VOL. 03. NO. 01 JANUARI 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

ANALISIS CLUSTERING PRODUK BUAH MENGGUNAKAN ALGORITMA K-MEDOID DENGAN EVALUASI DAVIES BOULDIN INDEX

Raynard Kenneth Eden Kansil[1], Kaslani[2], Raditya Danar Dana[3], Mulayawan[4].

Program Studi Manajemen Informatika[13]
Program Studi Sistem Informasi[2]
Program Studi Sistem Informasi[4]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
raynardkansil75@gmail.com

(*) Corresponding Author : raynardkansil75@gmail.com
Published : 30 Januari 2026

Abstract—*Fruit products are horticultural commodities that play an important role in meeting community nutritional needs and contributing to the economy, particularly in the agribusiness retail sector. However, fruit product management at the retail level still faces several challenges, such as high product variety, differences in quality, price fluctuations, and the perishable nature of the products. In practice, fruit product grouping is generally conducted manually and subjectively based on the experience of store managers, which results in underutilization of available data. This condition leads to less effective decision-making related to product arrangement, inventory management, and marketing strategies. Therefore, a data-driven approach is required to group fruit products objectively and systematically. This study aims to analyze the clustering of fruit products using the K-Medoid algorithm and to evaluate the quality of the resulting clusters using the Davies–Bouldin Index (DBI). The data used in this research were obtained from Pusat Buah Pati and consist of attributes that represent product characteristics, including product name, barcode code, category, and unit of sale. The research methodology follows the Knowledge Discovery in Databases (KDD) framework, which includes data selection, preprocessing, data transformation, clustering, and result evaluation. The K-Medoid algorithm was selected due to its robustness in handling categorical data and its resistance to outliers compared to centroid-based clustering methods. To determine the optimal number of clusters, experiments were conducted by testing cluster numbers from K = 2 to K = 20, which were then evaluated using the Davies–Bouldin Index, where a lower DBI value indicates better clustering quality. The results of this study indicate that the K-Medoid algorithm is capable of grouping fruit products into distinct clusters with similar characteristics in a consistent manner. The evaluation of different K values produced varying Davies–Bouldin Index scores, enabling the objective determination of the optimal number of clusters. The resulting clusters represent groups of fruit products with similar categories and sales units, which can be utilized to support inventory management, product display arrangement, and marketing strategy formulation. Therefore, this study demonstrates that the application of the K-Medoid algorithm combined with Davies–Bouldin Index evaluation provides an effective data-driven solution for clustering fruit products.*

Keywords*: clustering, fruit products, K-Medoid, Davies–Bouldin Index, data mining.*

Abstrak- Produk buah merupakan komoditas hortikultura yang memiliki peran penting dalam pemenuhan kebutuhan gizi masyarakat serta kontribusi terhadap perekonomian, khususnya pada sektor agribisnis ritel. Namun, pengelolaan produk buah di tingkat pengecer masih menghadapi berbagai permasalahan, antara lain tingginya variasi jenis produk, perbedaan kualitas, fluktuasi harga, serta sifat produk yang mudah rusak. Pada praktiknya, pengelompokan produk buah masih banyak dilakukan secara manual dan subjektif berdasarkan pengalaman pengelola, sehingga belum mampu memanfaatkan data yang tersedia secara optimal. Kondisi ini menyebabkan pengambilan keputusan terkait penataan produk, pengelolaan stok, dan strategi pemasaran menjadi kurang efektif. Oleh karena itu, diperlukan pendekatan berbasis data untuk mengelompokkan produk buah secara objektif dan sistematis. Penelitian ini bertujuan untuk menganalisis pengelompokan produk buah menggunakan algoritma K-Medoid serta mengevaluasi kualitas cluster yang dihasilkan dengan menggunakan Davies–Bouldin Index (DBI). Data yang digunakan merupakan data produk buah yang diperoleh dari Pusat Buah Pati, dengan atribut yang mencerminkan karakteristik produk, seperti nama produk, kode barcode, kategori, dan satuan penjualan. Metode penelitian mengacu pada tahapan Knowledge Discovery in Databases (KDD), yang meliputi seleksi data,

preprocessing, transformasi data, proses clustering, dan evaluasi hasil. Algoritma K-Medoid dipilih karena memiliki keunggulan dalam menangani data kategorikal dan lebih robust terhadap outlier dibandingkan metode clustering berbasis centroid. Untuk menentukan jumlah cluster yang optimal, dilakukan pengujian jumlah cluster mulai dari K = 2 hingga K = 20, kemudian dievaluasi menggunakan Davies–Bouldin Index, di mana nilai DBI yang lebih kecil menunjukkan kualitas pengelompokan yang lebih baik. Hasil penelitian menunjukkan bahwa algoritma K-Medoid mampu mengelompokkan produk buah ke dalam beberapa cluster yang memiliki karakteristik serupa secara jelas dan konsisten. Pengujian berbagai nilai K menghasilkan nilai Davies–Bouldin Index yang bervariasi, sehingga memungkinkan penentuan jumlah cluster optimal secara objektif. Cluster yang terbentuk dapat merepresentasikan kelompok produk buah berdasarkan kesamaan kategori dan satuan penjualan, yang berpotensi dimanfaatkan dalam pengelolaan stok, penataan display produk, serta perumusan strategi pemasaran. Dengan demikian, penelitian ini membuktikan bahwa penerapan algoritma K-Medoid dengan evaluasi Davies–Bouldin Index dapat menjadi solusi analitis yang efektif dalam pengelompokan produk buah berbasis data.

Kata kunci: clustering, produk buah, K-Medoid, Davies–Bouldin Index, data mining.

INTRODUCTION

Produk buah merupakan salah satu komoditas hortikultura yang memiliki peran strategis dalam perekonomian nasional, khususnya di negara agraris seperti Indonesia. Buah tidak hanya berfungsi sebagai sumber gizi masyarakat, tetapi juga menjadi sumber pendapatan bagi petani, pedagang, distributor, serta pelaku usaha agribisnis lainnya. Menurut Kementerian Pertanian Republik Indonesia, subsektor hortikultura, termasuk buah-buahan, memberikan kontribusi signifikan terhadap Produk Domestik Bruto (PDB) sektor pertanian dan menyerap tenaga kerja dalam jumlah besar. Namun demikian, pengelolaan produk buah menghadapi berbagai tantangan kompleks, seperti fluktuasi harga, ketidakseimbangan antara permintaan dan pasokan, tingginya tingkat kerusakan (perishability), serta ketergantungan pada musim panen. Kondisi ini berdampak langsung pada stabilitas pendapatan petani dan efisiensi rantai pasok buah. Selain itu, keterbatasan informasi mengenai karakteristik produk buah—seperti kualitas, ukuran, tingkat kematangan, dan daya simpan—menyebabkan proses distribusi dan pemasaran belum optimal. Permasalahan tersebut sering kali berujung pada pemborosan hasil panen dan penurunan nilai ekonomi produk buah di tingkat produsen maupun konsumen. Oleh karena itu, diperlukan pendekatan berbasis data untuk memahami pola dan karakteristik produk buah secara lebih sistematis guna mendukung pengambilan keputusan yang lebih akurat dan berkelanjutan [1], [2]

Dalam praktiknya, pengelompokan produk buah selama ini umumnya dilakukan secara konvensional berdasarkan segmentasi pembeli atau pengalaman subjektif pelaku usaha. Pedagang dan distributor biasanya mengelompokkan buah berdasarkan kriteria sederhana, seperti harga jual, jenis buah, ukuran fisik, atau preferensi konsumen tertentu, tanpa melibatkan metode analisis data yang terstruktur. Pendekatan ini sangat bergantung pada intuisi dan pengalaman individu, sehingga hasil pengelompokan cenderung tidak konsisten dan sulit direplikasi. Selain itu, segmentasi manual sering kali mengabaikan keterkaitan antaratribut produk buah yang sebenarnya dapat memberikan informasi lebih mendalam. Akibatnya, potensi data yang tersedia tidak dimanfaatkan secara optimal untuk menghasilkan pengetahuan baru. Penelitian lain menunjukkan bahwa segmentasi pasar berbasis persepsi subjektif memiliki keterbatasan dalam mengidentifikasi pola tersembunyi pada data produk. Dalam konteks produk buah, pendekatan non-algoritmik ini berisiko menghasilkan keputusan pemasaran yang kurang tepat, seperti penentuan harga yang tidak sesuai atau distribusi produk yang tidak efisien. Oleh karena itu, dibutuhkan pendekatan analitis yang mampu mengelompokkan produk buah secara objektif dan berbasis data guna meningkatkan efektivitas pengelolaan dan pemasaran produk [3], [4]

Beberapa penelitian sebelumnya telah membahas penerapan teknik clustering dalam pengelompokan produk pertanian. Penelitian oleh Prasetyo dengan judul *"Penerapan Algoritma K-Means untuk Pengelompokan Komoditas Hortikultura"* membahas permasalahan pengelompokan produk hortikultura berdasarkan harga dan volume penjualan menggunakan algoritma K-Means. Hasil penelitian menunjukkan bahwa K-Means mampu membagi data ke dalam beberapa cluster yang merepresentasikan kategori produk dengan karakteristik serupa. Selanjutnya, penelitian oleh Sari dan Wijaya berjudul *"Analisis Segmentasi Produk Pertanian Menggunakan Metode Clustering"* mengangkat permasalahan ketidakefisienan distribusi produk pertanian akibat pengelompokan yang tidak tepat. Metode yang digunakan adalah hierarchical clustering dengan

jarak Euclidean. Selain itu, penelitian oleh Rahman et al. dalam *"Clustering Analysis of Agricultural Products Using K-Medoid"* mengkaji penggunaan algoritma K-Medoid untuk mengatasi sensitivitas terhadap outlier pada data produk pertanian. Penelitian tersebut menegaskan bahwa K-Medoid lebih robust dibandingkan K-Means dalam menangani data dengan nilai ekstrem. Namun, sebagian besar penelitian tersebut belum secara spesifik membahas evaluasi kualitas cluster menggunakan indeks validasi internal seperti Davies–Bouldin Index [2], [5], [6]

Berdasarkan kajian terhadap penelitian-penelitian terdahulu, dapat diidentifikasi adanya celah (research gap) yang relevan dengan penelitian ini. Sebagian besar penelitian pengelompokan produk buah atau pertanian masih berfokus pada algoritma K-Means, yang memiliki kelemahan dalam menangani outlier dan sangat bergantung pada nilai centroid awal. Meskipun beberapa penelitian telah menggunakan K-Medoid sebagai alternatif, evaluasi kualitas cluster sering kali hanya dilakukan secara visual atau deskriptif, tanpa menggunakan indeks validasi yang terukur secara kuantitatif. Selain itu, integrasi proses pengelompokan ke dalam kerangka Knowledge Discovery in Databases (KDD) belum dibahas secara komprehensif. Penelitian ini hadir untuk mengisi gap tersebut dengan menerapkan algoritma K-Medoid yang lebih robust, serta menggunakan Davies–Bouldin Index sebagai alat evaluasi untuk menentukan jumlah cluster terbaik secara objektif. Dengan demikian, penelitian ini tidak hanya menghasilkan pengelompokan produk buah, tetapi juga memberikan ukuran validitas cluster yang dapat dipertanggungjawabkan secara ilmiah [7]

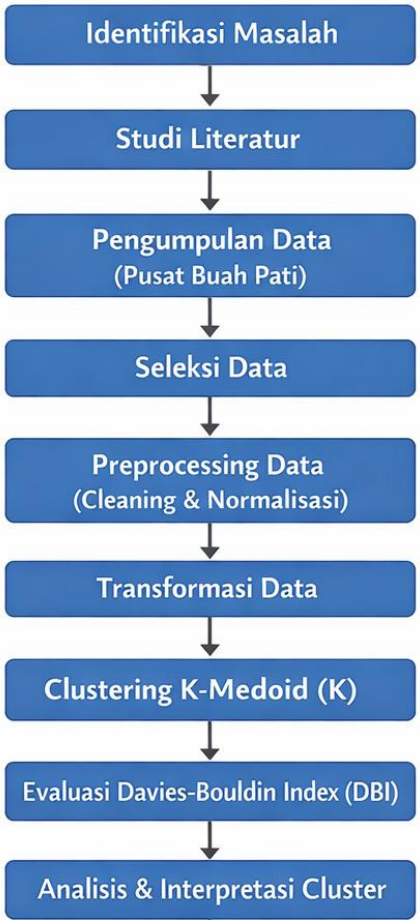
Penyelesaian permasalahan pengelompokan produk buah dalam penelitian ini mengacu pada tahapan Knowledge Discovery in Databases (KDD), yang meliputi seleksi data, preprocessing, transformasi data, data mining, dan evaluasi. Pada tahap data mining, algoritma K-Medoid digunakan untuk mengelompokkan produk buah berdasarkan atribut-atribut yang relevan, seperti harga, kualitas, dan tingkat permintaan. Pemilihan K-Medoid didasarkan pada kemampuannya dalam meminimalkan pengaruh outlier dan menghasilkan cluster yang lebih stabil. Selanjutnya, Davies–Bouldin Index digunakan untuk mengevaluasi kualitas cluster dengan mengukur rasio kedekatan intra-cluster dan jarak antar-cluster. Pendekatan ini memungkinkan penentuan jumlah cluster optimal secara objektif. Dengan mengintegrasikan K-Medoid dalam kerangka KDD, penelitian ini diharapkan mampu menghasilkan pengetahuan baru yang dapat digunakan sebagai dasar pengambilan keputusan strategis dalam pengelolaan produk buah[8]

Tujuan utama penelitian ini adalah untuk menganalisis pengelompokan produk buah menggunakan algoritma K-Medoid dan mengevaluasi kualitas cluster yang dihasilkan dengan Davies–Bouldin Index. Secara khusus, penelitian ini bertujuan untuk mengidentifikasi pola dan karakteristik produk buah yang memiliki kesamaan atribut, menentukan jumlah cluster optimal, serta memberikan dasar analitis bagi pengambilan keputusan dalam pengelolaan dan pemasaran produk buah. Selain itu, penelitian ini juga bertujuan untuk membuktikan efektivitas algoritma K-Medoid dibandingkan metode konvensional dalam konteks pengelompokan produk buah. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi teoretis dalam bidang data mining serta kontribusi praktis bagi pelaku usaha agribisnis .

Implikasi dari penelitian ini diharapkan dapat dirasakan baik secara akademik maupun praktis. Secara akademik, penelitian ini memperkaya kajian mengenai penerapan algoritma clustering dalam sektor pertanian, khususnya produk buah, dengan menekankan penggunaan K-Medoid dan evaluasi Davies–Bouldin Index. Secara praktis, hasil pengelompokan dapat digunakan oleh petani, distributor, dan pengambil kebijakan untuk meningkatkan efisiensi distribusi, penentuan harga, dan strategi pemasaran produk buah. Selain itu, penelitian ini dapat menjadi dasar bagi pengembangan sistem pendukung keputusan berbasis data untuk pengelolaan produk hortikultura secara berkelanjutan. Dengan demikian, penelitian ini diharapkan mampu memberikan kontribusi nyata dalam meningkatkan daya saing dan nilai ekonomi produk buah di pasar

MATERIALS AND METHODS

Tahapan perancangan dalam penelitian ini mengacu pada konsep Knowledge Discovery in Databases (KDD) yang digunakan sebagai kerangka kerja sistematis dalam proses analisis data.



Gambar 1 Diagram Alir Penelitian.

Flowchart tahapan penelitian pada Gambar 1 menggambarkan alur kerja penelitian secara sistematis mulai dari tahap awal hingga tahap akhir. Setiap tahapan disusun untuk memastikan proses penelitian berjalan terstruktur, terukur, dan sesuai dengan tujuan penelitian. Adapun penjelasan masing-masing tahapan adalah sebagai berikut:

1. **Identifikasi Masalah**

   Tahap pertama dalam penelitian ini adalah identifikasi masalah. Pada tahap ini peneliti mengkaji kondisi nyata yang terjadi pada pengelolaan produk buah, khususnya dalam proses pengelompokan produk. Permasalahan yang diidentifikasi adalah pengelompokan produk buah yang masih dilakukan secara manual dan subjektif, sehingga belum mampu memberikan informasi yang akurat dan konsisten terkait karakteristik produk. Tahap ini menjadi dasar dalam menentukan arah penelitian, metode yang digunakan, serta tujuan yang ingin dicapai.

2. **Studi Literatur**

   Tahap selanjutnya adalah studi literatur. Pada tahap ini peneliti mempelajari dan menelaah berbagai sumber pustaka yang relevan, seperti buku, jurnal ilmiah, dan penelitian terdahulu yang berkaitan dengan data mining, clustering, algoritma K-Medoid, Knowledge Discovery in Databases (KDD), serta metode evaluasi clustering menggunakan Davies–Bouldin Index. Studi literatur bertujuan untuk memperoleh landasan teori yang kuat serta memastikan bahwa metode yang digunakan sesuai dengan permasalahan penelitian.

3. Pengumpulan Data

   Pengumpulan data dilakukan secara langsung di objek penelitian, yaitu **Pusat Buah Pati**. Data yang dikumpulkan merupakan data primer yang mencerminkan kondisi riil produk buah yang dijual di lokasi penelitian. Atribut data meliputi informasi yang relevan dengan proses clustering, seperti jenis buah, harga, dan kualitas produk. Tahap ini bertujuan untuk memperoleh dataset yang akan digunakan sebagai bahan analisis dalam penelitian.

4. **Seleksi Data**

   Pada tahap seleksi data, peneliti melakukan pemilihan data dan atribut yang sesuai dengan tujuan penelitian. Tidak semua data hasil pengumpulan digunakan dalam proses clustering. Oleh karena itu, data yang dianggap tidak relevan atau tidak mendukung analisis akan dieliminasi. Tahap ini bertujuan untuk meningkatkan efisiensi proses pengolahan data dan memfokuskan analisis pada variabel yang memiliki pengaruh terhadap pengelompokan produk buah.

5. **Preprocessing Data**

   Tahap preprocessing data bertujuan untuk meningkatkan kualitas data sebelum dianalisis. Proses yang dilakukan pada tahap ini meliputi pembersihan data (data cleaning), penanganan data kosong (missing value), serta normalisasi data agar setiap atribut memiliki skala yang sebanding. Preprocessing merupakan tahap penting karena kualitas data sangat memengaruhi hasil clustering yang dihasilkan oleh algoritma K-Medoid.

6. **Transformasi Data**

   Pada tahap transformasi data, data diubah ke dalam format yang sesuai dengan kebutuhan algoritma clustering. Transformasi dapat berupa pengkodean data kategorikal menjadi numerik, penyesuaian tipe data, serta pembentukan dataset akhir yang siap digunakan dalam

proses clustering. Tahap ini memastikan bahwa data dapat diproses secara optimal oleh algoritma K-Medoid.

7. **Proses Clustering Menggunakan K-Medoid**

   Tahap ini merupakan inti dari penelitian, yaitu penerapan algoritma K-Medoid untuk mengelompokkan produk buah berdasarkan kemiripan karakteristiknya. Proses clustering dilakukan dengan menentukan jumlah cluster (K) dan menghitung jarak antar data untuk menentukan medoid sebagai pusat cluster. Algoritma K-Medoid dipilih karena lebih robust terhadap outlier dan mampu menghasilkan cluster yang lebih stabil pada data ritel buah.

8. **Evaluasi Menggunakan Davies–Bouldin Index**

   Hasil clustering yang diperoleh kemudian dievaluasi menggunakan Davies–Bouldin Index (DBI). Evaluasi ini bertujuan untuk mengukur kualitas cluster berdasarkan rasio jarak intra-cluster dan jarak antar-cluster. Nilai DBI yang lebih kecil menunjukkan hasil clustering yang lebih baik. Pada tahap ini dilakukan pengujian dengan beberapa nilai K untuk menentukan jumlah cluster optimal.

RESULTS AND DISCUSSION

Berdasarkan hasil pengujian jumlah klaster yang dilakukan secara bertahap mulai dari K = 2 hingga K = 20, diperoleh sejumlah temuan yang menunjukkan variasi kualitas pengelompokan pada setiap nilai K. Pengujian ini dilakukan untuk mengevaluasi kinerja model K-Medoid dalam membentuk klaster yang optimal, di mana setiap konfigurasi K menghasilkan nilai evaluasi yang berbeda. Hasil pengujian tersebut selanjutnya dianalisis untuk menentukan jumlah klaster terbaik yang mampu merepresentasikan struktur data secara efektif dan menghasilkan tingkat homogenitas intra-klaster serta heterogenitas antar-klaster yang optimal.

Tabel 1 Hasil pengujian

No	Jumlah Cluster (K)	Nilai DBI
1	2	1.554.982
2	3	1.621.141
3	4	1.628.485
4	5	1.628.322
5	6	1.602.413
6	7	1.601.868
7	8	1.591.104
8	9	1.593.045
9	10	1.599.318
10	11	1.608.536
11	12	1.603.086
12	13	1.588.360
13	14	1.588.932
14	15	1.595.166
15	16	1.587.551
16	17	1.590.260
17	18	1.588.695
18	19	1.588.373
19	20	1.587.385

Penentuan jumlah cluster (K) merupakan salah satu tahap paling krusial dalam analisis clustering, karena nilai K yang dipilih secara langsung memengaruhi kualitas, interpretabilitas, dan validitas hasil pengelompokan data. Pada penelitian ini, evaluasi jumlah cluster dilakukan menggunakan Davies Bouldin Index (DBI), yaitu salah satu metrik evaluasi internal yang mengukur kualitas cluster berdasarkan rasio kedekatan antar data dalam satu cluster (*intra-cluster*) dan jarak antar cluster (*inter-cluster*). Prinsip utama dari DBI adalah bahwa semakin kecil nilai DBI, maka semakin baik kualitas cluster yang dihasilkan, karena menunjukkan cluster yang kompak dan saling terpisah dengan jelas.

Berdasarkan tabel hasil pengujian, pengujian dilakukan pada rentang jumlah cluster K = 2 hingga K = 20. Setiap nilai K menghasilkan nilai DBI yang berbeda-beda, yang mencerminkan perubahan kualitas struktur cluster seiring dengan bertambahnya jumlah cluster. Dari hasil tersebut, dapat diamati bahwa nilai DBI terendah diperoleh pada K = 2, yaitu sebesar 1.554982. Nilai ini merupakan nilai minimum dibandingkan seluruh konfigurasi K lainnya, sehingga secara kuantitatif K = 2 dinyatakan sebagai jumlah cluster terbaik berdasarkan kriteria Davies Bouldin Index.

Jika dianalisis lebih lanjut, nilai DBI pada K = 3 hingga K = 5 justru mengalami peningkatan yang cukup signifikan, dengan nilai DBI berada di atas 1.62. Peningkatan ini menunjukkan bahwa penambahan jumlah cluster pada rentang tersebut belum mampu memperbaiki kualitas pengelompokan. Hal ini dapat terjadi karena data belum terpisah secara alami ke dalam lebih banyak kelompok, sehingga pemaksaan pembentukan cluster tambahan justru meningkatkan tumpang tindih antar cluster dan menurunkan kualitas pemisahan.

Pada rentang K = 6 hingga K = 10, nilai DBI mulai mengalami penurunan secara perlahan, berada pada kisaran 1.59 hingga 1.60. Kondisi ini menunjukkan bahwa penambahan jumlah cluster mulai mampu menangkap variasi internal data dengan lebih baik dibandingkan K = 3–5. Namun

demikian, nilai DBI pada rentang ini tetap lebih tinggi dibandingkan nilai DBI pada K = 2, sehingga secara objektif kualitas cluster yang dihasilkan masih belum lebih baik dibandingkan konfigurasi dua cluster.

Selanjutnya, pada rentang K = 11 hingga K = 20, nilai DBI cenderung stabil dan berfluktuasi dalam rentang yang relatif sempit, yaitu sekitar 1.587 hingga 1.608. Fenomena ini menunjukkan bahwa penambahan jumlah cluster pada rentang tersebut tidak memberikan peningkatan kualitas cluster yang signifikan. Dengan kata lain, meskipun jumlah cluster bertambah dan segmentasi data menjadi lebih rinci, peningkatan tersebut tidak diiringi dengan perbaikan struktur cluster yang berarti menurut metrik DBI. Kondisi ini mengindikasikan adanya batas alami dalam struktur data, di mana pemecahan cluster lebih lanjut tidak lagi menghasilkan pemisahan yang lebih baik.

Pemilihan K = 2 sebagai jumlah cluster terbaik juga sejalan dengan hasil analisis distribusi cluster sebelumnya. Pada K = 2, data terbagi menjadi dua kelompok utama dengan perbedaan karakteristik yang cukup jelas, yaitu kelompok produk dominan dan kelompok produk dengan karakteristik yang lebih beragam. Struktur ini mencerminkan pola alami data produk buah yang cenderung terbagi menjadi dua segmen besar. Dengan demikian, nilai DBI yang paling rendah pada K = 2 tidak hanya valid secara numerik, tetapi juga logis secara konseptual dan mudah diinterpretasikan.

Selain itu, pemilihan K = 2 memiliki keunggulan dari sisi interpretabilitas hasil. Cluster yang lebih sedikit memudahkan analisis karakteristik masing-masing cluster, terutama dalam konteks penelitian terapan seperti pengelompokan produk. Meskipun jumlah cluster yang lebih besar dapat memberikan detail yang lebih rinci, kompleksitas interpretasi juga meningkat dan berpotensi mengurangi kegunaan praktis hasil clustering. Oleh karena itu, pemilihan K = 2 merupakan kompromi terbaik antara kualitas cluster berdasarkan DBI dan kemudahan interpretasi hasil.

Secara keseluruhan, tabel hasil pengujian Davies Bouldin Index menunjukkan bahwa K = 2 merupakan jumlah cluster optimal dalam penelitian ini. Nilai DBI yang paling rendah mengindikasikan bahwa struktur data paling baik direpresentasikan dalam dua cluster utama. Hasil ini menjadi dasar yang kuat untuk tahap analisis lanjutan, seperti interpretasi karakteristik masing-masing cluster dan penarikan kesimpulan penelitian. Dengan demikian, pemilihan K = 2 tidak hanya didukung oleh nilai evaluasi kuantitatif, tetapi juga oleh konsistensi pola data dan pertimbangan analitis yang komprehensif.

CONCLUSION

Berdasarkan hasil penelitian dan pembahasan mengenai *Analisis Clustering Produk Buah Menggunakan Algoritma K-Medoid dengan Evaluasi Davies–Bouldin Index*, dapat ditarik beberapa kesimpulan sebagai berikut. Penelitian ini berhasil menerapkan algoritma K-Medoid untuk mengelompokkan data produk buah yang bersifat kategorikal dengan jumlah data sebanyak 3606 record. Algoritma K-Medoid dipilih karena memiliki keunggulan dalam menangani data dengan kemungkinan outlier serta menggunakan medoid yang merepresentasikan data aktual, sehingga hasil pengelompokan lebih stabil dan mudah diinterpretasikan dibandingkan metode berbasis centroid.

Tahapan preprocessing yang dilakukan dalam penelitian ini terbukti berperan penting dalam meningkatkan kualitas data sebelum proses clustering. Proses pembersihan data, penanganan nilai tidak konsisten, serta penyesuaian atribut kategorikal menghasilkan dataset yang lebih bersih, konsisten, dan siap digunakan dalam perhitungan jarak. Penggunaan jarak Gower memungkinkan pengolahan data kategorikal secara efektif, sehingga tingkat kemiripan antar produk buah dapat dihitung secara proporsional dan adil.

Hasil Exploratory Data Analysis (EDA) menunjukkan bahwa data produk buah memiliki tingkat keragaman yang tinggi, khususnya pada atribut nama produk dan kode barcode, serta distribusi kategori yang tidak merata. Kategori apel Fuji menjadi kategori dominan, diikuti oleh anggur merah, jeruk, dan beberapa kategori lainnya. Kondisi ini menunjukkan bahwa struktur data produk buah bersifat heterogen dan kompleks, sehingga membutuhkan metode clustering yang robust untuk mengungkap pola tersembunyi dalam data.

Pengujian jumlah cluster dilakukan pada rentang K = 2 hingga K = 20 dengan menggunakan Davies–Bouldin Index sebagai metode evaluasi internal. Hasil pengujian menunjukkan bahwa nilai Davies–Bouldin Index terendah diperoleh pada K = 2, yaitu sebesar 1,554982. Nilai ini mengindikasikan bahwa konfigurasi dua cluster menghasilkan kualitas pengelompokan terbaik, dengan tingkat kekompakan intra-cluster yang baik serta pemisahan antar cluster yang jelas dibandingkan konfigurasi K lainnya.

REFERENCE

[1] OECD and F. A. O, *Agricultural outlook 2023–2032*. OECD Publishing, 2023.

[2] A. Suryadi and E. Pratama, "Evaluasi clustering menggunakan Davies–Bouldin index pada data UMKM," *Jurnal Informatika*, vol. 11, no. 2, pp. 145–154, 2024.
[3] B. Santosa, *Data mining: Teknik pemanfaatan data untuk keperluan bisnis*. Graha Ilmu, 2021.
[4] R. A. Putra and T. Hidayat, "Implementasi Gower distance pada clustering data campuran," *Jurnal Ilmu Komputer dan Informasi*, vol. 17, no. 1, pp. 55–64, 2024.
[5] E. Prasetyo, *Data mining: Konsep dan aplikasi menggunakan Python*. Andi Publisher, 2021.
[6] Md. M. Rahman, Md. S. Islam, and Md. A. Hossain, "Clustering analysis of agricultural products using K-medoids algorithm," *J. Agric. Food Res.*, vol. 5, 2021.
[7] J. Han, M. Kamber, and J. Pei, *Data mining: Concepts and techniques*. Morgan Kaufmann, 2022.
[8] Y. Li and X. Wu, "Robust clustering methods for categorical data," *Knowl. Based. Syst.*, vol. 238, 2022.