

Analisis Ekspor Kopi Menggunakan Clustering K-Means dan Davies-Bouldin Index

Fenny Irawati^{1*}, Angelika Pratiwi Widya Nugroho², Arief Wibowo³

^{1,2,3}Fakultas Ekonomi dan Bisnis, Universitas Budi Luhur, Jakarta, Indonesia
Email: ¹2431600655@student.budiluhur.ac.id, ²2431600770@student.budiluhur.ac.id,
³arief.wibowo@budiluhur.ac.id
Penulis Korespondensi*

(received: 25-09-25, revised: 01-10-25, accepted: 16-10-25)

Abstrak

Indonesia merupakan salah satu produsen kopi terbesar di dunia, sehingga sektor ekspor menjadi pilar penting dalam perekonomian nasional. Aktivitas ekspor berperan dalam meningkatkan keuntungan, memperluas pangsa pasar, serta menjaga kestabilan harga komoditas dan nilai tukar. Penelitian ini mengkaji penerapan metode *K-Means Clustering* untuk menganalisis kinerja ekspor kopi berdasarkan negara tujuan. Data penelitian diperoleh dari catatan ekspor perusahaan Café Coffee pada periode 2023–2024, mencakup 40 negara tujuan beserta total kuantitas ekspor. Pengolahan data dilakukan melalui teknik *data mining clustering* dengan ukuran jarak *Euclidean Distance*. Hasil analisis menunjukkan bahwa algoritma *K-Means* berhasil mengelompokkan laba ekspor ke dalam tiga kategori, yaitu laba rendah, sedang, dan tinggi. Validasi model dilakukan menggunakan *Davies-Bouldin Index (DBI)* dengan nilai 0,422, yang mengindikasikan kualitas kluster yang baik dan dapat diterima.

Kata Kunci: Klasterisasi, Data Mining, K-Means, *Davies-Bouldin Index*, Segmentasi Pasar

Abstract

Indonesia is recognized as one of the world's leading coffee producers, making exports a central pillar of its national economy. Export activities play a crucial role in generating profits, expanding international markets, and maintaining the stability of both commodity prices and exchange rates. This study applies the *K-Means clustering* method to examine coffee export performance across different destination countries. The dataset was obtained from Café Coffee's export records during the 2023–2024 period, covering 40 countries along with their total export quantities. Data processing employed data mining clustering techniques with *Euclidean Distance* as the similarity measure. The findings reveal that the *K-Means* algorithm successfully classified export profits into three categories: low, medium, and high. Model validation using the *Davies-Bouldin Index (DBI)* resulted in a score of 0.422, which indicates that the clustering model achieved reliable and acceptable quality.

Keywords: Clustering, Data Mining, K-Means, *Davies-Bouldin Index*, Market Segmentation

1. PENDAHULUAN

Permasalahan ekspor kopi Indonesia terletak pada ketidakmerataan kontribusi antarnegara tujuan, di mana beberapa negara menyumbang pendapatan signifikan, sedangkan negara lain memiliki penyerapan pasar rendah, sehingga diperlukan identifikasi, segmentasi, dan prioritas pasar yang lebih efektif untuk mengoptimalkan manfaat ekonomi dan kesejahteraan petani.

Sejumlah penelitian sebelumnya telah menerapkan teknik klastering seperti K-Means untuk segmentasi data pertanian maupun perdagangan. Namun, kajian yang secara spesifik menyoroti pasar ekspor kopi masih terbatas. Selain itu, sebagian besar penelitian terdahulu hanya mengandalkan hasil klastering tanpa menambahkan evaluasi validitas kluster yang komprehensif. Keterbatasan tersebut dapat diatasi dalam penelitian ini mengintegrasikan algoritma K-Means dengan *Davies-Bouldin Index (DBI)* pengukuran validitas. Pendekatan ini diharapkan mampu menghasilkan pemetaan pasar yang lebih akurat, sekaligus menyediakan landasan analitis bagi penyusunan strategi ekspor yang berbasis bukti bagi pembuat kebijakan maupun eksportir.

Beberapa penelitian terdahulu telah memanfaatkan algoritma K-Means untuk pengelompokan data. Asmaya, Irawan, & Bahtiar (2024) melakukan penelitian mengenai optimalisasi penjualan produk pada Hopespace Coffee menggunakan algoritma metode K-Means Clustering. Hasil penelitian menunjukkan bahwa pengelompokan produk yang membantu perusahaan agar dapat menentukan strategi yang lebih tepat [1]. Hidayat & Kusniyati (2022) meneliti pengelompokan penjualan pada Cafe 47°Coffee dengan algoritma K-Means. Penelitian ini berhasil membagi data penjualan menjadi kelompok untuk memahami pola transaksi [2]. Hasan (2024) menilai kualitas kluster yang dihasilkan oleh algoritma K-Means dan DBSCAN menggunakan dua metrik utama, yaitu Silhouette Score dan Davies Bouldin Index (DBI), dan hasil penelitiannya menegaskan

adanya perbedaan signifikan dalam kualitas kluster antara kedua algoritma, sehingga masing-masing memiliki keunggulan dan keterbatasan tersendiri dalam penerapan analisis data [3]. Ros, Riad, & Guillaume (2023) menekankan bahwa *Davies-Bouldin Index* (DBI) klasik memiliki keterbatasan dalam menilai kualitas kluster secara optimal. Untuk mengatasi hal tersebut, mereka mengusulkan *Partitioning Davies-Bouldin Index* (PDBI) sebagai modifikasi yang mampu memberikan evaluasi kluster yang lebih akurat, efektif, dan representatif dibandingkan DBI standar [4]. Alfian, Kusrini, dan Hidayat (2024) melakukan *clustering* terhadap pegawai berdasarkan tingkat kedisiplinan menggunakan K-Means dengan evaluasi DBI hasil penelitian dapat membantu manajemen dalam penilaian kinerja pegawai [5].

Adapun perbedaan penelitian pada referensi diatas banyak fokus pada produk, penjualan, individu dan pegawai. Penelitian terdahulu bertujuan untuk membantu strategi pemasaran dan penilaian kinerja. Serta beberapa penelitian hanya memakai K-Means tanpa evaluasi, ada juga yang membandingkan dengan algoritma lain. Sementara penelitian ini menekankan pada negara tujuan ekspor kopi yang bertujuan untuk memetakan pasar ekspor kopi internasional. Penelitian ini secara khusus menggunakan K-Means dan *Davies-Bouldin Index* untuk menentukan kluster baik untuk dapat diterapkan.

Sebagian besar penelitian terdahulu lebih berfokus pada analisis regresi atau peramalan tren produksi kopi, namun belum banyak yang melakukan pemetaan pasar ekspor kopi berdasarkan negara tujuan menggunakan pendekatan *clustering*. Penelitian ini mengisi kesenjangan tersebut dengan menerapkan algoritma K-Means dan mengevaluasi kualitas hasilnya menggunakan *Davies-Bouldin Index* (DBI), sehingga diharapkan dapat memberikan gambaran yang lebih sistematis mengenai segmentasi pasar ekspor kopi Indonesia. Hal ini peneliti mengangkat topik ekspor kopi berdasarkan negara tujuan dengan proses yang dilakukan adalah *Clustering* dan dilanjutkan dengan evaluasi *Davies-Bouldin Index* (DBI). Fungsi dari dua istilah yang pertama mengidentifikasi minimalisasi varians intra-kluster dan yang kedua mengidentifikasi maksimalisasi jarak antar-kluster [4].

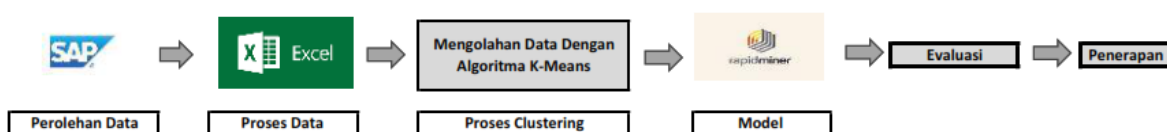
Penelitian terdahulu telah memberikan gambaran komprehensif mengenai berbagai metode pengelompokan beserta perkembangannya dari waktu ke waktu, dengan menguraikan karakteristik dan keterbatasan dua belas algoritma pengelompokan, termasuk beberapa representasi klasik dari kategori algoritma yang telah disebutkan sebelumnya, seperti k-means, BIRCH, dan DBSCAN [3]. Hidayat & Kusniyati (2022), dalam penelitiannya menggunakan metode *Data Mining Clustering* K-means yang di evaluasi dengan *Davies-Bouldin Index* (DBI) penggunaan pada software *RapidMiner* [2]. *Kmeans Clustering* adalah metode *data mining* atau metode menganalisa dari suatu data dengan dilanjutkan pada pemodelan dalam pengelompokan data. *Clustering* terdapat dua jenis data *clustering* yaitu *Hierarchical* dan *Non-Hierarchical* dan K-means termasuk kedalam metode data *clustering Non-Hierarchical* [4].

Penelitian ini memberikan informasi berbasis data mengenai prioritas negara tujuan ekspor kopi untuk mendukung kebijakan perdagangan dan strategi internasional Café Coffee. Melalui metode K-Means, negara tujuan dikluster berdasarkan potensi laba: kluster tinggi menjadi pasar utama yang perlu dipertahankan, kluster menengah dapat ditingkatkan melalui promosi atau perjanjian dagang, dan kluster rendah menjadi target ekspansi di masa depan. Pemetaan kluster ini juga mendukung pengelolaan logistik yang lebih efisien serta pengambilan keputusan untuk memaksimalkan keuntungan ekspor.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan metode kuantitatif dengan teknik clustering pada data mining. Algoritma K-Means diterapkan dan kualitas kluster yang dihasilkan dievaluasi menggunakan *Davies-Bouldin Index* (DBI). Pendekatan ini dipilih karena K-Means memiliki keunggulan dalam kesederhanaan perhitungan, efisiensi komputasi untuk dataset besar, serta kemampuan untuk memetakan objek berdasarkan kemiripan karakteristik. Evaluasi menggunakan DBI dilakukan untuk memastikan pemisahan kluster yang optimal serta keseragaman dalam setiap kluster. Tahapan bermula dari pengambilan data, pengumpulan dalam pemrosesan data, perhitungan K-Means kemudian dilanjutkan pemodelan serta pengujian penelitian dengan *Software RapidMiner*, evaluasi dari hasil pengelompokan yang di perhitungkan keakuratannya dengan *Davies-Bouldin Index* (DBI), lalu diakhiri pada penerapan hasil.



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan tahapan penelitian yang dilakukan sebagai berikut:

1. Data ekspor kopi di dapat dari system database SAP pada perusahaan
2. Data yang sudah diperoleh kemudian diolah di *Microsoft Excel*. Proses ini meliputi pembersihan data (data cleaning), normalisasi, pemilihan variabel yang relevan, serta menyiapkan dataset agar siap diproses lebih lanjut.
3. Data yang sudah siap selanjutnya diolah menggunakan algoritma K-Means *Clustering* dalam perhitungan di *Microsoft Excel*. Proses ini mengelompokkan negara tujuan ekspor kopi berdasarkan kesamaan karakteristik dalam *quantity*.
4. Pengolahan data selanjutnya dilakukan di software *RapidMiner* untuk menjalankan algoritma K-Means sehingga menghasilkan model pengelompokan yang bisa divisualisasikan.
5. Sebelum pemodelan, dilakukan analisis jumlah kluster ideal dengan *Elbow Method*.
6. Setelah pemodelan dan terbentuk kluster, dilakukan evaluasi dengan *Davies-Bouldin Index* (DBI). Evaluasi ini menentukan kualitas kluster, semakin rendah nilai DBI, semakin baik dalam pemisahan antar kluster.
7. Tahap terakhir adalah penerapan hasil *clustering*. Hasil kluster dapat digunakan oleh pemerintah atau perusahaan sebagai dasar pengambilan Keputusan dalam menentukan negara prioritas ekspor, strategi pemasaran, serta pemetaan pasar kopi internasional.

2.2 Pengolahan Data

Variabel yang digunakan dalam proses *clustering* meliputi volume ekspor kopi (*Cubic Meter*) pertumbuhan ekspor tahunan periode 2023-2024. Data diperoleh dari *software* SAP yang digunakan oleh Café Coffee dalam mengelola keuangannya. Data berupa hasil penjualan ekspor kopi pada periode tahun 2023-2024 dengan total pengiriman kopi ke 40 Negara Tujuan. Data populasi negara tidak digunakan sebagai variabel utama dalam *clustering*, melainkan sebagai analisis tambahan untuk interpretasi hasil pada ekspor kopi. Dataset yang digunakan dalam penelitian ekspor kopi tampak pada Tabel 1.

Tabel 1. Dataset ekspor kopi Café Coffee berdasarkan populasi negara [5]

No	Negara	Populasi	2023	2024
			Quantity	(Cubic Meter)
1	Afrika Selatan	61,673,081	440	858
2	Amerika Serikat	343,603,404	13,786	8,723
3	Argentina	46,337,520	250	450
4	Australia	26,958,054	283	535
5	Bangladesh	176,421,509	155	170
6	Belanda	17,722,333	5,230	7,364
7	Belarus	9,412,540	800	1,600
8	Brasil	218,803,058	740	450
9	China	1,424,381,924	1,980	3,450
10	Filipina	120,864,358	160	270
11	Gabon	2,532,885	150	180
12	Haiti	12,009,506	150	180
13	India	1,454,606,724	6,936	2,268
14	Inggris	69,570,331	158	150
15	Italia	58,518,843	4,246	4,648
16	Jamaika	2,823,199	120	170
17	Jepang	121,960,408	797	380
18	Jerman	83,199,069	770	776
19	Kanada	39,431,447	450	600
20	Kolombia	52,610,722	86	338
21	Korea Selatan	51,690,479	150	150
22	Laos	5,657,229	58	108

23	Liberia	5,657,229	110	110
24	Malaysia	35,028,030	2,520	3,450
25	Myanmar	55,336,783	80	80
26	Oman	4,780,706	58	43
27	Pantai Gading	32,711,547	19	108
28	Papua Nugini	10,701,144	2,660	2,560
29	Peru	35,015,825	120	120
30	Prancis	65,003,384	735	2,720
31	Qatar	2,757,220	206	343
32	Rusia	143,494,210	922	4,204
33	Singapura	6,089,541	3,326	2,387
34	Spanyol	47,420,024	78	58
35	Swiss	8,904,610	96	230
36	Taiwan	23,974,031	170	137
37	Thailand	71,953,054	686	1,165
38	Timor Leste	1,399,184	208	312
39	Uni Emirat Arab	37,989,969	388	841
40	Vietnam	100,103,979	1,822	729

2.3 Data Selection

Tahap ini berfokus pada pemilihan variabel yang relevan untuk proses *clustering*. Volume ekspor kopi (*cubic meter*) menggambarkan besarnya kontribusi negara tujuan. Pertumbuhan ekspor tahunan mengukur potensi pasar. Populasi negara tujuan digunakan untuk analisis tambahan, bukan variabel utama *clustering*. Pemilihan variabel ini didasarkan pada pertimbangan bahwa analisis *clustering* bertujuan memetakan negara tujuan berdasarkan besaran, nilai, dan dinamika ekspor.

Data selection adalah proses memilih data dari sumber data yang lebih besar untuk digunakan dalam analisis atau tugas tertentu. *Data selection* adalah langkah penting dalam proses data mining dan analisis data, di mana data yang tidak relevan atau tidak perlu dihilangkan untuk meningkatkan efisiensi dan efektivitas analisis, data diolah sedemikian rupa [6].

2.4 Data Cleaning (Pre-Processing)

Data cleaning dalam penelitian ini dilakukan untuk memastikan data yang digunakan bebas dari kesalahan dan inkonsistensi. Proses yang dilakukan antara lain menghapus negara tujuan ekspor yang memiliki data tidak lengkap dan menghilangkan data duplikat pada catatan ekspor tahunan. Format satuan variabel yaitu volume ekspor dalam cubic meter. Menghapus data yang dapat memengaruhi hasil *clustering*, data negara dengan volume ekspor sangat kecil dan tidak berkelanjutan.

Data *cleansing* atau *scrubbing*, berkaitan dengan pendeteksian dan penghapusan kesalahan dan ketidakkonsistenan dari data untuk meningkatkan kualitas data [7]. Masalah kualitas data dalam koleksi data tunggal, seperti file dan basis data, misalnya, karena salah eja selama entri data, informasi yang hilang atau data tidak valid lainnya. Ketika beberapa sumber data perlu diintegrasikan, misalnya, dalam gudang data, sistem basis data terfederasi atau sistem informasi berbasis web global, kebutuhan untuk pembersihan data meningkat secara signifikan. Ini karena sumber sering kali berisi data yang berlebihan dalam representasi yang berbeda. Untuk menyediakan akses ke data yang akurat dan konsisten, konsolidasi representasi data yang berbeda dan penghapusan informasi duplikat menjadi perlu dan Gudang data memerlukan dan menyediakan dukungan ekstensif untuk pembersihan data [4].

2.5 Data Transformation

Dalam penelitian ini, data terlebih dahulu ditransformasikan agar sesuai dengan algoritma K-Means. Proses transformasi dilakukan melalui normalisasi menggunakan *Min-Max Scaling*, sehingga seluruh variabel berada pada rentang nilai yang sama dan tidak dipengaruhi oleh perbedaan skala. Selain itu, dihitung pula pertumbuhan ekspor tahunan (2023–2024) sebagai variabel turunan untuk melihat dinamika perubahan dari waktu ke waktu.

Data yang telah ditransformasi kemudian diproses dengan algoritma K-Means *clustering* untuk mengelompokkan data ekspor kopi berdasarkan kesamaan karakteristik. Hasil klasterisasi dievaluasi menggunakan *Davies-Bouldin Index* (DBI), di mana nilai DBI mendekati 0 menunjukkan kualitas klaster yang baik.

2.6 Data Mining

Data mining merupakan proses mengekstraksi pola tersembunyi dari kumpulan data besar agar dapat mendukung pengambilan keputusan [8]. Dalam konteks perdagangan, data mining banyak dimanfaatkan untuk memprediksi tren pasar, segmentasi konsumen, dan analisis daya saing ekspor [9]. Perkembangan fenomena *data explosion* menjadi salah satu latar belakang munculnya konsep *data mining*. Akumulasi data yang tersimpan dalam basis data terus meningkat seiring dengan aktivitas penyimpanan informasi, namun sering kali hanya berfungsi sebagai arsip atau laporan tanpa dimanfaatkan lebih lanjut. Melalui penerapan *data mining*, kumpulan data tersebut dapat diproses untuk menemukan pola-pola tersembunyi yang tidak terlihat secara langsung. Pola yang dihasilkan selanjutnya dapat diinterpretasikan sebagai informasi maupun pengetahuan baru yang bernilai bagi pengambilan keputusan.

Data mining adalah proses untuk menemukan pola dan informasi yang bermanfaat dari kumpulan data berukuran sangat besar [10]. Proses data mining mencakup tahapan pengumpulan data serta analisis pola, yang dalam konteks lebih luas dikenal sebagai *knowledge discovery*, *knowledge extraction*, *data* atau *pattern analysis*, dan *information harvesting* [11]. Dengan demikian, metode data mining sangat relevan untuk digunakan dalam memetakan negara tujuan ekspor kopi.

2.7 Clustering

Clustering merupakan salah satu metode *unsupervised learning* yang digunakan untuk mengelompokkan data berdasarkan kemiripan tertentu [12]. Prinsip K-Means *Clustering* adalah membagi data ke dalam k kelompok dengan meminimalkan jarak antara titik data dengan pusat klaster (centroid) [13]. *Clustering* adalah salah satu pendekatan *unsupervised* dalam machine learning [9]. Metode ini terbagi menjadi dua kategori berdasarkan ilmu statistik yaitu pendekatan berbasis model probabilitas dan pendekatan nonparametrik. Pada pendekatan berbasis model probabilitas, diasumsikan bahwa data berasal dari kombinasi model probabilitas yang berbeda, sehingga metode pengelompokan menggunakan model campuran.

Dalam pemilihan fitur, fitur-fitur pembeda dipilih dari beberapa kandidat. Keunggulan pendekatan ini adalah kemampuannya menangani data yang kompleks dan campuran, serta menghasilkan hasil yang lebih akurat untuk data yang sesuai dengan model probabilitas. Namun, kekurangannya adalah pendekatan ini membutuhkan asumsi tentang distribusi data yang mungkin tidak selalu sesuai dengan data nyata, sehingga bisa membatasi penggunaannya. *Distance Clustering*, yang disebut juga dengan *distance-based clustering*, merupakan metode yang sangat populer untuk melakukan *clustering* objek dan telah menghasilkan hasil yang baik [9].

Melalui penerapan *clustering*, data dapat diklasifikasikan berdasarkan tingkat kepadatan, pola distribusi secara menyeluruh dapat diidentifikasi, serta hubungan antaratribut yang bermakna dapat ditemukan. Dalam konteks *data mining*, fokus utama terletak pada pengembangan metode klastering yang mampu bekerja secara efektif dan efisien pada basis data berukuran besar. Agar metode tersebut optimal, diperlukan sejumlah karakteristik, antara lain kemampuan untuk diskalakan (*scalability*), fleksibilitas dalam menangani berbagai jenis atribut, ketahanan terhadap data berdimensi tinggi, toleransi terhadap *noise*, serta hasil yang mudah diinterpretasikan. Bagian ini selanjutnya memaparkan metode yang digunakan dalam studi. Jelaskan dengan ringkas, tetapi tetap akurat seperti teknik pengerjaan yang dilakukan [10].

2.8 K-Means

K-Means merupakan algoritma clustering yang unggul karena kesederhanaannya dan efisiensinya dalam menangani dataset besar, meskipun kelemahannya adalah ketergantungan pada inisialisasi centroid dan penentuan jumlah klaster di awal [11]. K-means adalah algoritma yang membutuhkan parameter masukan sebanyak k dan membagi kelompok kedalam k *cluster* sehingga dapat mengetahui tingkat kemiripan anggota antara tingkat kemiripan anggota yang tinggi dan tingkat kemiripan anggota yang rendah. Untuk mengetahui kemiripan *cluster* mengukur kedekatan objek pada *centroid cluster* [12].

Algoritma K-means mencoba mempartisi kumpulan data dengan meminimalkan jumlah kuadrat galat jarak dari pusat massa ke setiap titik. Secara khusus, pengelompokan K-means mengevaluasi jarak antara titik data dan pusat massanya (pusat titik data dalam klaster) secara iteratif hingga pusat massa berhenti diperbarui. Titik-titik data dipartisi ke dalam berbagai kelompok dengan jarak antara titik data dan pusat massa yang

diminimalkan [7].

K-Means adalah algoritma iteratif yang membagi data ke dalam k kelompok, dengan tujuan meminimalkan *within-cluster variance* atau inerti. Pada setiap iterasi, algoritma menghitung ulang centroid berdasarkan mean dari data dalam cluster, dan mengalokasikan kembali data ke centroid terdekat. Proses ini diulang hingga konvergensi, di mana tidak ada perubahan signifikan dalam cluster. Dalam Hidayat & Kusniyati (2022), evaluasi cluster dalam K-Means mencakup metode seperti *Elbow Method*, *Silhouette Score* dan *Davies-Bouldin Index* [2]. Kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat

K-means adalah algoritma pengelompokan bersifat iteratif dan bekerja sebagai berikut:

1. Masukkan k, jumlah klaster yang dibentuk.
2. k "pusat" awal diinisialisasi secara acak dari domain data.
3. k klaster dibentuk dengan menempatkan setiap observasi ke pusat terdekatnya.
4. Langkah-langkah berikut diulang hingga klaster berhenti berubah:
 - a. Untuk setiap klaster, rata-rata observasi dalam klaster tersebut diambil sebagai pusat [13].
 - b. Setiap titik data dipindahkan ke klaster yang pusatnya paling dekat dengan titik data tersebut.

Nilai rata-rata yang dihitung dari semua objek dalam klaster menentukan pusat klaster tersebut. Ukuran jarak tradisional yang digunakan dalam algoritma K-means untuk menentukan kesamaan suatu objek dengan rata-rata klaster adalah jarak *Euclidean*.

1. Langkah dalam pengelompokan data adalah:
2. Pilihan jumlah klaster.
3. Klaster dilakukan secara random.

Setiap data klaster terpusat terdekat berdasarkan jarak antar obyek. Tahapan ini jarak dihitung dengan menentukan kesamaan dan tidak kesamaan data dengan Metode Jarak *Euclidean*.

(Euclidean Distance) dengan rumus pada persamaan (1).

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

d (x,y) adalah jarak antara data x ke data y

xi adalah variable data

yi adalah variable data pada titik pusat

Sedangkah untuk pembaharuan suatu titik centroid dapat dilakukan dengan rumus pada persamaan (2).

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

μ_k = Titik centroid *cluster* ke-K

N_k = Banyaknya data pada *cluster* ke-K

X_q = Data ke-q pada *cluster* ke-K

4. Menentukan pusat klaster baru berdasarkan keanggotaan terbaru dengan menghitung nilai rata-rata objek pada masing-masing klaster, atau alternatifnya menggunakan median.
5. Menghitung kembali jarak setiap objek terhadap pusat klaster yang diperbarui. Proses ini dilakukan secara iteratif hingga tidak terjadi perubahan komposisi klaster, yang menandai bahwa pengklasteran telah mencapai kondisi konvergen dan selesai.

2.9 Tahapan Algoritma Clustering K-Means

Tahapan dalam Clustering K-Means penentuan suatu entitas yang menjadi beberapa kelas, sehingga kelas tersebut tidak adanya perubahan. Proses tersebut juga dapat dilakukan melalui penerapan tahapan perhitungan jarak pada algoritma, dengan memanfaatkan berbagai persamaan, salah satunya menggunakan metode *Euclidean Distance*. Cara kerja algoritma K-Means terlihat pada Gambar 2.



Gambar 2. Tahap Algoritma Clustering K-Means [14]

Gambar 2 menjelaskan tahap algoritma K-Means dengan langkah-langkah sebagai berikut:

1. Start
Proses algoritma dimulai.
2. Menentukan Banyaknya Cluster (k)
Pada tahap ini ditentukan jumlah kelompok (cluster) yang diinginkan, sesuai tujuan analisis negara tujuan pada ekspor kopi
3. Menentukan Centroid
Algoritma memilih titik awal (centroid) secara acak, centroid ini akan menjadi pusat awal cluster.
4. Mengelompokkan Objek Berdasarkan Jarak Minimum
Setiap data atau objek dihitung jaraknya terhadap semua centroid. Objek kemudian ditempatkan pada cluster dengan centroid terdekat dengan metode Euclidean distance.
5. Menentukan Centroid Baru
Setelah semua objek terkelompok, centroid diperbarui dengan menghitung rata-rata (mean) dari semua objek dalam masing-masing cluster. Centroid baru ini akan menggantikan centroid lama.
6. Apakah Centroid Berubah?
 - a. Jika ya (berubah), maka algoritma kembali ke langkah pengelompokan objek dan iterasi berlanjut.
 - b. Jika tidak (stabil atau tidak berubah), maka proses *clustering* berhenti karena sudah menemukan data akhir pada pengelompokan.
7. End
Algoritma selesai dijalankan, hasil akhirnya berupa kelompok data (cluster) yang terbentuk berdasarkan kedekatan objek dengan centroid masing-masing.

2.10 Davies-Bouldin Index (DBI)

Evaluasi kualitas kluster penting dilakukan untuk mengukur seberapa baik suatu algoritma *clustering* bekerja. Davies dan Bouldin memperkenalkan *Davies-Bouldin Index* (DBI) sebagai ukuran validitas kluster berdasarkan dua kriteria diantaranya kepadatan data dalam kluster (intra-cluster) dan jarak antar kluster (inter-cluster). Nilai DBI yang lebih rendah menunjukkan kualitas pengelompokan yang lebih baik. Ros, Riad, dan Guillaume (2023) mengembangkan variasi DBI untuk memperbaiki sensitivitas terhadap distribusi data [4]. Sementara Shalsadilla, dkk (2023) menekankan efektivitas DBI dalam menentukan jumlah kluster optimal pada analisis wilayah kemiskinan di Indonesia [14].

Davies dan Bouldin memperkenalkan sebuah metode evaluasi yang kemudian dikenal dengan nama *Davies-Bouldin Index* (DBI). Indeks ini termasuk dalam kategori evaluasi internal kluster, di mana kualitas hasil kluster dinilai berdasarkan tingkat kedekatan antar anggota kluster serta pemisahan antar kluster. DBI menjadi salah satu ukuran yang banyak digunakan untuk menilai validitas hasil pengelompokan, dengan didefinisikan sebagai tingkat kedekatan data terhadap pusat kluster tempat data tersebut bergabung [15].

Indeks Davies-Bouldin menyatakan fungsi dari dua istilah, yang pertama mengidentifikasi minimalisasi varians intra cluster dan yang kedua adalah maksimalisasi jarak antar cluster [4]. *Indeks Davies-Bouldin (DBI)* digunakan dalam proses data mining untuk mengevaluasi kualitas hasil klusterisasi. Nilai DBI biasanya dihitung setelah data dikelompokkan menggunakan algoritma K-Means, dengan jumlah cluster yang telah ditentukan sebelumnya. Evaluasi DBI mempertimbangkan jarak antar cluster (inter-cluster) dan jarak dalam cluster (intra-cluster), di mana tujuan idealnya adalah memaksimalkan jarak antar cluster sekaligus meminimalkan jarak antar anggota dalam cluster yang sama. Semakin kecil nilai DBI, semakin baik kualitas cluster yang terbentuk [15]. Formula perhitungan *Sum of Squares* dijelaskan pada persamaan (3).

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (3)$$

Jumlah data dalam setiap cluster, yang dilambangkan sebagai m_i , dan posisi centroid cluster, c_i , digunakan untuk menghitung ketidakmiripan atau jarak antar data dalam proses pengelompokan. Validasi yang dilakukan pun mengikuti prinsip yang sama, yaitu menilai sejauh mana pengelompokan mencerminkan kedekatan atau perbedaan antar cluster. Salah satu metode yang digunakan adalah Sum of Square Between Cluster (SSB), yaitu matriks yang menilai jarak antar cluster dengan cara membandingkan posisi centroid masing-masing cluster. Dengan demikian, SSB memberikan ukuran kuantitatif untuk menilai perbedaan antara cluster ke- i dan cluster ke- j berdasarkan jarak centroid c_i dan c_j , sehingga hasil pengelompokan dapat dianalisis secara objektif [16]. Sementara antara dua cluster, misalnya cluster i dan j , digunakan formula *sum of square between cluster* (SSB) mengukur jarak antara centroid c_i dan c_j seperti pada persamaan (4).

$$SSB_{i,j} = d(C_i, C_j) \quad (4)$$

Didefinisikan $R_{i,j}$ sebagai rasio yang menggambarkan seberapa baik perbandingan antara cluster ke- i dan cluster ke- j . Nilai rasio ini diperoleh dengan mempertimbangkan dua komponen utama, yaitu kohesi dan separasi. Kohesi mencerminkan tingkat keterikatan data di dalam cluster, sehingga semakin kecil nilainya, semakin rapat data dalam cluster tersebut. Sebaliknya, separasi mengukur jarak antar cluster, di mana nilai yang lebih besar menunjukkan pemisahan yang lebih jelas antara cluster. Dengan demikian, cluster yang ideal adalah cluster yang memiliki kohesi minimal sekaligus separasi maksimal. Rumus $R_{i,j}$ kemudian digunakan untuk menghitung ukuran rasio ini secara kuantitatif (5).

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (5)$$

$$R_{i,j} \geq 0$$

$$R_{i,j} = R_{j,i}$$

$$\text{Jika } SSW_j \geq SSW_r \text{ dan } SSB_{i,j} = SSB_{i,r}, \text{ maka } R_{i,j} > R_{i,r}$$

$$\text{Jika } SSW_j = SSW_r \text{ dan } SSB_{i,j} \leq SSB_{i,r}, \text{ maka } R_{i,j} > R_{i,r}$$

Nilai-nilai *Davies-Bouldin Index* (DBI) didapatkan dari persamaan (6).

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \quad (6)$$

K adalah jumlah cluster yang digunakan.

$R_{i,j}$ adalah sebaran dalam-cluster dan jarak antar cluster.

Dari syarat-syarat perhitungan yang didefinisikan di atas, dapat dilihat bahwa semakin kecil nilai SSW maka hasil clustering yang didapat juga lebih baik. Secara esensial, DBI menginginkan nilai sekecil (non-negatif ≥ 0) mungkin untuk menilai baiknya cluster yang diperoleh.

3. HASIL DAN PEMBAHASAN

3.1 Ekspor Kopi

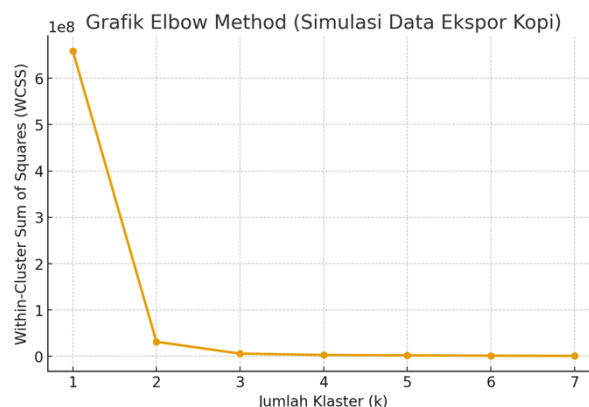
Indonesia merupakan salah satu produsen kopi terbesar dunia dengan kontribusi penting bagi perdagangan internasional. Berdasarkan laporan International Coffee Organization, Indonesia menempati peringkat keempat produsen kopi terbesar setelah Brasil, Vietnam, dan Kolombia. Namun, ekspor kopi Indonesia menunjukkan fluktuasi antar negara tujuan. Beberapa pasar tradisional seperti Amerika Serikat, Jerman, dan Jepang menyerap

volume besar, sementara kawasan lain seperti Timur Tengah dan Asia Tenggara menunjukkan pertumbuhan yang belum konsisten. Hal ini menunjukkan perlunya analisis kuantitatif untuk memetakan pasar ekspor kopi secara lebih sistematis [16]. Ekspor kopi Indonesia dipengaruhi oleh variabel makro ekonomi seperti produksi, kurs rupiah, dan permintaan global. Hasil regresi menunjukkan ada fluktuasi ekspor yang signifikan antar-negara tujuan [17]. Faktor utama yang memengaruhi ekspor adalah luas lahan panen, harga produsen, dan nilai tukar. Model regresi menjelaskan sekitar 53% variasi ekspor pada kopi. Menekankan bahwa Indonesia menghadapi kompetisi kuat dengan Brasil, Vietnam, dan Kolombia. Tantangan ekspor juga berasal dari fluktuasi harga global, perubahan iklim, dan kualitas kopi [18].

Dengan K-Means negara tujuan ekspor dapat dipetakan ke dalam cluster berdasarkan faktor harga, nilai tukar, dan permintaan, sehingga membantu identifikasi pasar. Dibedakan negara tujuan yang sensitif terhadap harga, kualitas, maupun faktor non-tarif. Hal ini bisa jadi dasar strategi diferensiasi pasar. Data variabel ekspor per negara dapat dikelompokkan untuk menemukan *cluster* negara tujuan dengan karakteristik yang dapat menguntungkan perusahaan Café Coffee.

3.2 Dataset

Dataset merupakan sekumpulan data yang akan digunakan dalam pelatihan model *Machine Learning* [19]. Dataset dapat berwujud berbagai jenis data, seperti citra, suara, sinyal, maupun teks, tergantung pada tujuan pengembangan model. Data tersebut dapat diperoleh melalui proses pengumpulan secara langsung ataupun dengan memanfaatkan sumber yang secara khusus menyediakan dataset untuk kepentingan pembelajaran maupun penelitian. Dalam penelitian ini, penelitian menggunakan dataset dari Café Coffee atas penjualan ekspor kopi periode 2023-2024. Untuk memvalidasi jumlah kluster yang optimal, digunakan metode Elbow dengan nilai K dari 2 hingga 6. Berdasarkan grafik Elbow, titik siku terlihat pada $K = 3$. Hal ini menunjukkan bahwa tiga kluster, yaitu laba rendah, laba sedang, dan laba tinggi, merupakan jumlah kluster yang paling tepat, sebagaimana terlihat pada Gambar 3.



Gambar 3. Penentuan Kluster Optimal dengan Elbow Method.

Gambar 3 menunjukkan penurunan tajam pada nilai WCSS hingga $k = 3$, kemudian penurunannya melandai setelahnya dan membentuk pola “siku” (elbow) yang jelas pada titik $k = 3$. Diketahui bahwa nilai WCSS menurun drastis dari $k=1 \rightarrow k=2 \rightarrow k=3$. Setelah $k=3$, penurunan WCSS tidak signifikan lagi. Dengan demikian, $k=3$ merupakan jumlah kluster optimal karena menyeimbangkan antara homogenitas dalam kluster dan efisiensi model.

Hasil pemodelan klusterisasi menggunakan algoritma K-Means dengan jumlah kluster tiga (laba rendah, sedang, dan tinggi) menunjukkan adanya diferensiasi yang jelas terhadap karakteristik pasar ekspor kopi Café Coffee. Nilai *Davies-Bouldin Index (DBI)* sebesar 0,442 menandakan kualitas pemisahan kluster yang baik dan stabil. Namun, analisis ini tidak hanya berhenti pada deskripsi statistik; secara ekonomi, hasil kluster memberikan gambaran tentang struktur pasar ekspor Indonesia. Negara yang masuk dalam kluster laba tinggi, seperti Amerika Serikat, merepresentasikan pasar utama dengan permintaan tinggi dan potensi stabilitas devisa yang besar. Sebaliknya, kluster laba sedang berisi negara-negara dengan permintaan moderat yang dapat ditingkatkan melalui kebijakan promosi, penguatan merek, atau perjanjian perdagangan. Adapun kluster laba rendah mencerminkan pasar potensial yang perlu strategi penetrasi jangka panjang. Visualisasi grafik *Line Plot* menunjukkan tren penurunan pada kluster laba tinggi dan peningkatan gradual pada kluster sedang, menandakan perlunya strategi diversifikasi ekspor agar tidak bergantung pada satu pasar dominan. Secara metodologis, hasil

K-Means ini masih bersifat *partition-based* dan sensitif terhadap inisialisasi centroid, sehingga dibandingkan dengan algoritma lain seperti DBSCAN atau *Hierarchical Clustering*, model ini lebih efisien untuk data homogen namun kurang optimal dalam mendeteksi variasi kompleks antarnegara. Dengan demikian, meskipun hasil kluster valid secara statistik, analisis lanjutan dengan pendekatan multi-algoritmik diperlukan untuk memperdalam pemahaman implikasi ekonomi dari segmentasi pasar ekspor kopi Indonesia.

3.3 Implementasi Clustering K-Means dengan Manual

1. Menentukan centroid awal

Langkah awal dalam proses algoritma K-Means penentuan jumlah *cluster* menjadi 3 dan kemudian centroid dilakukan secara acak atau secara manual yang diambil dari data. Pada penelitian ini centroid yang di ambil adalah cluster 1 data ke-4, cluster 2 data ke-11 dan *cluster* 3 data ke-2, seperti pada Tabel 2.

Tabel 2. Centroid awal

Centroid Awal Secara Acak		
C1	283.00	535.20
C2	150.00	180.00
C3	13,785.80	8,723.00

2. Menghitung jarak pada centroid dengan rumus Euclidean Distance

$$\text{Jarak data ke 1 atas centroid 1} = \sqrt{(440 - 283)^2 + (858 - 535.20)^2} = 358.96$$

$$\text{Jarak data ke 1 atas centroid 2} = \sqrt{(13,785.80 - 150)^2 + (8,723 - 180)^2} = 15,791.32$$

$$\text{Jarak data ke 1 atas centroid 3} = \sqrt{(250 - 13,785.80)^2 + (450 - 8723)^2} = 91.36$$

Perhitungan ini terus berlanjut sampai data ke-40 dan dibawah ini hasil data yang sudah dihitung pada data ke-1 sampai ke-40, berikut ini data atas perhitungan hasil jarak sebagaimana terlihat pada Tabel 3.

Tabel 3. Perhitungan Jarak Centroid

		2023	2024				
No	Country	Quantity (Cubic Meter)		C1	C2	C3	Jarak
1	Afrika Selatan	440.00	858.00	358.96	737.42	15,490.92	358.96
2	Amerika Serikat	13,785.80	8,723.00	15,791.32	16,090.93	-	-
3	Argentina	250.00	450.00	91.37	287.92	15,863.81	91.37
4	Australia	283.00	535.20	-	379.28	15,791.32	-
5	Bangladesh	155.00	170.00	386.98	11.18	16,092.00	11.18
6	Belanda	5,230.20	7,363.70	8,432.27	8,798.52	8,662.91	8,432.27
7	Belarus	800.00	1,600.00	1,183.68	1,561.70	14,811.08	1,183.68
8	Brasil	740.00	450.00	464.87	648.85	15,447.83	464.87
9	China	1,980.00	3,450.00	3,372.81	3,747.24	12,929.87	3,372.81
10	Filipina	160.00	270.00	292.34	90.55	16,034.83	90.55
11	Gabon	150.00	180.00	379.28	-	16,090.93	-
12	Haiti	150.00	180.00	379.28	-	16,090.93	-
13	India	6,935.60	2,268.40	6,874.67	7,099.70	9,412.07	6,874.67
14	Inggris	158.00	150.00	404.97	31.05	16,100.10	31.05
15	Italia	4,246.00	4,647.80	5,711.29	6,061.23	10,373.77	5,711.29
16	Jamaika	120.00	170.00	399.93	31.62	16,121.66	31.62
17	Jepang	796.80	380.00	536.73	677.02	15,437.61	536.73

18	Jerman	770.00	776.40	543.46	860.29	15,249.90	543.46
19	Kanada	450.00	600.00	179.13	516.14	15,614.95	179.13
20	Kolombia	86.40	338.40	278.18	170.69	16,061.60	170.69
21	Korea Selatan	150.00	150.00	407.51	30.00	16,106.87	30.00
22	Laos	57.60	108.00	483.02	117.14	16,207.46	117.14
23	Liberia	110.00	110.00	459.05	80.62	16,162.03	80.62
24	Malaysia	2,520.00	3,450.00	3,674.27	4,038.54	12,438.76	3,674.27
25	Myanmar	80.00	80.00	498.41	122.07	16,203.41	122.07
26	Oman	57.60	43.20	541.17	165.08	16,242.00	165.08
27	Pantai Gading	19.20	108.00	502.09	149.31	16,240.00	149.31
28	Papua Nugini	2,660.00	2,560.00	3,122.49	3,458.97	12,718.73	3,122.49
29	Peru	120.00	120.00	446.05	67.08	16,148.24	67.08
30	Prancis	735.00	2,720.00	2,231.07	2,606.50	14,365.21	2,231.07
31	Qatar	206.40	343.00	206.90	172.48	15,956.96	172.48
32	Rusia	921.60	4,203.60	3,723.57	4,096.92	13,634.98	3,723.57
33	Singapura	3,326.00	2,386.80	3,562.06	3,867.42	12,229.26	3,562.06
34	Spanyol	78.00	57.60	519.74	142.01	16,217.06	142.01
35	Swiss	96.00	230.00	357.93	73.59	16,110.30	73.59
36	Taiwan	170.00	137.00	413.92	47.42	16,096.88	47.42
37	Thailand	686.00	1,164.80	747.53	1,121.22	15,123.86	747.53
38	Timor Leste	208.00	312.00	235.46	144.18	15,971.90	144.18
39	Uni Emirat Arab	388.40	841.40	323.83	703.05	15,543.81	323.83
40	Vietnam	1,822.20	728.80	1,551.33	1,759.95	14,388.71	1,551.33

3. Pengelompokan Hasil Jarak

Berdasarkan hasil perhitungan jarak, diperoleh pengelompokan data yang mengacu pada jarak terdekat dari pusat kluster. Tabel 4 menunjukkan hasil dari proses pengelompokan tersebut.

Table 4. Hasil Pengelompokan Jarak

No	Country	C1	C2	C3
1	Afrika Selatan	√		
2	Amerika Serikat			√
3	Argentina	√		
4	Australia	√		
5	Bangladesh		√	
6	Belanda	√		
7	Belarus	√		
8	Brasil	√		
9	China	√		
10	Filipina		√	
11	Gabon		√	
12	Haiti		√	
13	India	√		
14	Inggris		√	
15	Italia	√		
16	Jamaika		√	
17	Jepang	√		
18	Jerman	√		
19	Kanada	√		

20	Kolombia	√
21	Korea Selatan	√
22	Laos	√
23	Liberia	√
24	Malaysia	√
25	Myanmar	√
26	Oman	√
27	Pantai Gading	√
28	Papua Nugini	√
29	Peru	√
30	Prancis	√
31	Qatar	√
32	Rusia	√
33	Singapura	√
34	Spanyol	√
35	Swiss	√
36	Taiwan	√
37	Thailand	√
38	Timor Leste	√
39	UEA	√
40	Vietnam	√

4. Menentukan Centroid Baru

Setelah centroid baru ini dihitung, iterasi berikutnya menggunakan nilai-nilai ini untuk menentukan pengelompokan data kembali. Proses ini terus berulang sampai tidak ada perubahan dalam pengelompokan data atau posisi centroid. Tabel 5, 6, 7 dan 8 menunjukkan hasil perhitungan pada centroid baru, centroid baru kluster 2, centroid baru kluster 3 centroid baru kluster 4.

Tabel 5. Menentukan Centroid Baru

Centroid Baru		
C1	1799.04	2071.75
C2	122.75	171.43
C3	13785.80	8723.00

Tabel 6. Menentukan Centroid Baru Kluster 2

Centroid k2		
C1	2,834.24	3,216.28
C2	254.87	332.61
C3	13,785.80	8,723.00

Tabel 7. Menentukan Centroid Baru Kluster 3

Centroid k3		
C1	3,172.71	3,672.26
C2	325.29	388.06
C3	13,785.80	8,723.00

Tabel 8. Menentukan Centroid Baru Kluster 4

Centroid k4		
C1	3,477.43	3,791.29
C2	338.50	463.28
C3	13,785.80	8,723.00

Penentuan *centroid* baru dilakukan secara iteratif hingga nilai rata-rata setiap klaster mencapai kondisi stabil. Pada penelitian ini, proses iterasi berhenti pada pembaruan *centroid* keempat, di mana tidak lagi terjadi perubahan posisi *centroid*. Tabel 9 menyajikan hasil akhir dari klaster ke-4.

Tabel 9. Hasil Perhitungan Jarak Atas Klaster 4

		2023	2024					
No	Country	Quantity (Cubic Meter)		C1	C2	C3	Jarak	Hasil
1	Afrika Selatan	440.00	858.00	4,222.57	407.56	15,490.92	407.56	C2
2	Amerika Serikat	13,785.80	8,723.00	11,427.35	15,781.40	-	-	C3
3	Argentina	250.00	450.00	4,645.48	89.49	15,863.81	89.49	C2
4	Australia	283.00	535.20	4,561.41	90.84	15,791.32	90.84	C2
5	Bangladesh	155.00	170.00	4,914.49	345.96	16,092.00	345.96	C2
6	Belanda	5,230.20	7,363.70	3,979.24	8,458.39	8,662.91	3,979.24	C1
7	Belarus	800.00	1,600.00	3,459.82	1,226.83	14,811.08	1,226.83	C2
8	Brasil	740.00	450.00	4,319.46	401.72	15,447.83	401.72	C2
9	China	1,980.00	3,450.00	1,535.83	3,408.08	12,929.87	1,535.83	C1
10	Filipina	160.00	270.00	4,837.85	263.10	16,034.83	263.10	C2
11	Gabon	150.00	180.00	4,910.51	340.27	16,090.93	340.27	C2
12	Haiti	150.00	180.00	4,910.51	340.27	16,090.93	340.27	C2
13	India	6,935.60	2,268.40	3,778.65	6,839.60	9,412.07	3,778.65	C1
14	Inggris	158.00	150.00	4,927.23	361.56	16,100.10	361.56	C2
15	Italia	4,246.00	4,647.80	1,150.79	5,725.27	10,373.77	1,150.79	C1
16	Jamaika	120.00	170.00	4,938.22	365.73	16,121.66	365.73	C2
17	Jepang	796.80	380.00	4,338.51	465.80	15,437.61	465.80	C2
18	Jerman	770.00	776.40	4,052.12	533.13	15,249.90	533.13	C2
19	Kanada	450.00	600.00	4,398.82	176.42	15,614.95	176.42	C2
20	Kolombia	86.40	338.40	4,839.57	281.34	16,061.60	281.34	C2
21	Korea Selatan	150.00	150.00	4,932.62	365.62	16,106.87	365.62	C2
22	Laos	57.60	108.00	5,026.11	452.92	16,207.46	452.92	C2
23	Liberia	110.00	110.00	4,989.13	420.74	16,162.03	420.74	C2
24	Malaysia	2,520.00	3,450.00	1,016.43	3,698.57	12,438.76	1,016.43	C1
25	Myanmar	80.00	80.00	5,031.52	462.31	16,203.41	462.31	C2
26	Oman	57.60	43.20	5,073.79	505.35	16,242.00	505.35	C2
27	Pantai Gading	19.20	108.00	5,052.32	477.68	16,240.00	477.68	C2
28	Papua Nugini	2,660.00	2,560.00	1,477.92	3,128.19	12,718.73	1,477.92	C1
29	Peru	120.00	120.00	4,975.00	406.92	16,148.24	406.92	C2
30	Prancis	735.00	2,720.00	2,944.24	2,291.28	14,365.21	2,291.28	C2
31	Qatar	206.40	343.00	4,752.92	178.66	15,956.96	178.66	C2
32	Rusia	921.60	4,203.60	2,588.87	3,785.49	13,634.98	2,588.87	C1

33	Singapura	3,326.00	2,386.80	1,412.63	3,553.17	12,229.26	1,412.63	C1
34	Spanyol	78.00	57.60	5,049.41	482.12	16,217.06	482.12	C2
35	Swiss	96.00	230.00	4,910.89	336.50	16,110.30	336.50	C2
36	Taiwan	170.00	137.00	4,928.78	367.23	16,096.88	367.23	C2
37	Thailand	686.00	1,164.80	3,832.82	782.87	15,123.86	782.87	C2
38	Timor Leste	208.00	312.00	4,774.37	199.79	15,971.90	199.79	C2
39	UEA	388.40	841.40	4,271.29	381.39	15,543.81	381.39	C2
40	Vietnam	1,822.20	728.80	3,481.18	1,507.27	14,388.71	1,507.27	C2

Tabel 10. Hasil Clustering K-Means

Klaster	Jumlah	Kategori Negara	Anggota Klaster
C1	8	Nilai Laba Rendah	Belanda, China, India, Italia, Malaysia, Papua Nugini, Rusia, Singapura
C2	31	Nilai Laba Sedang	Afrika Selatan, Argentina, Australia, Bangladesh, Belarus, Brasil, Filipina, Gabon, Haiti, Inggris, Jamaika, Jepang, Jerman, Kanada, Kolombia, Korea Selatan, Laos, Liberia, Myanmar, Oman, Pantai Gading, Peru, Prancis, Qatar, Spanyol, Swiss, Taiwan, Thailand, Timor Leste, Uni Emirat Arab, Vietnam
C3	1	Nilai Laba Tinggi	Amerika Serikat

3.4 Implementasi K-Means dengan RapidMiner

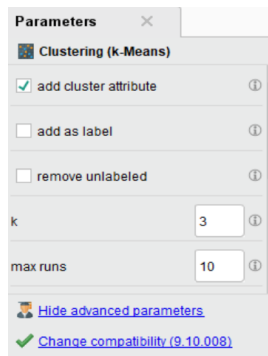
Dalam implementasi K-Means dengan *software RapidMiner* peneliti menyiapkan dataset yang filenya sudah format .xlsx (excel) pada prosesnya peneliti memasukan data tersebut total ada 40 negara. Gambar 4 menunjukkan hasil dari pemodelan.

Row No.	Country	cluster	2023.0	2024.0
1	Afrika Selatan	cluster_1	440	858
2	Amerika Ser...	cluster_2	13786	8723
3	Argentina	cluster_1	250	450
4	Australia	cluster_1	283	535
5	Bangladesh	cluster_1	155	170
6	Belanda	cluster_0	5230	7364
7	Belarus	cluster_1	800	1600
8	Brasil	cluster_1	740	450
9	China	cluster_0	1980	3450
10	Filipina	cluster_1	160	270
11	Gabon	cluster_1	150	180
12	Haiti	cluster_1	150	180
13	India	cluster_0	6936	2268

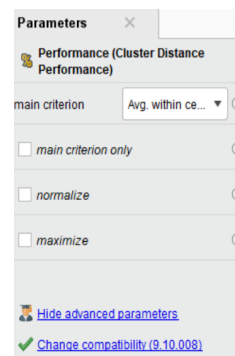
ExampleSet (40 examples, 2 special attributes, 2 regular attributes)

Gambar 4. Dataset Dalam *RapidMiner*

Berikutnya menjelaskan proses pengelompokan atau pengklusteran dengan menggunakan *RapidMiner* yang dengan pemilihan operator dengan menggunakan tool K-Means dengan kluster yang ditentukan ada 3 kluster. Untuk menentukan Klasterisasi dan Menghubungkan dengan *Performance (cluster distance performance)* tool ini digunakan untuk mengukur kualitas hasil pengelompokan untuk di lanjutkan proses K-Means dengan kluster yang ditentukan ada 3 kluster, K=0 adalah kluster laba rendah, K=1 kluster laba sedang dan K=2 kluster laba tinggi.

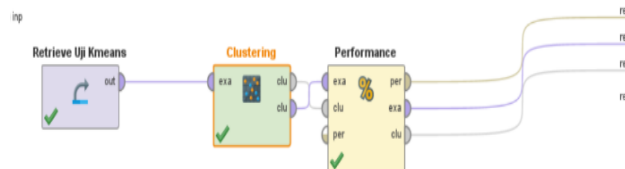


Gambar 5. Tampilan K-Means Kluster 3



Gambar 6. Performance (cluster distance performance)

Gambar 5 dan Gambar 6 merupakan proses *clustering* menggunakan parameter K-means pada tahap awal ialah *Retrive Uji Kmeans* tahap ini untuk mengambil atau menyiapkan data yang diuji menggunakan K-Means selanjutnya Tahap kedua ialah *clustering* pengelompokan menjadi tiga *cluster* dan Setelah proses *clustering* selanjutnya tahap terakhir ialah *performance* untuk mengukur dan mengevaluasi seberapa baik cluster model jarak bekerja.



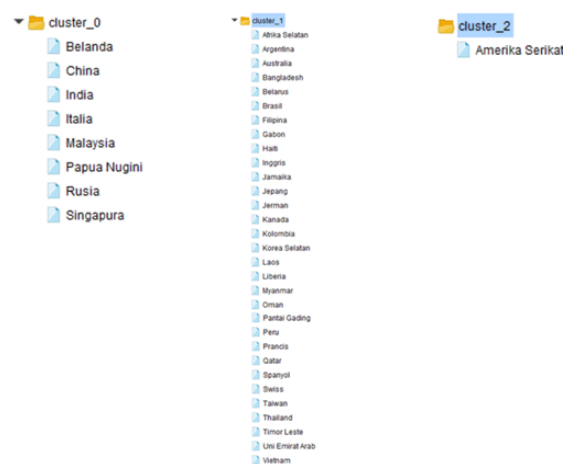
Gambar 7. Tampilan Parameter

Gambar 7 menunjukkan tampilan kluster model dapat dilihat setelah proses seleksi K-mean sudah diseleksi. Adapun Hasil dari masing masing kelompok ada 3 kluster antara lain kluster 0 (pertama) ada 8 kluster berisi negara dengan laba ekspor kategori rendah, kluster 1 (kedua) negara ekspor kopi dengan kategori laba sedang terdapat 31 negara dan kluster 2 (ketiga) negara dengan laba ekspor tinggi dengan satu anggota negara.

Cluster Model

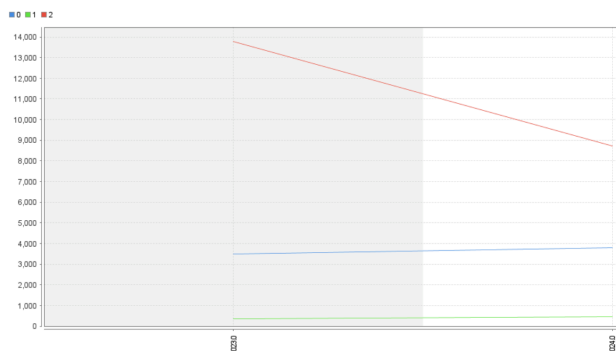
```
Cluster 0: 8 items
Cluster 1: 31 items
Cluster 2: 1 items
Total number of items: 40
```

Gambar 8. Tampilan kluster Model



Gambar 9. Tampilan Hasil Cluster Model

Gambar 8 dan 9 merupakan hasil tampilan kluster model dengan nama nama negara yang sudah dikelompokan oleh K Means dan terdapat 3 kluster yang dapat kita kategorikan menjadi tingkat ekspor laba rendah, sedang dan tertinggi. hasil pada pengelolaan kluster perhitungan manual dan rapidminer sama. Berarti kluster model sudah ideal. Dan dapat dilanjutkan untuk diimplementasikan sesuai kebutuhan perusahaan dalam mengclusterkan prioritas utama pada negara.



Gambar 10. Diagram Line Plot K-Means 3 Cluster

Gambar 10 pada diagram Line Plot K-Means 3 Cluster, dapat dideskripsikan:

1. Sumbu X (horizontal) menunjukkan periode waktu (2023 – 2024).
2. Sumbu Y (vertikal) menunjukkan nilai ekspor kopi per Cluster dalam meter cubic
3. Warna garis:
 - a. Biru = Cluster 1 (Laba Rendah)
 Berada di tingkat menengah sekitar 3.500 – 3.900 cubic meter dalam pemesanan ekspor kopi total ada 8 negara. Menunjukkan meningkat sedikit dan relatif stabil. Negara dalam klaster ini dapat dianggap sebagai pasar sekunder yang cukup konsisten dan berpotensi untuk ditingkatkan eksponya.
 - b. Hijau = Cluster 2 (Laba Sedang)
 Berada di tingkat paling rendah sekitar 300 – 400 cubic meter dalam pemesanan ekspor kopi total ada 31 negara. Negara dalam klaster ini adalah pasar kecil, namun bisa dianggap sebagai pasar jangka panjang cenderung tumbuh kenaikan, bisa dijadikan target pengembangan ke depan.
 - c. Merah = Cluster 3 (Laba Tinggi)
 Memiliki nilai ekspor tertinggi pada awal periode sekitar 14.000 cubic meter dalam pemesanan ekspor kopi total ada 1 negara. Namun mengalami tren penurunan signifikan hingga sekitar 9.000 pada tahun 2024. Ini menunjukkan bahwa negara dalam klaster 3 adalah pasar utama tetapi butuh strategi mempertahankan.

Tabel 10 menjelaskan evaluasi klaster menggunakan perhitungan *Davies-Bouldin Index* (DBI) diperoleh hasil 0.442 atas 3 *clustering*. Adapun perbandingan terkait beberapa *clustering* 4 dengan DBI hasil 0.601, *clustering* 5 dengan DBI hasil 0.463 Jika nilai DBI semakin kecil atau semakin mendekati 0, maka hasil klaster yang diperoleh semakin bagus dan menunjukkan bahwa hasil klaster yang diperoleh adalah relatif sangat baik.

Tabel 10. Visualisasi Pengukuran *Davies-Bouldin Index*

Jumlah Klaster	Nilai <i>Davies-Bouldin Index</i>
3	0.442
4	0.601
5	0.463

4. KESIMPULAN

Kesimpulan yang didapat dari penelitian ini yaitu dapat melakukan pengelompokan negara yang memiliki tingkat ekspor kopi berdasarkan laba dapat terselesaikan dengan menggunakan Algoritma K-Means Melalui proses *Data Mining* pada K-Means yang di evaluasi dengan nilai *Davies-Bouldin Index* sebesar 0.442 adalah nilai yang sangat baik. Diperoleh pemahaman yang lebih mendalam mengenai kelompok negara yang memberikan kontribusi laba tinggi, sedang, dan rendah. Hasil evaluasi performa klaster menunjukkan bahwa pemisahan data berdasarkan karakteristik laba relatif optimal dan membantu mengidentifikasi pasar ekspor yang

paling menguntungkan dan memberikan informasi untuk Café Coffee agar dapat berinovasi dan menjaga kualitas kopi yang diekspor. Nilai Laba Rendah ada 8 Negara yaitu Belanda, China, India, Italia, Malaysia, Papua Nugini, Rusia, Singapura, nilai laba sedang total ada 31 Negara yaitu Afrika Selatan, Argentina, Australia, Bangladesh, Belarus, Brasil, Filipina, Gabon, Haiti, Inggris, Jamaika, Jepang, Jerman, Kanada, Kolombia, Korea Selatan, Laos, Liberia, Myanmar, Oman, Pantai Gading, Peru, Prancis, Qatar, Spanyol, Swiss, Taiwan, Thailand, Timor Leste, Uni Emirat Arab, Vietnam dan nilai laba tinggi hanya ada 1 negara yaitu Amerika Serikat dengan tingkat populasi 343,603,404 paling terpadat di benua Amerika Utara. Negara yang memiliki nilai laba tinggi dapat kita prioritaskan karena negara tersebut secara terus menerus pembelian kopi pada perusahaan secara berulang. Evaluasi kualitas *clustering* melalui DBI menghasilkan nilai yang relatif baik, sehingga dapat disimpulkan bahwa model mampu memisahkan negara tujuan ekspor berdasarkan kemiripan karakteristik. Temuan ini memberikan implikasi praktis bagi pemerintah, asosiasi kopi, maupun eksportir untuk merancang strategi ekspor yang lebih tepat sasaran: mempertahankan pasar utama stabil, mengantisipasi risiko di pasar fluktuatif, dan mendorong penetrasi ke pasar potensial.

Keterbatasan pada penelitian ini variabel yang digunakan masih terbatas pada volume cubic meter dan pertumbuhan ekspor pertahun pemesanan dari banyaknya negara pada dataset perusahaan Café Coffee. Data yang digunakan bersifat cross-section dua tahun periode 2023 -2024, sehingga belum menggambarkan jangka panjang. Penelitian ini memperkuat penggunaan *K-Means-DBI* dalam analisis perdagangan komoditas pertanian sebagai metode segmentasi yang efisien dan terukur. Namun, untuk memperdalam kontribusi ilmiah, penelitian selanjutnya disarankan mengintegrasikan variabel ekonomi makro dan non-ekonomi, memperluas periode data menjadi multiyear, serta melakukan perbandingan dengan algoritma lain seperti *DBSCAN*, *Hierarchical Clustering*, dan *Fuzzy C-Means* guna memperoleh pemetaan pasar yang lebih komprehensif dan prediktif terhadap dinamika ekspor kopi Indonesia.

DAFTAR PUSTAKA

- [1] D. Asmaya, B. Irawan, and A. Bahtiar, "Optimalisasi Penjualan Produk Melalui Analisis Data Dan Pengelompokan Produk," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 4336–4343, 2024, doi: 10.36040/jati.v8i3.9805.
- [2] R. Hidayat and H. Kusniyati, "Analisis Clustering Dalam Pengelompokan Penjualan Menggunakan Algoritma K-Means Pada Cafe 47°Coffee," *CESS (Journal Comput. Eng. Syst. Sci. Comput. Eng. Syst. Sci.*, vol. 7, no. 2, pp. 420–434, 2022, [Online]. Available: www.jurnal.unimed.ac.id
- [3] K. Dbscan and Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan Dbscan," *KAKIFIKOM (Kumpulan Artik. Karya Ilm. Fak. Ilmu Komputer)*, vol. 06, no. 01, pp. 60–74, 2024.
- [4] F. Ros, R. Riad, and S. Guillaume, "Neurocomputing PDBI : A partitioning Davies-Bouldin index for clustering evaluation," vol. 528, pp. 178–199, 2023, doi: 10.1016/j.neucom.2023.01.043.
- [5] W. Alfian, Kusri, and T. Hidayat, "Analisis Clustering Pegawai Berdasarkan Tingkat Kedisiplinan Menggunakan Algoritma K-Means dan Davies-Bouldin Index," *J. Electr. Eng. Comput.*, vol. 6, no. 2, pp. 437–448, 2024, doi: 10.33650/jeeecom.v6i2.9556.
- [6] F. M. Anto, L. S. Abimanyu, and T. Herdi, "Penerapan Algoritma Naïve Bayes Dengan Feature Selection Pada Data Penjualan Konstruksi," *J. Ilm. FIFO*, vol. 15, no. 2, p. 102, 2024, doi: 10.22441/fifo.2023.v15i2.002.
- [7] Y. Jumaryadi, R. Fajriah, U. Salamah, B. Priambodo, and A. Lystha, "Machine Learning Approaches to Sentiment Analysis of Mental Health Discussions on Platform X," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 13, no. 2, pp. 235–246, 2025, doi: 10.33558/piksel.v13i2.11350.
- [8] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. Elsevier Inc, 2012.
- [9] T. Li, Y. Ma, and T. Endoh, "Normalization-based validity index of adaptive K-means clustering for multi-solution application," *IEEE Access*, vol. 8, pp. 9403–9419, 2020, doi: 10.1109/ACCESS.2020.2964763.
- [10] Y. E. Praptiningsih, W. W. Ariestya, I. Astuti, and S. Nurulita, "Data Mining untuk Klasifikasi Diagnosa Kanker Payudara Dengan Menerapkan Algoritma C4.5," *J. Ilm. FIFO*, vol. 15, no. 1, p. 37, 2023, doi: 10.22441/fifo.2023.v15i1.005.
- [11] K. Dong, I. Romanov, C. McLellan, and A. F. Esen, "Recent text-based research and applications in railways: A critical review and future trends," *Eng. Appl. Artif. Intell.*, vol. 116, no. April, p. 105435, 2022, doi: 10.1016/j.engappai.2022.105435.
- [12] - Arientawati, Y. Jumaryadi, and A. Wibowo, "Analisis Pengelompokan Gangguan TIK Pada Sistem Pencatatan Layanan Menggunakan Algoritma K-Means dan Metode Elbow," *Techno.COM*, vol. 22, no. 2, pp. 348–357, 2023.

-
- [13] I. T. R. Yanto, R. Setiyowati, N. Azizah, and Rasyidah, "A framework of mutual information kullback-leibler divergence based for clustering categorical data," *Int. J. Informatics Vis.*, vol. 5, no. 1, pp. 11–15, 2021, doi: 10.30630/joiv.5.1.462.
 - [14] N. Shalsadilla, S. Martha, H. Perdana, N. Satyahadewi, and E. Sulistianingsih, "Penentuan Jumlah Cluster Optimum Menggunakan Davies Bouldin Index dalam Pengelompokan Wilayah Kemiskinan di Indonesia," *Statistika*, vol. 23, no. 1, pp. 63–72, 2023.
 - [15] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020, doi: 10.1088/1757-899X/725/1/012128.
 - [16] T. L. Putri *et al.*, "Penerapan data mining pada clustering data harga rumah dki jakarta menggunakan algoritmak-means," vol. 8, no. 1, pp. 1174–1179, 2024.
 - [17] U. Herdira and D. B. Hakim, "Factors Affecting the Volume of Indonesian Coffee Export to The International Market," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 4099, pp. 348–354, 2021, doi: 10.32628/ijsrset218367.
 - [18] V. D. Permana, M. Patiung, and N. S. Wisnujati, "A Comprehensive Analysis of Indonesia's Coffee Exports to the Global Market from 1991 to 2021," *Sci. J. Multi Discip. Sci.*, vol. 03, no. 2, pp. 89–99, 2024.
 - [19] A. D. Aprianto and D. Ramayanti, "Penggunaan Klasifikasi Objek dalam Aplikasi Android untuk Melestarikan Kuliner Khas Indonesia," *J. Ilm. FIFO*, vol. 16, no. 1, p. 25, 2024, doi: 10.22441/fifo.2024.v16i1.003.