan terdapat juga usia dari 36-45 tahun yang terdapat pada klaster ini. Pada klaster ini konsumsi makanan berprotein, serta buah dan sayur cukup rutin, namun tidak setinggi klaster "Peduli kesehatan & Lingkungan". Klaster ini juga memiliki pengeluaran untuk pangan lokal dan organik terdapat pada tingkatan yang tidak terlalu tinggi dan tidak terlalu rendah. Dan juga frekuensi pembelian produk pangan lokal yang tinggi, namun konsumsi produk organik dan pangan sehat belum seintens klaster kesehatan. Dari sisi perilaku kesehatan, klaster ini cukup sering dalam membaca label gizi sebelum membeli, tetapi belum sesering klaster kesehatan. Sensitivitas terhadap harga pangan sehat di klaster ini juga tidak terlalu tinggi dan tidak terlalu rendah. Responden bersedia membayar lebih tinggi untuk produk lokal, tetapi tidak selalu pangan organik atau sehat. Klaster ini juga cukup peduli terhadap isu lingkungan, meskipun dalam praktiknya masih perlu ditingkatkan.

Sara kebijakan untuk klaster "Pro-Pangan Lokal" adalah dengan mendorong integrasi promosi pangan lokal dan gaya hidup sehat. Para pemangku kepentingan juga dapat meningkatkan edukasi agar klaster ini tidak hanya bangga terhadap produk pangan lokal, tetapi juga dapat edukasi pentingnya kesehatan, keamanan pangan, dan keberlanjutan. Pemangku

kepentingan juga dapat memperluas jangkauan pasar produk lokal dan dapat meningkatkan nilai tambah pada produk lokal melalui pelabelan sehat atau organik.

Tabel 4. 8. Tabel Hasil Ketiga Metode

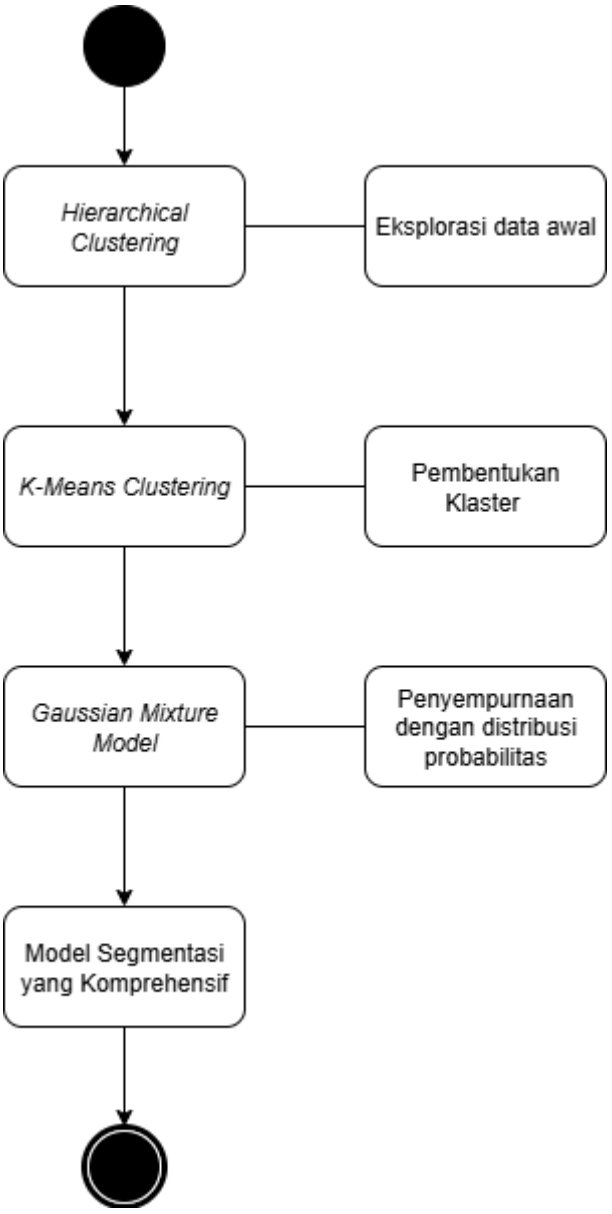
Metoode	Jumlah Klaster	Ciri Umum Klaster	Penamaan/Label Klaster	Hasil
Hierarchical Clustering	2 (Optimal)	Klaster belum terlabel (label belum terbentuk, hanya berbasis kemiripan pola)	Belum diberi label khusus (*Hierarchical clustering* hanya memetakan kemiripan/struktur data)	Menggambarkan struktur data dan hubungan antar responden. Mudah melihat hubungan antar klaster dari dendogram
K-Means	3 (Menyesuaikan segmentasi pola konsumsi)	Muncul klaster dengan profil konsumsi yang berbeda, seperti konvensional, pro-lokal, dan sadar kesehatan	Klaster 0 : Konvensional dan pragmatis Klaster 1 : Peduli Kesehatan Klaster 2 : Pro-Pangan Lokal	Hasil klaster yang jelas dan bisa diinterpretasikan. Sesuai dengan pembentukan segmen.
Gaussian Mixture Model	3 (Menyesuaikan metode sebelumnya)	Profil klaster hampir mirip dengan *K-Means*, namun lebih fleksibel karena setiap responden memiliki probabilitas keanggotaan di tiap klaster.	Klaster 0 : Konvensional dan pragmatis Klaster 1 : Peduli Kesehatan Klaster 2 : Pro-Pangan Lokal	Lebih realistis (*soft clustering*) dan cocok untuk data konsumsi yang dinamis.

Berdasarkan tabel di atas *hierarchical clustering* digunakan untuk tahap segmentasi karena bisa memetakan hubungan kemiripan antara responden tanpa perlu menentukan jumlah klaster di awal. Dari hasil *dendogram*, ditemukan 2 klaster optimal berdasarkan jarak pemotong horizontal terpanjang. Namun, metode ini belum memberikan label klaster, hasilnya masih berupa struktur/pohon kemiripan antar responden. Karena itu, interpretasi lebih mendalam mengenai profil klaster masih terbatas dan perlu dilanjutkan menggunakan metode selanjutnya

untuk mendapatkan segmen yang lebih jelas. Setelah menjalani HC, *K-Means* akan digunakan untuk mengelompokkan data secara lebih eksplisit dengan menggunakan K=3 untuk kebutuhan segmentasi. Hasil analisis membentuk 3 klaster utama yang bisa diinterpretasikan. Klaster 0 yaitu konsumen konvensional dan pragmatis (cenderung memilih makanan praktis, kurang memperhatikan asal pangan), klaster 2 yaitu konsumen pro-lokal (memiliki preferensi tinggi terhadap produk pangan lokal), dan klaster 3 yaitu konsumen dengan kesadaran tentang kesehatan (memilih pangan berdasarkan kandungan gizi dan kesehatan). Klaster *K-Means* mudah diinterpretasikan dan bisa dijadikan dasar segmentasi masyarakat berdasarkan pola konsumsi. Sebagai pelengkapnya, GMM digunakan untuk mendapatkan hasil klaster yang lebih dinamis dan realistis, karena GMM memberikan probabilitas keanggotaan setiap individu pada masing-masing klaster. Hasil analisis GMM umumnya menghasilkan komposisi klaster yang mirip dengan *K-Means*. Namun, GMM mampu menangkap fenomena bahwa satu data bisa memiliki dua klaster sekaligus atau biasa disebut *soft clustering*.

Secara keseluruhan, integrasi ketiga metode yang sudah digunakan dalam penelitian ini membentuk kerangka segmentasi yang saling melengkapi. Prosesnya dimulai dari *hierarchical clustering* (HC), yang dapat melakukan pemetaan hubungan alami antar responden dan membantu dalam menentukan jumlah klaster optimal secara visual. Kemudian, *K-Means clustering* dijalankan untuk pemetaan secara eksplisit, di mana data dapat dikelompokka ke dalam klaster-klaster dengan berdasarkan kemiripan karakteristik tertentu, sehingga dapat menghasilkan segmentasi yang terstruktur dan tegas. Tahap terakhir dilakukan oleh GMM, yang diterapkan sebagai penyempurnaan segmentasi dengan pendekatan probabiliastik. Pada GMM, setiap responden memiliki peluang keanggotaan di lebih dari satu klaster (*soft clustering*).

Dengan demikian, kerangka model segmentasi ini menghasilkan pembagian kelompok yang memberikan gambaran menyeluruh mengenai keragaman pola konsumsi masyarakat. Setiap tahapan saling melengkapi, dimulai dari HC untuk mengeksplorasi data awal, *K-Means* sebagai pembentukan klaster, serta GMM yang menyempurnakan berdasarkan distribusi probabilitas. Hasil akhirnya dapat dijadikan dasar yang kuat untuk perumusan kebijakan pangan yang tepat sasaran, karena kebijakan dapat disesuaikan dengan kerakteristik dan kecenderungan pola konsumsi dari masing-masing klaster. Sebagai gambaran, alur penggunaan model segmentasi ini dapat diilustrasikan sebagai berikut.



Gambar 4. 21. Diagram Gambaran Penerapan Ketiga Metode

Diagram ini menunjukkan bagaimana setiap metode memiliki peran dalam membangun model pola konsumsi yang saling melengkapi. Dimulai dari eksplorasi data, pembentukkan klaster, hingga penyempurnaan hasil akhir. Pendekatan yang terintegrasi ini akan menjadi kekuatan utama dalam penelitian ini dalam membaca keberagaman dan kompleksitas perilaku konsumsi masyarakat secara lebih relevan.

BAB V
KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini bertujuan membangun model segmentasi pola konsumsi masyarakat dengan memanfaatkan tiga metode *machine learning*, yaitu *hierarchical clustering*, *k-means clustering*, dan *Gaussian Mixture Model* (GMM), ketiga model ini digunakan untuk saling melengkapi satu sama lain, bukan untuk membandingkan satu sama lain. Penelitian ini berhasil membentuk segmentasi pola konsumsi masyarakat melalui tiga metode utama, yaitu *hierarchical clustering*, *K-Means*, dan *Gaussian Mixture Model* (GMM). *Hierarchical clustering* digunakan sebagai langkah awal untuk mengeksplorasi struktur data dan menentukan jumlah klaster yang paling tepat, yang hasilnya dapat memetakan struktur kemiripan responden, tetapi belum dapat langsung memberikan label klaster secara tegas.

Hasil evaluasi dari algoritma *hierarchical clustering* sebelum dan sesudah pemilihan atribut adalah 0,1194 dan 0,3398. Selanjutnya, *K-Means* memberikan segmentasi yang lebih jelas dengan membagi responden ke dalam tiga klaster berdasarkan preferensi pangan lokal, perhatian pada kesehatan, dan kecenderungan konsumsi pangan praktis. Label klaster *k-means* sama dengan label klaster yang terdapat di GMM. Nilai *silhouette score* yang menunjukkan evaluasi pembagian klaster sebelum dan sesudah pemilihan atribut adalah masing-masing 0,1443 dan 0,3513. Metode GMM kemudian memperkuat hasil dengan pendekatan probabilistik, sehingga mampu menangkap keragaman responden yang memiliki kecenderungan pada lebih dari satu klaster. Hasil penelitian menunjukkan bahwa masyarakat dapat dikelompokkan ke dalam tiga segmen utama berdasarkan pola konsumsi, preferensi pangan lokal, serta kepedulian terhadap isu kesehatan dan keberlanjutan.

1. Klaster Peduli Kesehatan dan Lingkungan

   Anggota klaster ini memiliki konsumsi protein baik hewani maupun nabati paling tinggi (masing-masing 2,77 dan 3,00), jarang membeli makanan instan (2,10), serta memiliki perhatian yang besar terhadap pangan keberlanjutan (3,00), dan ketergantungan terhadap pangan impor yang rendah (1,63). Klaster ini mencerminkan konsumen yang sadar akan gizi, peduli terhadap lingkungan, dan kecenderungan kuat untuk memilih pangan yang sehat sera berkontribusi dalam sistem pangan.
2. Klaster Pro Pangan Lokal

Klaster ini memiliki anggota dengan konsumsi protein hewani dan nabati yang sedang atau seimbang (masing-masing 2,42 dan 2,14), pembelian makanan siap saji atau instan yang cukup moderat (2,40), dan tingkat partisipasi pada kegiatan pertanian belum tampak signifikan. Tetapi, ketergantungan pangan impor yang relatif rendah (1,91), serta perhatian terhadap pangan keberlanjutan yang cukup (1,91). Kelompok ini sudah mulai mengutamakan produk lokal dan memperhatikan aspek keberlanjutan, walaupun perilaku konsumsi sehat belum diterapkan secara optimal.

3. Klaster Konvensional dan Pragmatis

   Klaster memiliki konsumsi protein hewani dan nabati yang paling rendah (1,99 dan 2,08), dan yang paling sering membeli makanan instan (3,04), serta ketergantungan terhadap pangan impor yang paling tinggi (3,05). Partisipasi kegiatan pertanian lebih tinggi dibanding klaster lain (1,00), tetapi perhatian terhadap pangan berkelanjutan masih rendah (1,88)

5.2. Saran

Saran untuk penelitian selanjutnya adalah untuk memperluas cakupan wilayah dan menambah jumlah responden agar hasil segmentasi lebih representatif. Kemudian saran juga untuk mengembangkan aplikasi sistem pendukung keputusan tentang ketahanan pangan dengan berdasarkan data dari segmentasi, agar pemangku kepentingan dapat membuat kebijakan untuk pangan berkelanjutan.

Saran untuk sisi masyarakat adalah diharapkan agar lebih kritis dalam memilih pangan, dengan mempertimbangkan aspek kesehatan, keberlanjutan, dan dukungan terhadap produk lokal. Perlu untuk meningkatkan kesadaran kolektif untuk membangun sistem pangan yang lebih kuat dan berkelanjutan demi generasi mendatang.

Daftar Pustaka

Abdella, G. M., Kucukvar, M., Onat, N. C., Al-Yafay, H. M., & Bulak, M. E. (2020). Sustainability assessment and modeling based on supervised machine learning techniques: The case for food consumption. *Journal of Cleaner Production*, *251*, 119661. https://doi.org/10.1016/j.jclepro.2019.119661

Akande, O., Asani, E. O., & Dautare, B. (2024). Customer Segmentation Through RFM Analysis and K-Means Clustering: Leveraging Data-Driven Insights for Effective Marketing Strategy. *Ceddi Journal of Information System and Technology (JST)*, *3*(1), 14–25. https://doi.org/10.56134/jst.v3i1.81

Alamtaha, Z., Djakaria, I., & Yahya, N. I. (2023). *Implementasi Algoritma Hierarchical Clustering dan Non-Hierarchical Clustering untuk Pengelompokkan Pengguna Media Sosial*. *4*(1), 33–43. https://doi.org/10.20956/ejsa.vi.24830

Almulhim, A. I., & Abubakar, I. R. (2024). A segmentation approach to understanding water consumption behavioral patterns among households in Saudi Arabia for a sustainable future. *Resources, Environment and Sustainability*, *15*(August 2023), 100144. https://doi.org/10.1016/j.resenv.2023.100144

An, P., Wang, Z., & Zhang, C. (2022). Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection. *Information Processing & Management*, *59*(2), 102844. https://doi.org/10.1016/j.ipm.2021.102844

Ashari, A. M. (2023). Dampak Perubahan Iklim terhadap Ketahanan Pangan dan Adaptasinya oleh Masyarakat Pesisir. *Empiricism Journal*, *4*(2), 426–431. https://doi.org/10.36312/ej.v4i2.1611

Biresselioglu, M. E., Kentmen-Cin, C., Demir, M. H., Savas, Z. F., Solak, B., Onder, B., Ceviker-Cinar, G., & Ozcureci, B. (2023). How to Exploit Sustainable Food Consumption Habits of Individuals: Evidence from a Household Survey in Izmir, Türkiye. *Sustainability (Switzerland)*, *15*(10). https://doi.org/10.3390/su15108271

Delabre, I., Rodriguez, L. O., Smallwood, J. M., Scharlemann, J. P. W., Alcamo, J., Antonarakis, A. S., Rowhani, P., Hazell, R. J., Aksnes, D. L., Balvanera, P., Lundquist, C. J., Gresham, C., Alexander, A. E., & Stenseth, N. C. (2021). Actions on sustainable food production and consumption for the post-2020 global biodiversity framework. *Science Advances*, *7*(12). https://doi.org/10.1126/sciadv.abc8259

Drewnowski, A., Finley, J., Hess, J. M., Ingram, J., Miller, G., & Peters, C. (2020). Toward Healthy Diets from Sustainable Food Systems. *Current Developments in Nutrition*, *4*(6),

nzaa083. https://doi.org/10.1093/cdn/nzaa083

Garai-Fodor, M. (2021). Food Consumption Patterns, in a Values-based Approach, for Generation Z. *Acta Polytechnica Hungarica*, *18*(11), 117–134. https://doi.org/10.12700/APH.18.11.2021.11.7

Gazdecki, M., Goryńska-Goldmann, E., Kiss, M., & Szakály, Z. (2021). Segmentation of Food Consumers Based on Their Sustainable Attitude. *Energies*, *14*(11), 3179. https://doi.org/10.3390/en14113179

Knorr, D., Augustin, M. A., & Tiwari, B. (2020). Advancing the Role of Food Processing for Improved Integration in Sustainable Food Chains. *Frontiers in Nutrition*, *7*(April), 1–8. https://doi.org/10.3389/fnut.2020.00034

Li, R., & Roe, B. E. (2024). Segmenting U.S. consumers by food waste attitudes and behaviors: Opportunities for targeting reduction interventions. *Sustainable Production and Consumption*, *45*(January), 348–358. https://doi.org/10.1016/j.spc.2024.01.015

Maciejewski, G., & Lesznik, D. (2022). Consumers Towards the Goals of Sustainable Development: Attitudes and Typology. *Sustainability*, *14*(17), 10558. https://doi.org/10.3390/su141710558

Mensah, K., Wieck, C., & Rudloff, B. (2024). Sustainable food consumption and Sustainable Development Goal 12: Conceptual challenges for monitoring and implementation. *Sustainable Development*, *32*(1), 1109–1119. https://doi.org/10.1002/sd.2718

O'Neill, C., McCarthy, M. B., O'Reilly, S., & Alfnes, F. (2023). Food interests, preferences and behaviours: a profile of the sustainable food consumer. *British Food Journal*, *125*(13), 352–374. https://doi.org/10.1108/BFJ-09-2022-0762

Patria, A. S., Ekohariadi, E., & Sutiadiningsih, A. (2023). *Validity and Reliability Test of Literacy Skills Assessment Instrument for Graphic Design Student*. Atlantis Press SARL. https://doi.org/10.2991/978-2-38476-008-4_206

Piligrimiene, Z., Banyte, J., Dovaliene, A., Gadeikiene, A., & Korzilius, H. (2021). Sustainable consumption patterns in different settings. *Engineering Economics*, *32*(3), 278–291. https://doi.org/10.5755/j01.ee.32.3.28621

Piracci, G., Casini, L., Contini, C., Stancu, C. M., & Lähteenmäki, L. (2023). Identifying key attributes in sustainable food choices: An analysis using the food values framework. *Journal of Cleaner Production*, *416*(June), 137924. https://doi.org/10.1016/j.jclepro.2023.137924

Radde, H. A., Nurrahmah, A, N., & Nur Aulia Saudi, A. (2021). Uji Validitas Konstrak dari Emotion Regulation Questionnaire Versi Bahasa Indonesia dengan Menggunakan

Confirmatory Factor Analysis Constract Validity Test of Emotions Regulation Questionnaire of Indonesian Version Using Confirmatory Factor Analysis. *Jurnal Psikologi Karakter*, *1*(2), 156–160. https://journal.unibos.ac.id/jpk

Ran, Y., Nilsson Lewis, A., Dawkins, E., Grah, R., Vanhuyse, F., Engström, E., & Lambe, F. (2022). Information as an enabler of sustainable food choices: A behavioural approach to understanding consumer decision-making. *Sustainable Production and Consumption*, *31*, 642–656. https://doi.org/10.1016/j.spc.2022.03.026

Ruben, R., Cavatassi, R., Lipper, L., Smaling, E., & Winters, P. (2021). Towards food systems transformation—five paradigm shifts for healthy, inclusive and sustainable food systems. *Food Security*, *13*(6), 1423–1430. https://doi.org/10.1007/s12571-021-01221-4

Sayili, U., Sak, K., Aydin, S. N., Kara, B., Turgut, D., & Bisgin, O. (2024). Development, validity and reliability of the healthy lifestyle behavior scale. *Discover Public Health*, *21*(1). https://doi.org/10.1186/s12982-024-00186-x

Schäufele-Elbers, I., & Janssen, M. (2023). Consumer segmentation based on three dimensions of sustainable food consumption: a simultaneous analysis of meat, organic food, and sweet snack purchases based on household panel data in Germany. *Frontiers in Nutrition*, *10*. https://doi.org/10.3389/fnut.2023.1140636

Siemon, M. S. N., Shihavuddin, A. S. M., & Ravn-Haren, G. (2021). Sequential transfer learning based on hierarchical clustering for improved performance in deep learning based food segmentation. *Scientific Reports*, *11*(1), 813. https://doi.org/10.1038/s41598-020-79677-1

Singhal, A., Singh, P., Lall, B., & Joshi, S. D. (2020). Modeling and prediction of COVID-19 pandemic using Gaussian mixture model. *Chaos, Solitons & Fractals*, *138*, 110023. https://doi.org/10.1016/j.chaos.2020.110023

Tabianan, K., Velu, S., & Ravi, V. (2022). K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. *Sustainability (Switzerland)*, *14*(12), 1–15. https://doi.org/10.3390/su14127243

Ummami, R., & Winarno, B. (2023). Gaussian Mixture Model dengan Algoritme Expectation Maximization untuk Pengelompokan Data Distribusi Air Bersih di Jawa Barat. *PRISMA, Prosiding Seminar Nasional Matematika*, *6*, 745–750. https://journal.unnes.ac.id/sju/index.php/prisma/

Wongprawmas, R., Mora, C., Pellegrini, N., Guiné, R. P. F., Carini, E., Sogari, G., & Vittadini, E. (2021). Food Choice Determinants and Perceptions of a Healthy Diet among Italian Consumers. *Foods*, *10*(2), 318. https://doi.org/10.3390/foods10020318

Zhang, Y., Ai, Q., & Li, Z. (2021). Grouping of dynamic electricity consumption behaviour: An f -divergence based hierarchical clustering model. *IET Generation, Transmission & Distribution*, *15*(22), 3164–3175. https://doi.org/10.1049/gtd2.12248

Lampiran

Code program *Google Colab* sebelum pemilihan atribut
https://colab.research.google.com/drive/1S9AYQdYyQMxzec1y7AhV_Ag58n4QJNKZ?usp=sharing

Code program *Google Colab* setelah pemilihan atribut
https://colab.research.google.com/drive/1LCbQv9D0IoGuNzjWNEcKtfISX1PFxeSN?usp=sharing

Data kuesioner asli dan data hasil mapping/pembersihan
https://docs.google.com/spreadsheets/d/1hJ9JXuIwZHuJ4pdZf9M1j2BU-TDs1SFWyW4R5jn74Ew/edit?usp=sharing