IMPLEMENTASI CLUSTERING UNTUK PENGELOMPOKAN TOTAL KONSUMSI KOPI DOMESTIK BERDASARKAN NEGARA MENGGUNAKAN ALGORITMA K-MEANS

Syamsa Shahira Julyinda, Eva Mahdyta Kiswana
Sistem Informasi, Institut Teknologi Telkom Purwokerto
Jl. DI Panjaitan No.128, Purwokerto Selatan, Indonesia
21103078@ittelkom-pwt.ac.id

ABSTRAK

Penelitian ini berfokus pada implementasi algoritma *K-Means* untuk mengelompokkan negara berdasarkan total konsumsi kopi domestik. Data yang digunakan dalam penelitian ini diambil dari 25 negara yang tersedia di situs Kaggle.com. Latar belakang penelitian ini adalah untuk memahami pola konsumsi kopi di berbagai negara dengan tujuan merumuskan strategi pemasaran dan distribusi kopi yang lebih efektif, khususnya untuk pasar Indonesia. Penelitian ini melakukan pengujian menggunakan metode *Elbow* untuk menentukan jumlah *cluster* optimal dan evaluasi dengan *Silhouette Score* untuk menilai kualitas *clustering*. Hasil pengelompokan menunjukkan terbentuknya dua *cluster* utama, satu *cluster* terdiri dari 24 negara dengan konsumsi kopi rendah hingga sedang, sementara Brazil dengan konsumsi kopi yang sangat tinggi membentuk *cluster* tersendiri. *Silhouette Score* rata-rata sebesar 0.91 menunjukkan bahwa kualitas *clustering* yang dihasilkan sangat baik, yang dapat dijadikan dasar dalam pengambilan keputusan terkait strategi pemasaran dan distribusi kopi.

Kata kunci : *Clustering, K-Means, Konsumsi Kopi, Principal Component Analysis (PCA), Silhouette Score*

1. PENDAHULUAN

Tanaman kopi, salah satu komoditas tropis yang penting, telah memberikan kontribusi signifikan terhadap perekonomian banyak negara, khususnya negara-negara penghasil kopi. Pertumbuhannya yang optimal dipengaruhi oleh faktor-faktor seperti daerah persebaran, curah hujan, angin dan iklim. Hal ini menjadikan kopi sebagai komoditas ekspor yang menguntungkan sejak masa kolonial[1].

Konsumsi kopi global pun terus meningkat, mencapai 166 juta karung pada tahun 2021 menurut International Coffee Organization (ICO), menunjukkan tren yang stabil[2].

Namun, pemahaman yang lebih mendalam tentang pola konsumsi kopi di tingkat negara sangat penting untuk merumuskan strategi yang efektif di pasar kopi. Perkembangan kopi mendorong minat dan kebutuhan masyarakat untuk mengkonsumsinya. Hal tersebut juga mendasari pertumbuhan jumlah kafe dan kedai terus meningkat tiap tahun. Kondisi menciptakan tantangan masing-masing bagi distribusi kopi dari petani hingga ke distributor yang sangat mempengaruhi tinggi rendahnya harga kopi. Fluktuasi harga yang terjadi mempengaruhi harga produk hasil olahan kopi para konsumen[3].

Untuk memahami pola konsumsi kopi di berbagai negara, diperlukan analisis yang komprehensif. Salah satu metode yang dapat digunakan adalah *clustering*. *Clustering* adalah proses pengelompokkan data yang tidak teratur dengan memperhatikan kesamaan antara satu data dengan data lainnya. Proses *clustering* ini bertujuan untuk mengumpulkan data yang serupa dalam satu kelompok, tanpa membiarkan data tersebut muncul di kelompok lain[4].

Dengan melakukan *clustering*, kita dapat mengidentifikasi kelompok negara dengan karakteristik konsumsi kopi yang mirip, sehingga memungkinkan pengembangan strategi pemasaran yang lebih efektif. *Clustering* juga dapat membantu dalam merancang kebijakan yang tepat dalam distribusi dan pengelolaan persediaan kopi di pasar global.

Penelitian ini menggunakan metode K-Means, yang merupakan salah satu algoritma *clustering* yang paling populer dan mudah diimplementasikan. Algoritma ini bekerja dengan mengelompokkan data berdasarkan kesamaan tanpa panduan awal (*unsupervised*). Algoritma ini membagi seluruh data ke dalam kelompok-kelompok yang memiliki kemiripan, algoritma ini menghitung jarak antara setiap data dengan pusat data (*centroid*) untuk menilai seberapa mirip data tersebut. Metode ini bertujuan untuk meminimalkan variassi dalam satu kelompok (*cluster*) dan memaksimalkan perbedaan dengan kelompok lainnya, sehingga pengelompokan data menjadi lebih efektif[5].

Implementasi *clustering* ini dapat digunakan untuk mengelompokkan negara-negara berdasarkan total konsumsi kopi domestik. *Algoritma K-Means* digunakan untuk membagi data ke dalam k kelompok (*clusters*), di mana setiap negara akan masuk ke dalam *cluster* berdasarkan kesamaan total konsumsi kopi domestiknya[6].

2. TINJAUAN PUSTAKA

2.1. *Clustering*

Clustering adalah salah satu metode dalam data mining yang bersifat tanpa arahan (*unsupervised*). Dalam proses pengelompokan data, terdapat dua jenis

clustering yang sering digunakan, yaitu *hierarchical* (hirarki) *clustering* dan *non-hierarchical* (non-hirarki) *clustering*[7].

Menurut Metisen dan Sari (2015), *clustering* merupakan metode yang digunakan untuk membagi data menjadi beberapa kelompok berdasarkan kesamaan-kesamaan yang sudah ditentukan sebelumnya. *Cluster* merupakan sekelompok objek data yang memiliki kesamaan satu sama lain di dalam *cluster* yang sama dan memiliki perbedaan dengan objek-objek di cluster lain. Objek-objek ini akan dikelompokkan ke dalam satu atau lebih cluster, sehingga objek-objek dalam satu cluster memiliki tingkat kesamaan yang tinggi antara satu dengan lainnya[6].

2.2. *Elbow Method*

Metode *Elbow* adalah teknik yang digunakan untuk menentukan jumlah *cluster* optimal dengan cara memeriksa persentase perbandingan antara jumlah *cluster* (k) yang membentuk siku pada suatu titik[8].

Metode ini membantu menentukan jumlah cluster yang paling tepat dengan memilih nilai cluster awal, lalu menambahkannya secara bertahap untuk membentuk model data yang optimal. Persentase perhitungan yang dihasilkan digunakan sebagai pembanding antara jumlah cluster yang ditambahkan, di mana hasil persentase yang berbeda dari setiap nilai cluster dapat ditampilkan dalam bentuk grafik untuk memudahkan interpretasi[9].

2.3. Algoritma *K-Means*

K-means adalah algoritma yang sederhana, dan ukuran kesamaan memainkan peran penting dalam proses *clustering* dengan menggunakan *K-means*[10].

Algoritma ini mengelompokkan data dengan karakteristik yang serupa ke dalam satu *cluster*. Langkah pertama dari algoritma *K-Means* adalah menentukan jumlah *cluster* dan menginisiasi titik pusat awal. Langkah kedua adalah memasukkan semua data ke dalam *cluster* terdekat. Kedekatan objek ditentukan oleh jarak menggunakan teori jarak *Euclidean* dengan rumus sebagai berikut[11]:

$$d(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \cdots + (x_{ki} - x_{kj})^2} \quad (1)$$

Keterangan:

$d(i,j)$: jarak objek ke-i ke pusat *cluster* j
x_{ki} : data ke-i pada atribut data ke-k
x_{kj} : titik pusat ke-j pada atribut ke-k

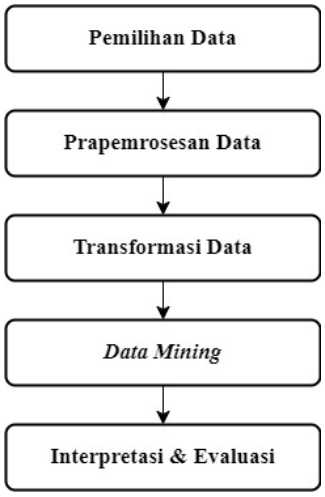
Langkah ketiga adalah menghitung ulang titik pusat menggunakan rata-rata data dalam satu cluster. Langkah terakhir adalah mengulangi langkah kedua dan ketiga sampai konvergen, yaitu sampai tidak ada lagi perpindahan objek dari satu cluster ke cluster lainnya.

2.4. *Silhouette Score*

Pengujian *silhouette score* dilakukan dengan cara membagi data ke dalam beberapa kelompok dengan jumlah kluster yang berbeda-beda, kemudian membandingkan nilai *silhouette score* untuk setiap pengelompokan tersebut. *Silhouette score* yang tinggi mengindikasikan bahwa *clustering* yang dilakukan baik, sementara nilai yang rendah menunjukkan adanya penyebaran data yang tidak homogen di dalam satu atau lebih kelompok. Oleh karena itu, pengujian *silhouette score* membantu dalam menentukan jumlah kluster yang optimal untuk melakukan *clustering* pada suatu dataset[12].

3. METODE PENELITIAN

Diagram alir yang ditunjukkan pada Gambar 1 menggambarkan metodologi yang digunakan dalam penelitian ini, yaitu *Knowledge Discovery in Database* (KDD). KDD merupakan sebuah metodologi penelitian yang digunakan dalam proses *data mining*. Tujuan utama KDD adalah menemukan dan mengidentifikasi pola dalam basis data guna memperoleh informasi baru. Tahapan KDD meliputi pemilihan data (*selection*), prapemrosesan (*preprocessing*), transformasi data (*transformation*), penambangan data (*data mining*), serta interpretasi dan evaluasi hasil (*interpretation & evaluation*)[13].



Gambar 1. Alur Penelitian

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder. Data yang digunakan merupakan data konsumsi kopi yang berjumlah 54 data berisi id, nama negara, jenis biji kopi, konsumsi kopi setiap tahun, dan total konsumsi konsumsi. Dataset ini diambil dari situs kaggel.com yang diunduh pada tanggal 10 Juni 2024.

3.2. Pemilihan Data

Pada tahap ini, dilakukan pemilihan data yaitu menyeleksi dataset yang awalnya berisi 54 negara yang terdapat nama negara, jenis biji kopi, konsumsi kopi setiap tahun, dan total konsumsi domestik diseleksi menjadi 25 negara yang terdapat nama negara, total konsumsi domestik dan penambahan *latitude* serta *longitude*.

3.3. Prapemrosesan Data

Pada tahap ini, dilakukan pembersihan data yang meliputi pengecekan nilai yang hilang, pengecekan data yang duplikat, dan penghapusan atribut yang tidak relevan dengan tujuan data mining.

3.4. Transformasi Data

Pada tahap ini, data dinormalisasi menggunakan teknik standarisasi dan kemudian dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA). Standarisasi data diperlukan sebelum menerapkan PCA agar interval data menjadi lebih proporsional[14]. PCA digunakan untuk meningkatkan kualitas hasil *clustering*.

3.5. *Data Mining*

Data mining adalah proses yang memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin (*machine learning*) untuk menggali dan menemukan informasi berharga serta pengetahuan yang relevan dari berbagai *database*[7].

Salah satu teknik yang digunakan adalah *clustering*, yaitu mengelompokkan data berdasarkan kesamaan. Algoritma *K-Means* adalah salah satu algoritma *clustering* yang populer[15].

Untuk menentukan jumlah kelompok yang optimal dalam *K-Means*, dalam penelitian ini dapat menggunakan metode *elbow*[16].

3.6. Interpretasi dan Evaluasi Hasil

Dengan melakukan interpretasi untuk pengambilan keputusan, penelitian ini dapat memastikan bahwa hasil *cluster* yang paling optimal akan dievaluasi menggunakan *Silhouette Score*. Silhouette Score merupakan metode yang sangat efektif untuk menilai kualitas clustering.

4. HASIL DAN PEMBAHASAN

4.1. Pemilihan Data

Pada tahap ini, dilakukan pemilihan data agar relevan dengan tujuan *data mining*.

Tabel 1. Dataset Hasil Seleksi

Country	Total_domestic_consumption	latitude	longitude
Angola	46500000	-11,2027	17,8739
Bolivia (Plurinational State of)	75180000	-16,2902	-63,5887
Brazil	27824700000	-14,235	-51,9253
Burundi	3412020	-3,3731	29,9189
Ecuador	381540000	-1,8312	-78,1834
Indonesia	4920480000	-0,7893	113,9213
Madagascar	588705960	-18,7669	46,8691
Malawi	2340000	-13,2543	34,3015
Papua New Guinea	3608400	-6,31499	143,9556
Paraguay	35100000	-23,4425	-58,4438
Peru	402000000	-9,19	-75,0152
Rwanda	2139960	-1,9403	29,8739
Timor-Leste	294000	-8,8742	125,7275
Zimbabwe	8595960	-19,0154	29,1549
Congo	5360040	-0,228	15,8277
Cuba	384006000	21,5218	-77,7812
Dominican Republic	642823380	18,7357	-70,1627
Haiti	600600000	18,9712	-72,2852
Philippines	2807280000	12,8797	121,774
Tanzania	76425060	-6,36903	34,88882
Zambia	991920	-13,1339	27,8493
Cameroon	143450940	7,3697	12,3547
Central African Republic	24794400	6,6111	20,9394
Colombia	2536776384	4,5709	-74,2973
Costa Rica	665335200	9,7489	-83,7534

4.2. Prapemrosesan Data

Tahap selanjutnya merupakan tahap *preprocessing*. Pada tahap ini akan dilakukan pembersihan data. Dari 25 atribut sebelumnya akan diseleksi apakah terdapat *missing value* atau tidak. Data dikatakan *missing value* apabila apabila atribut-atribut dalam *dataset* tidak berisi nilai atau kosong.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 4 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Country                   25 non-null     object
 1   Total_domestic_consumption  25 non-null     int64
 2   latitude                  25 non-null     float64
 3   longitude                 25 non-null     float64
dtypes: float64(2), int64(1), object(1)
memory usage: 928.0+ bytes
```

Gambar 2. Pengecekan *missing value* dan duplikat pada tools Google Colab

Dapat diketahui dari Gambar 2 bahwa tidak terdapat *missing value* dan duplikat pada *dataset*.

4.3. Transformasi Data

Setelah pembersihan data, tahap berikutnya dilakukan transformasi data. Pada tahap ini dilakukan standarisasi dan reduksi data. Data konsumsi kopi domestik dari 25 negara yang dipilih dinormalisasi menggunakan teknik standarisasi dengan *StandardScaler* dari library *scikit-learn* menggunakan Google Colab. Proses ini mengubah distribusi nilai total konsumsi kopi domestik sehingga memiliki *mean* nol dan varians satu. Hasil standarisasi data dapat dilihat pada tabel berikut:

Tabel 2. Hasil standarisasi data

Total_domestic_consumption_scaled
-0,300849463
-0,295590824
4,7924397
-0,308749887
-0,239417995
0,592822176
-0,201432952
-0,308946447
-0,308713879
-0,302939717
-0,235666539
-0,308983126
-0,309321593
-0,307799382
-0,308392706
-0,23896584
-0,191510218
-0,199252113
0,205355073
-0,295362535
-0,309193626
-0,283072963
-0,304829307
0,155756714
-0,18738255

Setelah data dinormalisasi pada tabel 2, dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) untuk mengurangi kompleksitas data sambil mempertahankan sebanyak mungkin informasi variabilitas dari data asli. PCA digunakan dengan tujuan untuk mempermudah proses *clustering* dengan *algoritma K-Means*, berikut adalah hasil reduksi dimensi menggunakan *Principal Component Analysis* (PCA):

Tabel 3. Hasil reduksi dimensi menggunakan *Principal Component Analysis* (PCA)

PCA_component
-0,300849463
-0,295590824
4,7924397
-0,308749887
-0,239417995
0,592822176
-0,201432952
-0,308946447
-0,308713879
-0,302939717
-0,235666539
-0,308983126
-0,309321593
-0,307799382
-0,308392706
-0,23896584
-0,191510218
-0,199252113
0,205355073
-0,295362535
-0,309193626
-0,283072963
-0,304829307
0,155756714
-0,18738255

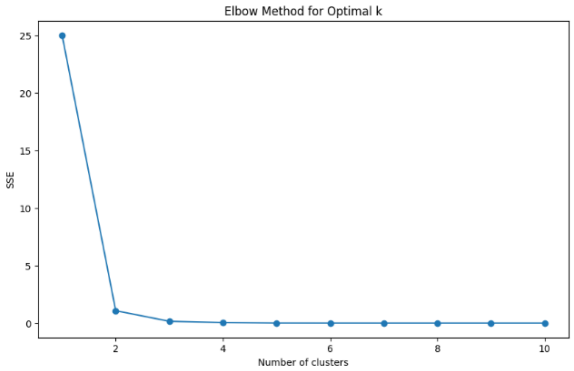
4.4. *Data Mining*

Data mining adalah proses penerapan model untuk mencapai tujuan pengelompokan data (*clustering*). Pada tahap ini, *clustering* akan dilakukan menggunakan *algoritma k-means* dengan menggunakan jarak *euclidean*. Jumlah *cluster* akan ditentukan dengan metode *elbow* berdasarkan nilai *Sum of Square Error* (SSE).

Tabel 4. Hasil Nilai SSE dan selisih

Klaster	Nilai SSE	Selisih
K=1	25.000000000000004	0
K=2	**1.0755434607332155**	**23.92445653926679**
K=3	0.15677218330280848	0.918771277430407
K=4	0.04346319192751054	0.11330899137529794
K=5	0.005333132064165235	0.038130059863345305
K=6	0.0021453566610930893	0.0031877754030721457
K=7	0.0009153580533638835	0.0012299986077292058
K=8	0.0003471878697902369	0.0005681701835736466
K=9	0.0001968103812401 2124	0.00015037748855011565
K=10	0.00008006119342189087	0.00011674918781823037

Tabel 4, merupakan hasil nilai SSE yang telah dilakukan pada K=1 sampai K=10 yang dapat disimpulkan selisih paling besar pada K=2.



Gambar 3. Hasil menentukan *cluster* optimal menggunakan *elbow method*

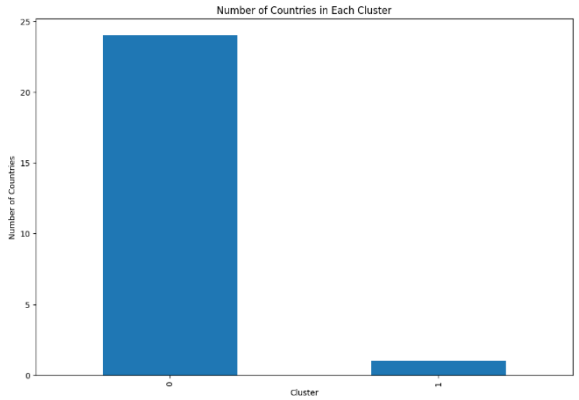
Dapat dilihat dari Gambar 3, grafik membentuk siku pada K=2. Maka dapat disimpulkan bahwa nilai K yang paling optimal adalah pada saat nilai SSE memiliki selisih paling besar dan grafik menunjukkan penurunan paling signifikan, yaitu pada K=2. Selanjutnya dilakukan *clustering* dengan rekomendasi parameter yaitu, K=2.

Tabel 5. Hasil *clustering*

Country	Total_domestic_consumption	latitude	longitude	cluster
Angola	46500000	**-11,2027**	**17,8739**	0
Bolivia (Plurinational State of)	75180000	**-16,2902**	**-63,5887**	0

Country	Total_domestic_consumption	latitude	longitude	cluster
Brazil	2,78E+10	**-14,235**	**-51,9253**	1
Burundi	3412020	**-3,3731**	**29,9189**	0
Ecuador	3,82E+08	**-1,8312**	**-78,1834**	0
Indonesia	4,92E+09	**-0,7893**	**113,9213**	0
Madagascar	5,89E+08	**-18,7669**	**46,8691**	0
Malawi	2340000	**-13,2543**	**34,3015**	0
Papua New Guinea	3608400	**-6,31499**	**143,9556**	0
Paraguay	35100000	**-23,4425**	**-58,4438**	0
Peru	4,02E+08	**-9,19**	**-75,0152**	0
Rwanda	2139960	**-1,9403**	**29,8739**	0
Timor-Leste	294000	**-8,8742**	**125,7275**	0
Zimbabwe	8595960	**-19,0154**	**29,1549**	0
Congo	5360040	**-0,228**	**15,8277**	0
Cuba	3,84E+08	**21,5218**	**-77,7812**	0
Dominican Republic	6,43E+08	**18,7357**	**-70,1627**	0
Haiti	6,01E+08	**18,9712**	**-72,2852**	0
Philippines	2,81E+09	**12,8797**	**121,774**	0
Tanzania	76425060	**-6,36903**	**34,88882**	0
Zambia	991920	**-13,1339**	**27,8493**	0
Cameroon	1,43E+08	**7,3697**	**12,3547**	0
Central African Republic	24794400	**6,6111**	**20,9394**	0
Colombia	2,54E+09	**4,5709**	**-74,2973**	0
Costa Rica	6,65E+08	**9,7489**	**-83,7534**	0

Dari hasil *clustering* pada tabel 3, kemudian dilakukan visualisasi dengan diagram batang dan peta interaktif, sebagai berikut:



Gambar 4. Visualisasi hasil *clustering* dengan diagram batang

Diagram batang menunjukkan jumlah negara dalam setiap *cluster*, memberikan gambaran yang jelas bahwa sebagian besar negara termasuk dalam *Cluster* 0, sementara hanya satu negara yang berada di *Cluster* 1.

Gambar 5. Visualisasi hasil *clustering* dengan peta interaktif

Peta interaktif memberikan representasi geografis dari hasil *clustering*, memperlihatkan distribusi negara berdasarkan konsumsi kopi domestik mereka.

4.5. Interpretasi dan Evaluasi Hasil

Setelah melakukan *clustering* menggunakan *algoritma K-Means* dengan jumlah *cluster* optimal ditentukan oleh Metode *Elbow*, diperoleh dua *cluster* untuk data total konsumsi kopi domestik dari 25 negara. Evaluasi hasil *clustering* ini menggunakan *Silhouette Score* untuk menilai kualitas pengelompokan ini dengan menggunakan Google Colab. Berikut adalah persamaan yang digunakan dan hasil dari *Silhouette Score*:

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (2)$$

Keterangan:
$s(i)$: Silhouette Score
$a(i)$: jarak rata-rata antara titik data i dengan semua titik data lain dalam cluster yang sama
x_{kj} : jarak rata-rata antara titik data i dengan semua titik data di cluster terdekat yang berbeda.

```
Silhouette Score: 0.9248027966977631
```

Gambar 5. Hasil *Silhouette Score* menggunakan Google Colab

Silhouette Score memberikan nilai rata-rata sebesar 0.91, yang menunjukkan bahwa hasil *clustering* memiliki pemisahan yang baik antara *cluster* yang satu dengan yang lainnya. *Cluster* 0 terdiri dari 24 negara yang memiliki total konsumsi kopi domestik yang relatif rendah hingga sedang, sedangkan *Cluster* 1 hanya terdiri dari 1 negara yaitu Brazil, yang memiliki total konsumsi kopi domestik yang sangat tinggi dibandingkan negara-negara lain.

5. KESIMPULAN DAN SARAN

Algoritma K-Means berhasil mengelompokkan 25 negara berdasarkan total konsumsi kopi domestik menjadi dua *cluster* yang berbeda. Satu negara, yaitu Brazil, memiliki konsumsi kopi yang sangat tinggi dan terpisah dalam *cluster* tersendiri. Evaluasi *clustering* menggunakan *Silhouette Score* menunjukkan nilai rata-rata sebesar 0.91, yang menandakan bahwa pengelompokan yang dilakukan memiliki pemisahan yang baik antara *cluster* yang satu dengan yang lainnya. *Cluster* 0 terdiri dari 24 negara dengan

konsumsi kopi rendah hingga sedang, sedangkan *Cluster* 1 hanya terdiri dari Brazil dengan konsumsi kopi yang sangat tinggi. Hasil *clustering* ini dapat digunakan untuk mengembangkan strategi pemasaran yang lebih efektif dengan fokus pada negara-negara di masing-masing cluster sesuai dengan karakteristik konsumsi mereka.

Disarankan untuk melakukan penelitian lanjutan dengan dataset yang lebih besar dan pada penelitian selanjutnya dapat mencoba menggunakan metode *clustering* lain seperti DBSCAN atau *Agglomerative Clustering* untuk membandingkan hasil dan mencari metode yang paling optimal.

DAFTAR PUSTAKA

[1] M. W. Chandra and N. S. Fadjar, "Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Konsumsi Kopi Di Malang Tahun 2022 (Studi Kasus: JL. Ikan Tombro - Kota Malang)," *JOURNAL OF DEVELOPMENT ECONOMIC AND SOCIAL STUDIES*, vol. 2, no. 1, pp. 1–12, 2023, doi: 10.21776/jdess.2023.02.1.8.

[2] "International Coffee Organization (ICO), 'World coffee consumption,' International Coffee Organization, 2021.," *International Coffee Organization*, 2024. Accessed: Aug. 21, 2024. [Online]. Available: https://www.ico.org.

[3] A. Ghifari, I. Ihsan, and L. Muh. Asfan Mujahid, "Arahan Pola Distribusi Pemasaran Kopi Pada Warung Kopi Dan Kafe Di Kota Makassar," *Jurnal WKM*, vol. 10, no. 1, pp. 1–8, 2022.

[4] H. Pratiwi and A. P. Wahyu Wibowo, "Implementasi Algoritma K-Means Untuk Mengklaster Kelompok Sektor Perkebunan Di Indonesia," *JOISIE Journal Of Information System And Informatics Engineering*, vol. 6, no. 1, pp. 1–10, 2022.

[5] A. Rohmah, F. Sembiring, and A. Erfina, "Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung)," *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)* , pp. 1–9, 2021, [Online]. Available: https://www.alfasoleh.com/2019/11/k-means-clustering-contoh

[6] S. A. Rahma, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Journal of Information Technology Research*, vol. 1, no. 1, pp. 1–5, 2020.

[7] N. Karolina, "Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus : RSU.Bangkatan)," *JOURNAL OF INFORMATION AND TECHNOLOGY UNIMOR*, pp. 1–7, 2021, [Online]. Available: www.kaputama.ac.id

[8] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrogaman Python," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 5, no. 1, pp. 1–7, 2021.

[9] A. R. Said, D. Arifianto, and H. A. Al Faruq, "Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Tanaman Pangan Dengan Algoritma Fuzzy C-Means Dan Metode Elbow," *Jurnal Smart Teknologi*, vol. 2, no. 1, pp. 1–12, 2020.

[10] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pengelompokkan Data Obat Dengan Silhouette Coefficient," *Journal of Applied Informatics and Computing (JAIC)*, vol. 5, no. 2, pp. 1–8, 2021, [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[11] E. Rahmawati, B. Nurina Sari, and M. Jajuli, "Implementasi Algoritma K-Means Pada Pemetaan Daerah Terdampak Tanah Longsor Di Jawa Tengah," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 3, pp. 1–7, 2024.

[12] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)*, vol. 3, no. 1, pp. 1–11, 2024, doi: 10.58602/jima-ilkom.v3i1.26.

[13] Y. S. Siregar, D. Handoko, M. Khairani, N. I. Syahputri, and H. Harahap, "Implementasi Data Mining Klasifikasi Algoritma Chaid Dalam Menentukan Pola Penerima Mahasiswa Baru," *Digital Transformation Technology (Digitech) | e*, vol. 3, no. 2, pp. 1–12, 2023, doi: 10.47709/digitech.v3i2.3612.

[14] S. Dewi and M. A. I. Pakereng, "Implementasi Principal Component Analysis Pada K-Means Untuk Klasterisasi Tingkat Pendidikan Penduduk Kabupaten Semarang," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 4, pp. 1–10, Nov. 2023, doi: 10.29100/jipi.v8i4.4101.

[15] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, pp. 1–10, 2022, doi: 10.5281/zenodo.7052276.

[16] N. A. Maori, "Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering," *Jurnal SIMETRIS*, vol. 14, no. 2, pp. 1–11, 2023