



Penerapan Algoritma K-Means Clustering untuk Pengelompokan Karakteristik Limbah Makanan

Imaulidia Rediutami¹, Achmad Mufliq², Syahri Mu'min³

^{1,2,3}Sistem Informasi, Ilmu Komputer, Universitas Nahdlatul Ulama Sidoarjo

121422021.student@unusida.ac.id . 2achmadmufliq.it@unusida.ac.id , 3syahri.si@unusida.ac.id

Abstract

Food waste is one of the environmental problems that can have a significant impact on environmental, economic, and social aspects. The large pile of food waste is caused by inefficient production and consumption patterns, as well as low public and government understanding of the characteristics of food waste that are complex and represented by various sectors such as households, retail, and fast food and beverage businesses. To analyze complex food waste data, an analytical method is needed that is able to simplify complex data into easy-to-understand information. This study aims to optimize the use of the K-Means Clustering algorithm in the grouping of food waste data based on the similarity of its characteristics using Jupyter Notebook. The data used for this study includes several food waste indicators such as food waste per capita, per ton and food waste from various sectors. Before clustering, the data will go through a pre-processing stage which includes checking the completeness of the data and standardizing the data using Robust Scaler to equalize the scale of all its attributes. Then, determine the number of clusters by using the evaluation of the Elbow Method and Silhouette Score models. After that, the final clustering process is carried out for grouping the attributes with the number of clusters that have been determined from the results of the second model evaluation from two clusters. After that, the final clustering results will be evaluated by using the Silhouette Score and Davies-Bouldin Index methods. Based on the results of the evaluation, the Silhouette Score produced a value of 0.966 and the Davies-Bouldin Score with a value of 0.231. Last, the results of clustering are expected to be used as a basis for designing targeted educational programs and as a basis for formulating accurate and efficient waste management policies.

Keywords: Clustering, K-Means Clustering, Waste, Food Waste, Environment

Abstrak

Limbah makanan merupakan salah satu permasalahan lingkungan yang dapat memberikan dampak signifikan terhadap aspek lingkungan, ekonomi, serta sosial. Banyaknya timbunan limbah makanan disebabkan oleh pola produksi dan konsumsi yang tidak efisien, serta rendahnya pemahaman publik dan pemerintah terhadap karakteristik limbah makanan yang bersifat kompleks dan direpresentasikan oleh berbagai sektor seperti rumah tangga, ritel, dan bisnis makanan dan minuman cepat saji. Untuk menganalisis data limbah makanan yang kompleks tersebut, diperlukan metode analitis yang mampu menyederhanakan data kompleks menjadi informasi yang mudah dipahami. Penelitian ini bertujuan mengoptimalkan pemanfaatan algoritma K-Means Clustering dalam pengelompokan data limbah makanan berdasarkan kesamaan karakteristiknya menggunakan Jupyter Notebook. Data yang digunakan untuk penelitian ini mencakup beberapa indikator limbah makanan, seperti limbah makanan per kapita, per ton, dan limbah makanan dari berbagai sektor. Sebelum dilakukan klusterisasi, data akan melalui tahap pra-pemrosesan yang meliputi pengecekan kelengkapan data serta standarisasi data menggunakan Robust Scaler untuk menyamakan skala semua atributnya. Kemudian, menentukan jumlah kluster menggunakan evaluasi model Elbow Method dan Silhouette Score. Setelah itu, dilakukan proses klusterisasi akhir yaitu pengelompokan atribut dengan jumlah kluster yang sudah ditentukan dari hasil kedua evaluasi model yaitu dua kluster. Setelah itu, hasil klusterisasi akhir akan dievaluasi menggunakan metode Silhouette Score dan Davies-

Bouldin Index. Berdasarkan hasil evaluasi tersebut, Silhouette Score menghasilkan nilai 0,966 dan Davies-Bouldin Score dengan nilai 0,231. Dengan ini, hasil klusterisasi diharapkan mampu digunakan sebagai dasar perancangan program edukasi yang tepat sasaran dan sebagai dasar perumusan kebijakan pengelolaan sampah yang akurat dan efisien.

Kata kunci: Klusterisasi, K-Means Clustering, Sampah, Limbah Makanan, Lingkungan

© 2026 Author

Creative Commons Attribution 4.0 International License



1. Pendahuluan

Permasalahan limbah makanan telah menjadi isu global karena dampaknya yang signifikan terhadap degradasi lingkungan, kesejahteraan sosial, serta kerugian ekonomi, termasuk di tingkat nasional. Data menunjukkan bahwa di Indonesia, sampah makanan menjadi salah satu jenis sampah terbesar yang dihasilkan dari aktivitas rumah tangga. Sehingga menimbulkan tantangan tersendiri dalam pengelolaan sampah berkelanjutan serta target *Sustainable Development Goals (SDGs)* dalam pengurangan *food loss* dan *food waste* [1]. Permasalahan limbah makanan tidak hanya terletak pada besarnya volume yang dihasilkan, tetapi juga pada kompleksitasnya yang mencakup berbagai jenis limbah makanan dari berbagai sektor yaitu sektor rumah tangga, sektor ritel, serta sektor bisnis makanan dan minuman cepat saji. Oleh karena itu, dibutuhkan pengetahuan yang mendalam mengenai karakteristik limbah makanan agar pemerintah maupun pemangku kebijakan terkait dapat memahami permasalahan secara menyeluruh dan merumuskan kebijakan yang efektif, baik berupa program edukasi untuk masyarakat maupun dasar penyusunan kebijakan untuk menentukan strategi penanganan limbah makanan.

Dalam pendekatan *data mining*, klusterisasi merupakan teknik analisis yang berfungsi untuk mengorganisasi data ke dalam kelompok-kelompok homogen berdasarkan ukuran kedekatan atau kesamaan karakteristik antar data [2]. Dalam teknik klusterisasi terdapat berbagai metode yang dapat diterapkan, salah satunya adalah algoritma K-Means Clustering [3]. K-Means adalah teknik pengelompokan data yang membagi objek ke dalam kluster-kluster dengan karakteristik serupa, sehingga pola timbunan sampah yang bersifat multidimensi dapat diinterpretasikan dalam bentuk kelompok yang lebih sederhana dan informatif.

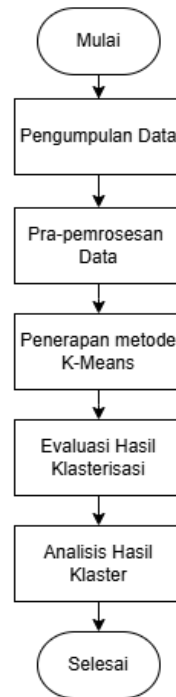
Beberapa penelitian terdahulu telah menerapkan algoritma K-Means untuk mengelompokkan data yang berkaitan dengan lingkungan, sehingga menghasilkan pola kluster yang mampu menggambarkan karakteristik dan perbedaan kondisi lingkungan di setiap kelompok. Pada penelitian yang membahas mengenai penerapan algoritma K-Means untuk pengelompokan data tingkat pencemaran sampah plastik di seluruh kabupaten di Indonesia menunjukkan bahwa algoritma ini dapat memetakan daerah-daerah berdasarkan tingkat pencemaran sampah plastik dan menghasilkan informasi kluster serta visualisasi kluster yang dapat memudahkan masyarakat dalam memahami perbedaan dan perubahan jumlah limbah plastik pada setiap kluster [4]. Penelitian lain yang membahas mengenai penerapan algoritma K-Means dalam ranah pengelolaan sampah mampu menghasilkan pola kluster yang optimal dan mudah diinterpretasikan. Hasil klusterisasi tersebut menunjukkan bahwa data jumlah produksi sampah di wilayah Jakarta yang, menghasilkan pola kluster yang optimal serta mudah dipahami sehingga dapat dimanfaatkan untuk membuat strategi pengelolaan produksi sampah di Jakarta dan menjadi landasan penentuan langkah strategis yang lebih efektif [5]. Penelitian yang menganalisis data jumlah volume sampah organik di Kota Magelang menghasilkan kluster yang dapat memanfaatkan sebagai dasar pengambilan keputusan dalam menangani permasalahan sampah organik di Kota Magelang [6].

Selain diterapkan dalam pengelolaan sampah, algoritma K-Means dapat diterapkan dalam kajian yang membahas mengenai polusi udara. Penelitian tersebut telah membuktikan bahwa penerapan algoritma K-Means efektif dalam mengklusterisasi data pemantauan pencemaran udara sehingga menghasilkan kluster yang dapat dimanfaatkan untuk membuat dasar kebijakan dalam upaya penanganan pencemaran lingkungan [7].

Berdasarkan hasil penelitian-penelitian di atas, membuktikan bahwa dengan menerapkan algoritma K-Means dalam proses klusterisasi pada data yang berkaitan dengan lingkungan seperti sampah plastik, sampah organik, serta data tingkat pencemaran udara dapat menghasilkan pola yang bermakna dan informatif sehingga dapat dimanfaatkan sebagai perumusan kebijakan yang didukung oleh temuan analitis. Namun, dari penelitian-penelitian tersebut belum ada penelitian yang membahas mengenai pengelompokan karakteristik limbah makanan. Oleh karena itu, penelitian ini dilakukan sebagai pendekatan alternatif berbasis data untuk mendukung keberlanjutan lingkungan yaitu dengan mengelompokkan karakteristik limbah makanan sebagai dasar peningkatan efisiensi pengelolaan sampah.

2. Metode Penelitian

Dalam penelitian ini, metode penelitian disusun secara sistematis, dimana setiap tahapannya saling berkaitan antara satu sama lain sehingga membentuk alur penelitian yang utuh [8]. Dengan itu, data mentah yang diperoleh dari proses pengumpulan data dapat diolah sehingga menghasilkan informasi yang bermakna yang dapat digunakan sebagai dasar pengambilan keputusan. Selain itu, metode penelitian dapat membantu mempermudah dalam pencapaian tujuan penelitian [9].



Gambar 1. Tahapan Penelitian

Gambar 1 merupakan gambar alur metode penelitian yang digunakan dalam penelitian ini, yang terdiri dari lima tahapan, yaitu tahap pengumpulan data, pra-pemrosesan data, penerapan metode K-Means, evaluasi hasil klusterisasi, serta analisis hasil klaster. Alur metode penelitian tersebut disusun secara sistematis dan saling berkaitan antara tahapan satu dengan yang lainnya untuk mencapai tujuan penelitian yaitu menghasilkan klaster yang representatif serta memastikan bahwa hasil klusterisasi karakteristik limbah makanan relevan dalam mendukung peningkatan efisiensi pengelolaan sampah. Penjelasan mengenai setiap tahapan dalam alur metode penelitian tersebut dapat diuraikan sebagai berikut:

2.1. Pengumpulan Data

Tahap pengumpulan data adalah tahap awal dari alur penelitian yang berfungsi sebagai dasar untuk seluruh tahapan analisis, yang bertujuan untuk mengumpulkan data mentah yang berkaitan dengan permasalahan penelitian [10]. Data yang dikumpulkan akan diolah dan dianalisis. Oleh karena itu, data yang dikumpulkan harus lengkap, dapat merepresentasikan kondisi nyata, serta sesuai dengan kebutuhan analisis agar hasil analisis dapat mencapai tujuan penelitian secara optimal. Dalam penelitian ini, data yang akan dianalisis adalah data yang berkaitan dengan karakteristik limbah makanan dalam berbagai sektor, yaitu limbah makanan rumah tangga, ritel, serta bisnis makanan dan minuman cepat saji yang berjumlah 214. Data diambil dari situs Kaggle yang berisi kumpulan data limbah makanan dari berbagai laporan dan publikasi, antara lain laporan *Food and Agriculture Organization (FAO)*, *United Nations Environment Programme (UNEP)*, jurnal ilmiah, dan organisasi riset terkait.

2.2. Pra-pemrosesan Data

Tahap pra-pemrosesan data adalah tahapan yang bertujuan untuk meningkatkan kualitas data sebelum dilakukan proses klusterisasi. Data yang diperoleh dari tahap pengumpulan data yang belum melalui tahap pra-pemrosesan data, memungkinkan memiliki beberapa permasalahan seperti nilai kosong, perbedaan format antar variabel, serta

perbedaan skala antar variabel yang dapat mempengaruhi keakuratan hasil analisis. Oleh karena itu, pada tahap ini, data biasanya dibersihkan melalui beberapa proses pra-pemrosesan data.

Dalam penelitian ini, pra-pemrosesan data dimulai dari pemilihan atribut, cek nilai kosong, kemudian standarisasi data. Setelah data dikumpulkan, dilakukan proses pemilihan atribut yang mencerminkan karakteristik limbah makanan di suatu negara sebelum melalui proses klusterisasi. Pemilihan atribut merupakan tahapan penting dalam penelitian karena atribut yang digunakan sangat menentukan kualitas hasil klusterisasi. Atribut yang tidak relevan dengan topik penelitian akan menghasilkan kluster yang bias, tidak stabil, dan sulit diinterpretasikan. Oleh karena itu, pemilihan atribut dalam penelitian ini dilakukan secara konseptual dengan mempertimbangkan tujuan penelitian, ketersediaan data, serta karakteristik algoritma K-Means. Setelah pemilihan atribut, dilakukan pemeriksaan nilai kosong, kemudian data akan disamakan skala antar variabelnya menggunakan Robust Scaler. Berikut merupakan rumus Robust Scaler:

$$X_{scaled} = \frac{X - \text{Median}}{IQR} \quad (1)$$

Keterangan:

- X = nilai data asli
- Median = nilai tengah dari seluruh data di setiap variabel
- IQR = rentang antara kuartil ketiga dan pertama

2.3. Penerapan Metode K-Means

Dalam penelitian ini, setiap tahapan klusterisasi mulai dari tahap pra-pemrosesan data hingga analisis hasil kluster dilakukan menggunakan Jupyter Notebook dengan bahasa pemrograman Python. Algoritma K-Means merupakan algoritma yang diterapkan pada penelitian ini untuk menentukan jumlah kluster optimal serta mengelompokkan data berdasarkan tingkat kemiripan atribut dengan jumlah kluster yang sudah ditentukan [11]. Algoritma K-Means adalah salah satu metode klusterisasi yang mampu mengelompokkan data dalam jumlah besar. K-Means memiliki keunggulan dibandingkan algoritma klusterisasi lainnya, antara lain memiliki kecepatan komputasi lebih tinggi dan efisien untuk mengelola data dalam jumlah besar, lebih sederhana dan mudah dipahami, serta dapat menghasilkan kluster dengan tingkat kemiripan yang tinggi dalam kluster dan keterpisahan antar kluster yang jelas. Dalam proses klusterisasi menggunakan algoritma K-Means, jumlah kluster tidak ditentukan secara otomatis oleh algoritma. Oleh karena itu, dibutuhkan evaluasi model untuk menentukan jumlah kluster yang sesuai dengan karakteristik data yang dianalisis [12]. Dalam penelitian ini, evaluasi model yang diterapkan untuk menentukan jumlah kluster adalah Elbow Method dan Silhouette Score. Berikut merupakan rumus Elbow Method:

$$WSS = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (2)$$

Keterangan:

- k = jumlah kluster
- n = jumlah data dalam kluster j
- $x_i^{(j)}$ = data ke i dalam kluster j
- c_j = centroid kluster j
- $\|x_i^{(j)} - c_j\|^2$ = kuadrat jarak Euclidean antara titik data dan pusat kluster

Selain itu, rumus Silhouette Score ditunjukkan sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

Keterangan:

- s_i = nilai Silhouette Score untuk data ke-i
- i = indeks atau penanda objek data
- $a(i)$ = rata-rata jarak antara data ke-i dengan seluruh data lain dalam kluster yang sama
- $b(i)$ = rata-rata jarak antara data ke-i dengan data pada kluster terdekat yang berbeda
- $\max\{a(i), b(i)\}$ = nilai maksimum antara $a(i)$ dan $b(i)$

Setelah mendapatkan jumlah kluster optimal, proses selanjutnya adalah melakukan klusterisasi dengan jumlah kluster yang sudah ditentukan. Umumnya, klusterisasi menggunakan algoritma K-Means dihitung menggunakan

rumus Euclidean Distance, yaitu rumus untuk menghitung jarak antar data dan titik *centroid* [13]. Berikut merupakan rumus Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Keterangan:

- x_i = nilai atribut data ke- i
- y_i = nilai *centroid* ke- i
- n = jumlah atribut

2.4. Evaluasi Hasil Klasterisasi

Tahap evaluasi hasil klasterisasi dalam penelitian ini berperan untuk menilai kualitas klaster yang dihasilkan dari proses klasterisasi [14]. Pada tahap ini, hasil klasterisasi dinilai berdasarkan jarak kerapatan data dalam klaster serta jarak keterpisahan antar klaster [15]. Hasil klasterisasi yang baik ditunjukkan oleh kerapatan data di dalam klaster serta tingkat keterpisahan antar klaster yang tinggi. Oleh karena itu, dengan penerapan evaluasi hasil klasterisasi dapat dipastikan bahwa klaster yang terbentuk memiliki struktur pengelompokan yang optimal sehingga mampu merepresentasikan perbedaan data secara jelas dan bermakna.

Dalam penelitian ini, evaluasi hasil klasterisasi dilakukan menggunakan metode Silhouette Score dan Davies-Bouldin Index. Penerapan metode Silhouette Score bertujuan untuk mengukur kualitas penempatan data dalam klaster dengan membandingkan jarak rata-rata data terhadap klaster asalnya dan jarak klaster terdekat.

Selain Silhouette Score, metode yang digunakan untuk evaluasi hasil klasterisasi dalam penelitian ini adalah Davies-Bouldin Index. Penerapan metode Davies-Bouldin Index bertujuan untuk menilai kualitas klasterisasi berdasarkan perbandingan nilai kerapatan klaster terhadap jarak pemisah antar klaster. Berikut merupakan rumus perhitungan Davies-Bouldin Index:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (5)$$

Keterangan:

- k = jumlah klaster
- S_i = rata-rata jarak data dalam klaster ke- i terhadap *centroid* klaster tersebut
- S_j = rata-rata jarak data dalam klaster ke- j terhadap *centroid* klaster tersebut
- M_{ij} = jarak antara *centroid* klaster ke- i dan klaster ke- j

2.5. Analisis Hasil Klaster

Tahap analisis hasil klaster merupakan tahap akhir dari tahapan metode penelitian yang berfungsi untuk mengidentifikasi klaster yang terbentuk dari proses klasterisasi menggunakan metode K-Means. Setelah melalui tahap evaluasi hasil klasterisasi, setiap klaster yang terbentuk akan diidentifikasi berdasarkan karakteristiknya. Setiap klaster memiliki karakteristik yang berbeda secara signifikan. Dalam penelitian ini, karakteristik setiap klaster mencakup jumlah dan tingkat volume limbah makanan yang dihasilkan dari berbagai sektor, yaitu sektor rumah tangga, ritel, serta bisnis makanan dan minuman cepat saji.

3. Hasil dan Pembahasan

Hasil dan pembahasan dalam penelitian ini adalah bagian yang menjelaskan mengenai pengolahan dan analisis data yang diperoleh berdasarkan hasil yang didapatkan dari seluruh tahapan metode penelitian. Pada bagian ini, hasil yang diperoleh dari seluruh tahapan metodologi tidak hanya ditampilkan, tetapi juga dijelaskan secara kritis untuk menunjukkan kejelasan proses untuk mencapai tujuan penelitian. Penyajian hasil analisis dijelaskan secara sistematis dan berurutan mengikuti alur analisis dan memahami bagaimana data mentah berkembang menjadi informasi yang memiliki nilai analitis. Dengan adanya penjelasan mengenai hasil analisis, bagian ini berkontribusi langsung dalam mendukung efisiensi pengelolaan sampah.

3.1. Hasil Pengumpulan Data

A	B	C	D	E	F	G	H	I	J	K	L	
Country	combined figures (kg/capita/year)	Household estimate (kg/capita/year)	Household estimate (tonnes/year)	Retail estimate (kg/capita/year)	Retail estimate (tonnes/year)	Food service estimate (kg/capita/year)	Food service estimate (tonnes/year)	Confidence in estimate	M49 code	Region	Source	
1	Afghanistan	126	82	3109153	16	594982	28	1051783	Very Low Confidence	4	Southern Asia	https://www...
2	Albania	127	83	238492	16	45058	28	79651	Very Low Confidence	8	Southern Europe	https://www...
3	Algeria	135	91	3918529	16	673360	28	1190335	Very Low Confidence	12	Northern Africa	https://www...
4	Andorra	123	84	6497	13	988	26	1971	Low Confidence	20	Southern Europe	https://www...
5	Angola	144	100	3169523	16	497755	28	879908	Very Low Confidence	24	Sub-Saharan Africa	https://www...
6	Antigua and Barbuda	113	74	7178	13	1244	26	2483	Low Confidence	28	Latin America and the Caribbean	https://www...
7	Argentina	116	72	3243563	16	700380	28	1238100	Very Low Confidence	32	Latin America and the Caribbean	https://www...
8	Armenia	137	93	275195	16	46259	28	81775	Very Low Confidence	51	Western Asia	https://www...
9	Aruba	113	74	7859	13	1362	26	2718	Low Confidence	533	Latin America and the Caribbean	https://www...
10	Australia	133	102	2953130	9	230248	22	549349	High Confidence	26	Australia and New Zealand	https://www...
11	Austria	76	39	349249	9	77289	28	254191	High Confidence	40	Western Europe	https://www...
12	Azerbaijan	137	93	934872	16	157148	28	277800	Very Low Confidence	31	Western Asia	https://www...
13	Bahamas	113	74	28792	13	4990	26	9960	Low Confidence	44	Latin America and the Caribbean	https://www...
14	Bahrain	171	132	216161	13	21026	26	41966	Low Confidence	48	Western Asia	https://www...
15	Bangladesh	84	65	10618233	16	2550080	3	544436	Very Low Confidence	50	Southern Asia	https://www...
16	Barbados	113	74	21215	13	3677	26	7339	Low Confidence	52	Latin America and the Caribbean	https://www...
17	Belarus	132	98	646396	16	147838	28	261341	Very Low Confidence	112	Eastern Europe	https://www...
18	Belgium	60	50	576036	10	112100	20	227371	Medium Confidence	56	Western Europe	https://www...
19	Belize	97	53	20564	16	6106	28	10784	Very Low Confidence	84	Latin America and the Caribbean	https://www...
20	Benin	144	100	1173297	16	184873	28	365280	Very Low Confidence	204	Sub-Saharan Africa	https://www...
21	Bermuda	113	74	4696	13	891	26	1598	Low Confidence	60	Northern America	https://www...
22	Bhutan	123	79	60000	16	11935	28	21998	Very Low Confidence	64	Southern Asia	https://www...
23	Bolivia (Plurinational State of)	124	83	922012	16	180068	28	318315	Very Low Confidence	68	Latin America and the Caribbean	https://www...
24	Bolivia (Plurinational State of)	127	83	272069	16	51628	28	92266	Very Low Confidence	70	Southern Europe	https://www...
25	Bosnia and Herzegovina	136	92	211802	16	36930	28	63693	Very Low Confidence	72	Sub-Saharan Africa	https://www...
26	Brazil	104	60	12578308	16	3300862	28	5835112	Very Low Confidence	76	Latin America and the Caribbean	https://www...
27	British Virgin Islands	113	74	2218	13	384	26	767	Low Confidence	92	Latin America and the Caribbean	https://www...
28	Bruni Darussalam	119	80	34739	13	5551	26	11990	Low Confidence	96	South-eastern Asia	https://www...
29	Bulgaria	112	68	478667	16	109483	28	193539	Very Low Confidence	100	Eastern Europe	https://www...
30	Burkina Faso	147	103	2086893	16	317831	28	561848	Very Low Confidence	854	Sub-Saharan Africa	https://www...
31	Burundi	147	103	1184127	16	160341	28	316799	Very Low Confidence	168	Sub-Saharan Africa	https://www...
32	Cabo Verde	144	100	547653	16	89011	28	15204	Very Low Confidence	132	Sub-Saharan Africa	https://www...
33	Cambodia	130	86	1423387	16	257853	28	458620	Very Low Confidence	116	South-eastern Asia	https://www...

Gambar 2. Data Limbah Makanan

Gambar 2 merupakan tampilan data mentah yang akan dianalisis. Data yang akan digunakan adalah data sekunder yang menggambarkan kondisi limbah makanan di berbagai negara. Data tersebut disajikan dalam berbentuk tabel numerik yang berisi beberapa variabel, yaitu *Country*, *Combined figures (kg/capita/year)*, *Household Estimate (kg/capita/year)*, *Household Estimate (tonnes/year)*, *Retail Estimate (kg/capita/year)*, *Retail Estimate (tonnes/year)*, *Food Service Estimate (kg/capita/year)*, *Food Service Estimate (tonnes/year)*, *Confidence In Estimate*, *M49 Code*, dan *Source*.

Data mentah tersebut diperoleh dari situs Kaggle yang bersumber dari Information is Beautiful pada topik *food waste* berupa visualisasi data dan infografis, kemudian disusun ulang ke dalam bentuk tabel numerik. Visualisasi data tersebut bersumber dari berbagai laporan dan publikasi, antara lain laporan *Food and Agriculture Organization (FAO)*, *United Nations Environment Programme (UNEP)*, jurnal ilmiah, dan organisasi riset terkait.

3.2. Hasil Pra-Pemrosesan Data

Pra-pemrosesan data dalam penelitian ini dimulai dengan pemilihan atribut, cek nilai kosong, kemudian standarisasi data. Setelah data dikumpulkan, dilakukan proses pemilihan atribut yang mencerminkan karakteristik limbah makanan di suatu negara sebelum melalui proses klusterisasi.

	Country	combined figures (kg/capita/year)	Household estimate (tonnes/year)	Retail estimate (tonnes/year)	Food service estimate (tonnes/year)
0	Afghanistan	126	3109153	594982	1051783
1	Albania	127	238492	45058	79651
2	Algeria	135	3918529	673360	1190335
3	Andorra	123	6497	988	1971
4	Angola	144	3169523	497755	879908
...
209	Venezuela (Boliv. Rep. of)	116	2065461	445994	788407
210	Viet Nam	120	7346717	1508689	2666991
211	Yemen	148	3026946	456099	806270
212	Zambia	122	1391729	279350	493822
213	Zimbabwe	144	1458564	229059	404920

214 rows x 5 columns

Gambar 3. Atribut yang Digunakan dalam Proses Klusterisasi

Gambar 3 menampilkan hasil atribut yang akan dikelompokkan berdasarkan karakteristiknya. Atribut yang digunakan untuk klusterisasi adalah *Combined figures (kg/capita/year)*, *Household Estimate (tonnes/year)*, *Retail Estimate (tonnes/year)*, dan *Food Service Estimate (tonnes/year)*. Keempat atribut tersebut mampu

menggambarkan limbah makanan dari sisi konsumsi individu maupun dari berbagai sektor ekonomi. Selain itu, atribut tersebut juga menggambarkan limbah makanan yang terjadi pada tahapan rantai pangan, mulai dari sebelum pengolahan, proses pengolahan, distribusi, hingga konsumsi akhir.

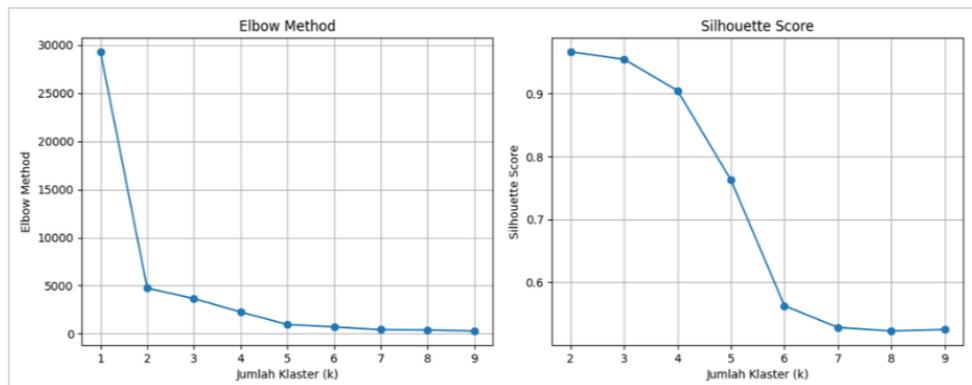
Selanjutnya, dilakukan pengecekan kelengkapan data dan standarisasi data menggunakan Robust Scaler. Berdasarkan Gambar 4 yang menampilkan hasil standarisasi data menunjukkan bahwa nilai setiap variabel sudah tidak sama dengan nilai awal variabel sebelum melalui tahap standarisasi. Nilai antar data limbah makanan dari sektor rumah tangga, ritel, serta bisnis makanan dan minuman cepat saji dengan satuan kg per kapita per tahun maupun ton per tahun menjadi tidak terlalu berbeda dibandingkan sebelumnya, sehingga mudah untuk dibandingkan.

	combined figures (kg/capita/year)	Household estimate (tonnes/year)	Retail estimate (tonnes/year)	Food service estimate (tonnes/year)
0	0.131868	1.267081	1.435732	1.419065
1	0.175824	-0.138040	-0.161461	-0.178863
2	0.527473	1.663251	1.663373	1.646807
3	0.000000	-0.251596	-0.289457	-0.306548
4	0.923077	1.296631	1.153347	1.136548
...
209	-0.307692	0.756218	1.003013	0.986144
210	-0.131868	3.341269	4.089493	4.074039
211	1.098901	1.226842	1.032362	1.015506
212	-0.043956	0.426442	0.519014	0.501924
213	0.923077	0.459156	0.372950	0.355793

214 rows × 4 columns

Gambar 4. Hasil Standarisasi Data

3.3. Penentuan Jumlah Kluster



Gambar 5. Evaluasi Model

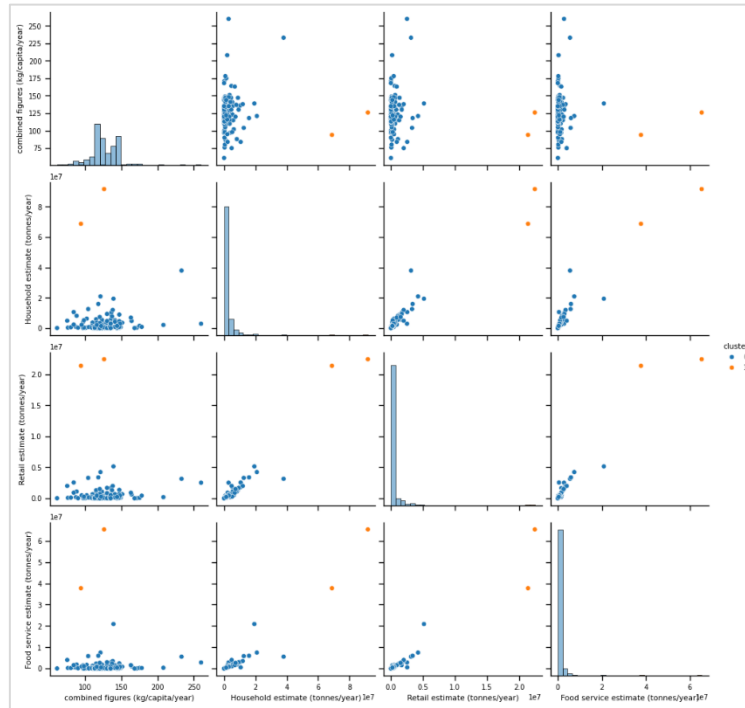
Gambar 5, menampilkan visualisasi hasil evaluasi model menggunakan metode Elbow Method dan Silhouette Score. Kedua metode tersebut digunakan untuk mencari jumlah kluster optimal dan memastikan bahwa jumlah kluster yang dipilih memiliki dasar kuantitatif. Hasil evaluasi menggunakan Elbow Method merupakan evaluasi model yang bekerja dengan menghitung nilai jumlah total variasi data dalam masing-masing kluster, atau biasa disebut dengan nilai Within-Cluster Sum of Squares (WCSS). Pada gambar grafik di atas, menunjukkan bahwa penurunan nilai WCSS terjadi secara signifikan dari K = 1 ke K = 2. Setelah itu nilai WCSS menjadi semakin kecil dan melandai. Jumlah kluster yang berada pada titik siku menunjukkan bahwa jumlah kluster tersebut merupakan jumlah kluster yang optimal yaitu tidak terlalu banyak dan tidak terlalu sedikit. Karena penurunan yang melambat atau tidak signifikan menunjukkan bahwa jumlah kluster sudah tidak memberikan peningkatan pada kualitas kluster. Berdasarkan grafik di atas, jumlah kluster yang berada pada “titik siku” adalah pada K = 2.

Pada evaluasi model menggunakan Silhouette Score, nilai Silhouette berada pada rentang -1 hingga 1, dimana jumlah kluster yang memiliki nilai yang mendekati 1 menunjukkan bahwa data dalam suatu kluster memiliki

tingkat kemiripan yang tinggi dan terpisah dengan baik antar kluster. Pada grafik di atas menunjukkan bahwa jumlah kluster yang memiliki nilai tertinggi atau yang mendekati 1 adalah $K = 2$.

Hasil dari evaluasi menggunakan Elbow Method dan Silhouette Score menunjukkan jumlah kluster optimalnya adalah $K=2$. Kedua metode tersebut mengindikasikan bahwa karakteristik limbah makanan terbagi menjadi dua kelompok utama yang memiliki perbandingan signifikan.

3.4. Klasterisasi dengan Jumlah Kluster yang Sudah Ditentukan



Gambar 6. Visualisasi Hasil Klasterisasi

Gambar 6 menampilkan visualisasi hasil klasterisasi dengan jumlah kluster yang diperoleh dari evaluasi model menggunakan Elbow Method dan Silhouette Score. Kedua metode tersebut menghasilkan jumlah kluster optimalnya adalah dua. Data akan dikelompokkan menjadi dua kluster menggunakan algoritma K-Means. Kemudian, hasil klasterisasi tersebut divisualisasikan menggunakan pairplot untuk melihat pola distribusi dan hubungan antar variabel dalam masing-masing kluster. Visualisasi tersebut disajikan dalam bentuk Scatterplot Matrix yang menampilkan hubungan antar variabel utama, yaitu *Combined figures (kg/capita/year)*, *Household Estimate (tonnes/year)*, *Retail Estimate (tonnes/year)*, dan *Food Service Estimate (tonnes/year)*, dengan warna yang berbeda pada setiap kluster. Setiap kluster diberi warna yang berbeda untuk mempermudah pembacaan visualisasi hasil, kluster 0 diberi warna biru sedangkan kluster 1 diberi warna oranye.

Berdasarkan visualisasi tersebut, terlihat bahwa kluster 0 memiliki jumlah data yang lebih banyak dibandingkan kluster 1. Data pada kluster 0 cenderung berada pada rentang nilai yang lebih rendah hingga menengah, serta memiliki sebaran titik yang lebih padat. Hal tersebut, memperlihatkan tingkat kemiripan atau keseragaman antar titik data dalam kluster. Sedangkan, kluster 1 memiliki jumlah data yang lebih sedikit, yang memiliki nilai ekstrem atau jauh lebih tinggi dibandingkan data lainnya. hal tersebut menunjukkan bahwa kluster 1 merepresentasikan kelompok dengan karakteristik limbah makanan yang berskala besar.

3.5. Evaluasi Hasil Klasterisasi

```

# Silhouette Score
sil_evaluasi_hasil_model = silhouette_score(data_standarisasi, labels)

# Davies-Bouldin Index
dbi_evaluasi_hasil_model = davies_bouldin_score(data_standarisasi, labels)

sil_evaluasi_hasil_model,dbi_evaluasi_hasil_model
    
```

[9] ✓ 0.0s Python

... (0.9664564564849194, 0.23135004415814106)

Gambar 7. Evaluasi Hasil Klasterisasi

Gambar 7 menampilkan kode dan hasil dari penerapan metode Silhouette Score dan Davies-Bouldin Index untuk evaluasi hasil klasterisasi menggunakan algoritma K-Means. Evaluasi hasil klasterisasi digunakan untuk menilai kualitas klasterisasi dengan jumlah kluster sebanyak dua. Evaluasi hasil klasterisasi dilakukan menggunakan dua metode, yaitu Silhouette Score dan Davies-Bouldin Index untuk memastikan bahwa kluster yang terbentuk menghasilkan kluster dengan tingkat kemiripan yang tinggi dalam kluster dan keterpisahan antar kluster yang jelas.

Berdasarkan perhitungan evaluasi hasil klasterisasi menggunakan Silhouette Score menghasilkan nilai sebesar 0,966. Nilai tersebut sangat mendekati 1, menunjukkan bahwa kluster yang terbentuk memiliki tingkat kemiripan yang sangat tinggi antar data dalam setiap kluster. Selain itu, nilai yang mendekati 1 juga menunjukkan bahwa jarak antar kluster jauh dan hampir tidak ada tumpang tindih antar kelompok.

Selain itu, perhitungan menggunakan Davies-Bouldin Index diperoleh nilai sebanyak 0,231. Nilai tersebut termasuk dalam kategori sangat baik, karena metode ini menganggap nilai yang lebih kecil adalah nilai kluster yang lebih baik. Indeks pada metode ini berfungsi untuk mengukur rasio antara penyebaran dalam kluster dan jarak antar pusat kluster.

3.6. Analisis Hasil Kluster

	combined figures (kg/capita/year)	Household estimate (tonnes/year)	Retail estimate (tonnes/year)	Food service estimate (tonnes/year)
Cluster				
0	127	1927346	350669	663016
1	110	80203188	21897896	51578281
Cluster				
0	212			
1	2			

Gambar 8. Karakteristik Setiap Kluster yang Terbentuk dari Proses Klasterisasi

Gambar 8 menampilkan karakteristik setiap kluster yang diperoleh dari proses klasterisasi terhadap data karakteristik limbah makanan menggunakan algoritma K-Means. Berdasarkan hasil klasterisasi tersebut, terbentuk dua kluster yaitu kluster 0 dan kluster 1. Kluster 0 terdiri atas 212 data, sedangkan kluster 1 terdiri atas 2 data. Pada kluster 0 limbah makanan yang dihasilkan dari sektor rumah tangga, retail, serta bisnis makanan dan minuman cepat saji dengan satuan kg per kapita per tahun menunjukkan nilai yang lebih besar daripada kluster 1, yaitu 127 pada kluster 0 dan 110 pada kluster 1. Sedangkan pada limbah makanan yang dihasilkan dari sektor rumah tangga, retail, serta bisnis makanan dan minuman cepat saji dengan satuan ton per tahun menunjukkan bahwa nilai pada kluster 1 lebih besar daripada kluster 0. Limbah makanan yang dihasilkan dari sektor rumah tangga dengan satuan ton per tahun menunjukkan nilai 80.203.188. Limbah makanan yang dihasilkan dari sektor retail dengan satuan ton per tahun menunjukkan nilai 21.897.896. Sedangkan, limbah makanan yang dihasilkan dari sektor bisnis makanan dan minuman cepat saji dengan satuan ton per tahun menunjukkan nilai 51.578.281.

Berdasarkan rata-rata nilai masing-masing variabel, terdapat perbedaan yang signifikan antara kedua kluster. Perbedaan tersebut, khususnya pada variabel produksi tahunan yang menunjukkan bahwa variabel tersebut memiliki selisih nilai rata-rata lebih besar dibandingkan variabel produksi per kapita, sehingga dapat diartikan bahwa produksi tahunan memiliki kontribusi yang lebih banyak dibandingkan variabel produksi per kapita dalam pembentukan kluster.

4. Kesimpulan

Berdasarkan seluruh tahapan analisis yang telah dilakukan, mulai dari pengumpulan data, pra-pemrosesan data, penerapan algoritma K-Means, evaluasi hasil klasterisasi, hingga analisis hasil klaster, dapat disimpulkan bahwa data limbah makanan yang digunakan dalam penelitian, secara alami membentuk dua klaster optimal dengan perbedaan karakteristik limbah makanan yang signifikan.

Hasil klasterisasi menggunakan algoritma K-Means dengan jumlah klaster optimal, menunjukkan bahwa klaster 0 terdiri atas data yang memiliki karakteristik limbah makanan dengan tingkat produksi tahunan yang relatif lebih rendah dibandingkan klaster 1. Sementara itu, klaster 1 terdiri atas jumlah data yang jauh lebih sedikit, namun menunjukkan karakteristik dengan tingkat produksi limbah makanan tahunan yang jauh lebih besar. Perbedaan karakteristik klaster yang terbentuk menunjukkan bahwa pemisah klaster lebih dipengaruhi total limbah makanan tahunan dibandingkan per kapita.

Dengan demikian, penelitian ini membuktikan bahwa algoritma K-Means efektif digunakan untuk mengidentifikasi pola tersembunyi pada data limbah makanan sehingga menghasilkan kelompok yang memiliki karakteristik limbah makanan yang berbeda secara signifikan. Oleh karena itu, hasil analisis ini dapat dimanfaatkan sebagai dasar dalam perumusan strategi pengelolaan limbah makanan yang lebih terarah dan tepat sasaran guna mendukung peningkatan efisiensi pengelolaan sampah.

Namun, penelitian ini memiliki keterbatasan pada data, variabel dan metode yang digunakan. Data yang digunakan masih kurang lengkap, ada beberapa negara yang tidak memiliki data karakteristik limbah makanan. Variabel yang digunakan dalam penelitian ini hanya mencakup jenis limbah makanan dari berbagai sektor yaitu sektor rumah tangga, ritel dan bisnis makanan dan minuman cepat saji, dimana variabel tersebut belum merepresentasikan karakteristik limbah makanan secara keseluruhan. Algoritma yang digunakan hanya algoritma K-Means yang memiliki kekurangan dibandingkan algoritma klasterisasi lainnya. Oleh karena itu, dibutuhkan data yang lebih banyak, penambahan variabel yang dapat merepresentasikan karakteristik limbah makanan secara lebih luas, serta melakukan klasterisasi pada data limbah makanan menggunakan algoritma lain seperti DBSCAN, Hierarchical Clustering, atau lainnya kemudian membandingkannya dengan hasil penerapan algoritma pada penelitian ini untuk mendapatkan hasil klasterisasi yang lebih optimal.

Daftar Rujukan

- [1] S. C. Lestari, A. Halimatussadiyah, F. Ekonomi, and U. Indonesia, "Kebijakan Pengelolaan Sampah Nasional : Analisis Pendorong Food Waste di Tingkat Rumah Tangga," 2015.
- [2] D. P. Sari, W. Buana, M. Febri, and M. Sari, "E ISSN : 2809-4069 Implementasi Data Mining pada Penjualan Barang dengan Teknik K Means," vol. 5, no. 1, pp. 106–112, 2025.
- [3] L. F. A. Zahroh, N. Rahaningsih, and R. D. Dana, "KLASTERISASI DATA KEGEMARAN MEMBACA MENGGUNAKAN ALGORITMA K-MEANS DI SMA AL-ISLAM CIREBON," vol. 8, no. 3, pp. 2692–2698, 2024.
- [4] C. Darmawan, Y. Setiyawan, R. A. Prasetyo, S. Khodijah, and Qurrota'Ayyun, "G-Tech : Jurnal Teknologi Terapan," vol. 8, no. 1, pp. 349–358, 2024.
- [5] R. Nugraha, N. Suarna, I. Ali, and D. Rohman, "OPTIMASI PENGELOLAAN SAMPAH MELALUI MODEL PENGELOMPOKAN DENGAN ALGORITMA K-MEANS," vol. 13, no. 1, pp. 646–652, 2025.
- [6] M. Nurrohman, Maimunah, and P. Sukmasetya, "Jurnal Sistem Klasterisasi Volume Sampah Organik di Kota Magelang menggunakan K-Means," vol. 5, pp. 146–153, 2023.
- [7] I. Mahendrasyah, A. Diana, R. C, and D. Mahdiana, "Implementation of K-Means Algorithm for Air Pollution Standards Index Clustering," vol. 14, no. 2, pp. 146–156, 2024.
- [8] P. Sakinah, A. E. Syaputra, Z. Rahman, M. Fajri, and H. F. Rachmansyah, "Optimalisasi Akreditasi Perguruan Tinggi dengan Orkestrasi Business Intelligence Berbasis K-Means dan OLAP," vol. 5, no. 3, pp. 550–561, 2025.
- [9] M. A. Septianto, A. Faqih, and A. R. Rinaldi, "KLASTERISASI DATA PRODUKSI PERTANIAN DI KABUPATEN CIREBON DENGAN ALGORITMA K-," vol. 13, no. 2, 2025.
- [10] E. R. Wedowati, F. S. Rejeki, D. Puspitasari, K. Handarini, and R. Rahmiati, "Jurnal Pengabdian UNDIKMA:," vol. 6, no. 2, pp. 350–358, 2025.
- [11] M. Amelia, A. Faqih, A. R. Rinaldi, and K. Sosial, "PENERAPAN METODE K-MEANS CLUSTERING DALAM PEMETAAN KEMISKINAN KABUPATEN / KOTA DI TEPAT," vol. 13, no. 2, 2025.
- [12] R. B. Trianto, A. S. Nugroho, E. Supriyadi, and J. Tengah, "Klasterisasi Menggunakan Algoritma K-Means dan Elbow pada Opini Masyarakat Tentang Kebijakan Sekolah Luring Tahun 2022," pp. 1–13, 2023.
- [13] F. Aviliana and P. Hendikawati, "Algoritma K-Means dan Analisis Komponen Utama untuk Mengatasi Multikolinearitas pada Pengelompokan Kabupaten Tertinggal," vol. 10, no. 3, pp. 294–306, 2025.
- [14] E. M. Fitri, R. R. Suryono, and A. Wantoro, "KLASTERISASI DATA PENJUALAN BERDASARKAN WILAYAH MENGGUNAKAN METODE K-MEANS PADA PT XYZ," vol. 11, no. 2, pp. 157–168, 2023.
- [15] I. Alfian, "Penerapan Metode K-Means Dalam Melakukan Pengelompokan Bencana Alam di Indonesia Dilakukan dengan Memanfaatkan Teknik Text Mining," pp. 139–147.