ANALISIS PERBANDINGAN K-MEANS DAN K-MEDOIDS UNTUK SEGMENTASI PRODUK BERDASARKAN DATA PENJUALAN

Muhammad Ali Ridla [1], Noval Prayoga [2]
[1,2] Sistem Informasi, Universitas Ibrahimy
el.riedla@gmail.com

ABSTRAK

Data penjualan produk dapat dimanfaatkan untuk mengidentifikasi karakteristik transaksi dan mendukung pengambilan keputusan terkait persediaan serta pemasaran. Namun, data transaksi yang digunakan dalam penelitian ini belum dikelompokkan berdasarkan karakteristik penjualannya sehingga pola kemiripan antartransaksi belum terlihat secara sistematis. Kondisi tersebut menjadi dasar penerapan teknik klasterisasi untuk membentuk kelompok berdasarkan atribut numerik yang tersedia. Penelitian ini secara khusus membandingkan kualitas hasil klasterisasi algoritma K-Means dan K-Medoids pada dataset penjualan yang sama. Perbandingan dilakukan dengan menggunakan atribut jumlah order, harga, dan total penjualan, serta mengevaluasi jumlah klaster K=2 sampai K=7 menggunakan Davies-Bouldin Index (DBI). Tujuannya adalah mengetahui nilai K yang menghasilkan DBI terendah pada masing-masing algoritma dan membandingkan hasilnya pada dataset penelitian. Hasil evaluasi menunjukkan bahwa K-Means memperoleh DBI terendah sebesar 0,429 pada K=4, sedangkan K-Medoids memperoleh DBI terendah sebesar 0,875 pada K=6. Karena DBI yang lebih rendah menunjukkan kualitas pemisahan klaster yang lebih baik, hasil tersebut menunjukkan bahwa K-Means menghasilkan nilai DBI yang lebih rendah daripada K-Medoids pada dataset penelitian ini. Temuan ini menjadi bukti empiris pada dataset yang digunakan, bukan kesimpulan bahwa K-Means selalu lebih baik untuk semua dataset.

Keyword : segmentasi produk, Davies-Bouldin Index, K-Means, K-Medoids, data mining

1. PENDAHULUAN

Perkembangan teknologi informasi menyebabkan jumlah data yang dihasilkan dari aktivitas penjualan semakin besar. Data penjualan yang tersimpan secara berkala dapat dianalisis untuk memperoleh informasi mengenai karakteristik penjualan dan mendukung pengambilan keputusan. Pada dataset penelitian ini, data transaksi memuat informasi seperti jumlah order, harga, dan total penjualan yang dapat digunakan untuk melihat pola kemiripan antartransaksi.

Salah satu pendekatan yang dapat digunakan untuk menganalisis data penjualan adalah data mining [1]. Dalam konteks penelitian ini, teknik clustering relevan karena tidak memerlukan label kelas sebelumnya dan dapat mengelompokkan objek berdasarkan kemiripan karakteristiknya [2]. Dengan demikian, clustering dapat digunakan untuk membentuk kelompok transaksi berdasarkan karakteristik penjualan yang serupa.

Dua algoritma clustering yang digunakan dalam penelitian ini adalah K-Means dan K-Medoids. K-Means menggunakan centroid sebagai pusat klaster dan dikenal memiliki proses komputasi yang relatif efisien [3]. Namun, centroid berupa nilai rata-rata dapat dipengaruhi oleh outlier. K-Medoids menggunakan objek data aktual sebagai pusat klaster sehingga lebih tahan terhadap pengaruh outlier dibandingkan K-Means [4]. Perbedaan mekanisme tersebut menjadi dasar untuk membandingkan kualitas hasil kedua algoritma pada dataset penelitian dengan prosedur preprocessing dan atribut yang sama.

Perbandingan kedua algoritma penting dilakukan karena hasil clustering dapat dipengaruhi oleh karakteristik data dan metode yang digunakan. Penelitian ini tidak mengasumsikan bahwa salah satu algoritma selalu menghasilkan klaster yang lebih baik. Sebaliknya, kualitas hasil ditentukan secara empiris menggunakan nilai DBI pada dataset yang diteliti.

Penelitian terdahulu telah menerapkan K-Means pada data penjualan untuk mengelompokkan barang berdasarkan volume penjualan dan mendukung pengelolaan persediaan [5]. Penelitian lain menerapkan K-Medoids pada penjualan smartphone untuk mengelompokkan produk berdasarkan pola penjualan [6]. K-Medoids juga digunakan untuk mengelompokkan data penjualan barang sebagai dasar pengoptimalan penjualan [7]. Studi-studi tersebut menunjukkan bahwa kedua algoritma telah digunakan dalam konteks penjualan, tetapi penerapan kedua algoritma secara berdampingan pada dataset yang sama memberikan konteks perbandingan yang berbeda karena preprocessing, atribut, dan rentang K dapat dikendalikan dalam satu eksperimen.

Berdasarkan penelitian terdahulu tersebut, masih diperlukan bukti empiris mengenai bagaimana K-Means dan K-Medoids menghasilkan kualitas klaster yang berbeda ketika diterapkan pada dataset penjualan dengan atribut dan preprocessing yang sama. Penelitian ini mengisi kebutuhan tersebut dengan membandingkan nilai DBI pada K=2 hingga K=7. Kontribusi penelitian bersifat empiris dan metodologis pada tingkat dataset, yaitu memberikan gambaran hasil perbandingan kedua

algoritma dalam kondisi eksperimen yang sama serta karakteristik klaster yang dihasilkan.

Dengan pendekatan tersebut, penelitian ini mengevaluasi hasil K-Means dan K-Medoids menggunakan Davies-Bouldin Index (DBI). Nilai DBI terendah digunakan sebagai dasar pemilihan jumlah klaster pada masing-masing algoritma. Hasil klaster kemudian diinterpretasikan berdasarkan nilai centroid agar karakteristik kelompok dapat dijelaskan secara deskriptif.

Penelitian ini diharapkan memberikan informasi mengenai perbedaan hasil K-Means dan K-Medoids pada dataset penjualan yang digunakan serta memberikan gambaran karakteristik kelompok berdasarkan jumlah order, harga, dan total penjualan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian mengenai penerapan algoritma clustering pada data penjualan telah dilakukan dengan berbagai pendekatan. Lesmana dan Gunawan melakukan perbandingan algoritma K-Means dan K-Medoids dalam melakukan clustering data penjualan pada PT. United Teknologi Integrasi. Penelitian tersebut menunjukkan bahwa kedua algoritma dapat digunakan untuk mengelompokkan data penjualan berdasarkan karakteristik transaksi, sehingga dapat memberikan informasi yang mendukung analisis penjualan [8].

Rahman dan Yustanti melakukan penelitian mengenai perbandingan algoritma K-Means, K-Medoids, dan Agglomerative untuk pengelompokan penjualan produk komputer. Penelitian tersebut menggunakan Elbow Method untuk menentukan jumlah cluster dan Silhouette Coefficient untuk mengevaluasi hasil clustering. Hasil penelitian membentuk beberapa kelompok produk berdasarkan karakteristik harga dan kuantitas penjualan [9].

Abidin membandingkan algoritma K-Means, K-Medoids, dan Fuzzy C-Means pada data penjualan PT. XYZ. Penelitian tersebut menunjukkan bahwa teknik clustering dapat digunakan untuk mengelompokkan data transaksi penjualan dan memberikan informasi yang dapat mendukung analisis serta strategi penjualan [10].

Fithriyah, Yaqin, dan Zaman menerapkan K-Means Clustering untuk melakukan segmentasi produk berdasarkan analisis Recency, Frequency, dan Monetary (RFM) pada data transaksi penjualan. Penelitian tersebut menunjukkan bahwa segmentasi produk dapat digunakan untuk membantu memahami karakteristik produk dan mendukung pengelolaan persediaan [11].

Wicaksono et al. menerapkan algoritma K-Medoids untuk melakukan klasterisasi data penjualan barang. Penelitian tersebut menggunakan hasil pengelompokan untuk membedakan karakteristik barang berdasarkan tingkat penjualannya sehingga dapat mendukung analisis penjualan dan pengelolaan stok [7].

Berdasarkan penelitian terdahulu tersebut, algoritma K-Means dan K-Medoids telah banyak diterapkan untuk mengelompokkan data penjualan. Penelitian sebelumnya juga menunjukkan bahwa penggunaan metode clustering dapat membantu mengidentifikasi kelompok produk berdasarkan karakteristik penjualan. Penelitian ini memiliki fokus pada perbandingan K-Means dan K-Medoids dengan menggunakan atribut jumlah order, harga, dan total penjualan. Kedua algoritma diuji pada dataset yang sama dengan beberapa jumlah cluster, yaitu K=2 sampai K=7, kemudian kualitas hasil clustering dibandingkan menggunakan Davies-Bouldin Index (DBI). Dengan demikian, penelitian ini berfokus pada perbandingan kualitas hasil clustering kedua algoritma berdasarkan karakteristik dataset penjualan yang digunakan.

2.2. Data Mining

Data mining merupakan proses untuk menemukan pengetahuan, pola, atau informasi yang bermanfaat dari kumpulan data dengan memanfaatkan teknik statistik, pembelajaran mesin, dan pengelolaan basis data [1]. Dalam penelitian ini, data mining digunakan sebagai pendekatan untuk menganalisis data penjualan dan menemukan kelompok dengan karakteristik penjualan yang serupa.

2.3. Clustering

Clustering merupakan teknik data mining yang mengelompokkan objek berdasarkan tingkat kemiripan karakteristiknya [2]. Objek dalam klaster yang sama diharapkan memiliki karakteristik yang lebih mirip dibandingkan objek pada klaster yang berbeda. Pada penelitian ini, clustering digunakan untuk mengelompokkan unit analisis berdasarkan atribut jumlah order, harga, dan total penjualan.

2.4. Algoritma K-Means

K-Means mengelompokkan data berdasarkan jarak setiap objek terhadap centroid . Tahapan umumnya meliputi pemilihan jumlah klaster, penentuan centroid awal, perhitungan jarak Euclidean, penempatan objek ke klaster terdekat, pembaruan centroid, dan pengulangan sampai kondisi berhenti tercapai. Pada penelitian ini, Euclidean distance digunakan sebagai ukuran jarak untuk menjaga keseragaman evaluasi antaralgoritma.

2.5. Algoritma K-Medoids

K-Medoids merupakan algoritma clustering yang menggunakan objek data aktual sebagai pusat klaster atau medoid [4]. Algoritma ini mengelompokkan objek berdasarkan jarak terhadap medoid dan memperbarui medoid melalui proses optimasi untuk memperoleh biaya clustering yang

lebih rendah. Penggunaan objek aktual sebagai pusat klaster membuat K-Medoids relatif lebih tahan terhadap pengaruh outlier dibandingkan pendekatan berbasis rata-rata.

2.6. Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) merupakan salah satu metrik internal untuk mengevaluasi kualitas hasil clustering [12]. DBI mempertimbangkan kekompakan dalam klaster dan pemisahan antar-klaster. Nilai DBI yang lebih rendah menunjukkan kombinasi klaster yang lebih kompak dan lebih terpisah. Penelitian ini menggunakan DBI sebagai metrik utama karena metrik yang sama dapat diterapkan pada kedua algoritma dan memungkinkan perbandingan langsung. Penggunaan satu metrik juga menjadi keterbatasan penelitian, sehingga metrik internal tambahan seperti Silhouette Coefficient dan Calinski-Harabasz Index disarankan pada penelitian berikutnya.

2.7. Segmentasi Berdasarkan Data Penjualan

Dalam penelitian ini, istilah segmentasi digunakan untuk menyebut pembentukan kelompok berdasarkan karakteristik numerik data penjualan. Atribut yang digunakan adalah jumlah order, harga, dan total penjualan. Karena data yang tersedia berupa catatan transaksi dan satu jenis produk dapat muncul pada beberapa transaksi, unit analisis penelitian harus dipahami sebagai transaksi penjualan, bukan produk unik, kecuali data telah diagregasi terlebih dahulu. Oleh karena itu, hasil klaster ditafsirkan sebagai segmentasi transaksi berdasarkan karakteristik penjualan.

3. METODE PENELITIAN

Penelitian ini menggunakan teknik clustering untuk mengelompokkan transaksi penjualan. K-Means dan K-Medoids diterapkan pada atribut numerik yang sama setelah preprocessing dan normalisasi. Davies-Bouldin Index (DBI) digunakan untuk mengevaluasi kualitas klaster dan membandingkan hasil kedua algoritma [13].

3.1. Pengumpulan Data

Data penelitian berupa catatan transaksi penjualan yang memuat tanggal, jenis produk, jumlah order, harga, dan total penjualan. Berdasarkan struktur data pada naskah, setiap baris merepresentasikan satu catatan transaksi dan terdapat pengulangan jenis produk pada tanggal atau transaksi yang berbeda. Untuk menjaga reproduktibilitas, informasi berikut perlu dicantumkan dari sumber data asli: sumber dataset [LENGKAPI SUMBER DATA], periode pengamatan [LENGKAPI PERIODE], jumlah seluruh transaksi [LENGKAPI JUMLAH], jumlah produk unik [LENGKAPI JUMLAH], prosedur pengumpulan data [LENGKAPI], serta kriteria inklusi dan eksklusi [LENGKAPI]. Informasi tersebut tidak dapat ditentukan dari dokumen yang tersedia sehingga tidak diisi dengan asumsi.

3.2. Preprocessing Data

Sebelum proses clustering, data dipersiapkan agar konsisten dan dapat digunakan dalam perhitungan jarak. Tahapan preprocessing meliputi data cleaning, seleksi atribut, transformasi seperlunya, dan normalisasi.

3.3. Data Cleaning

Data diperiksa untuk mengidentifikasi nilai kosong, data duplikat, dan nilai yang tidak konsisten. Jumlah data yang ditemukan dan tindakan penanganannya perlu dilaporkan berdasarkan hasil pemeriksaan dataset asli: data kosong [LENGKAPI JUMLAH], data duplikat [LENGKAPI JUMLAH], dan data tidak konsisten [LENGKAPI JUMLAH]. Jika tidak ditemukan masalah, tuliskan jumlahnya sebagai nol dan jelaskan bahwa data dipertahankan.

Tabel 1. Dataset Transaksi Penjualan Produk

No	Tanggal	Jenis Produk	Jumlah Order	Harga	Total
1	05/08/2022	Foodpak260	1000	1800	1800000
2	05/08/2022	FoodpakMatte245	1000	1900	1900000
3	05/08/2022	CraftLaminasi290	5000	750	3750000
4	05/08/2022	CraftLaminasi290	1000	1200	1200000
5	07/08/2022	Dupleks310	1000	1550	1550000
6	08/08/2022	Dupleks310	500	1900	900000
7	08/08/2022	Ivory270	1000	1700	1700000
8	09/08/2022	Kinstruk130	1000	1000	1000000
9	09/08/2022	HVS	1500	850	12755000
….	….	….	….	….	….
1077	15/11/2023	GreaseProof	1000	300	300000

3.4. Seleksi Atribut

Atribut jumlah order, harga, dan total penjualan dipilih karena secara langsung menggambarkan aktivitas penjualan dan tersedia dalam bentuk numerik. Tanggal dan jenis produk digunakan sebagai informasi identitas dan tidak dimasukkan ke perhitungan jarak. Namun, total penjualan secara matematis berkaitan dengan jumlah order dan harga. Oleh karena itu, penelitian ini menggunakan ketiga atribut sebagai representasi

karakteristik penjualan, tetapi hubungan antarfitur merupakan keterbatasan yang perlu dipertimbangkan dalam interpretasi hasil. Analisis korelasi atau uji redundansi fitur dapat ditambahkan jika data mentah tersedia.

Tabel 2. Seleksi Atribut Data Penjualan

No	Jumlah Order	Harga	Total
1	1000	1800	1800000
2	1000	1900	1900000
3	5000	750	3750000
4	1000	1200	1200000
5	1000	1550	1550000
6	500	1900	900000
7	1000	1700	1700000
8	1000	1000	1000000
9	1500	850	1275000
….	….	….	….
1077	1000	300	300000

3.5. Transformasi Data

Atribut jenis produk tidak digunakan dalam perhitungan clustering karena penelitian berfokus pada pembentukan kelompok berdasarkan karakteristik numerik penjualan. Pengkodean jenis produk hanya diperlukan apabila atribut tersebut digunakan pada proses lain; karena atribut ini dikeluarkan dari perhitungan clustering, transformasi kode produk tidak diperlukan untuk algoritma. Jenis produk tetap dapat digunakan pada tahap interpretasi untuk melihat komposisi setiap klaster apabila diperlukan.

Tabel 3. Transformasi Atribut Jenis Produk

No	Tanggal	Jenis Produk	Jumlah Order	Harga	Total
1	05/08/2022	1	1000	1800	1800000
2	05/08/2022	2	1000	1900	1900000
3	05/08/2022	3	5000	750	3750000
4	05/08/2022	3	1000	1200	1200000
5	07/08/2022	4	1000	1550	1550000
6	08/08/2022	4	500	1900	900000
7	08/08/2022	5	1000	1700	1700000
8	09/8/2022	6	1000	1000	1000000
9	09/08/2022	7	1500	850	1275000
….	….	….	….	….	….
1077	15/11/2023	23	1000	300	300000

3.6. Normalisasi Data

Normalisasi dilakukan menggunakan metode Min-Max untuk menyetarakan rentang nilai antaratribut sebelum perhitungan jarak. Rumus yang digunakan adalah x' = (x - x_min) / (x_max - x_min), dengan x sebagai nilai asli, x_min sebagai nilai minimum atribut, dan x_max sebagai nilai maksimum atribut. Hasil normalisasi berada pada rentang 0 sampai 1. Nilai minimum dan maksimum setiap atribut perlu dicantumkan berdasarkan dataset asli untuk memastikan reproduktibilitas.

Tabel 4. Hasil Normalisasi Atribut Data Penjualan

No	Tanggal	Jenis Produk	Jumlah Order	Harga	Total
1	05/08/2022	1	0,02487811	0,017022129	0,083090117
2	05/08/2022	2	0,02487811	0,01802343	0,087746007
3	05/08/2022	3	0,124890611	0,006508461	0,173879958
4	05/08/2022	3	0,02487811	0,011014319	0,055154782
5	07/08/2022	4	0,02487811	0,014518875	0,071450394
6	08/08/2022	4	0,012376547	0,01802343	0,043515059
7	08/08/2022	5	0,02487811	0,016020827	0,078434228
8	09/8/2022	6	0,02487811	0,009011715	0,045843003
9	09/08/2022	7	0,037379672	0,007509763	0,058646699
….	….	….	….	….	….
1077	15/11/2023	23	0,02487811	0,002002603	0,013251778

3.7. Algoritma K-Means

K-Means diterapkan pada data yang telah dinormalisasi menggunakan Euclidean distance. Untuk setiap nilai K, centroid awal dipilih sesuai konfigurasi implementasi yang digunakan. Setiap objek ditempatkan pada klaster dengan jarak terdekat, centroid diperbarui berdasarkan rata-rata anggota klaster, dan proses diulang sampai kondisi berhenti terpenuhi. Konfigurasi random seed dan jumlah pengulangan perlu dicantumkan dari pengaturan RapidMiner yang digunakan: random seed [LENGKAPI], jumlah runs/repetisi [LENGKAPI]. Nilai DBI yang dilaporkan harus dinyatakan apakah merupakan hasil satu run, nilai terbaik, atau rata-rata beberapa run.

a. Menentukan jumlah klaster (K).
b. Memilih titik pusat klaster awal secara acak.
c. Menentukan jarak antara setiap titik data dan titik pusat klaster menggunakan jarak Euclidean.
d. Menempatkan data ke dalam klaster terdekat.
e. Menentukan kembali titik pusat klaster yang baru.
f. Mengulangi proses tersebut hingga titik pusat klaster tidak lagi berubah.

Perhitungan jarak Euclidean menggunakan rumus $d(x,y)=\sqrt{(\sum(x_i-y_i)^2)}$. Penggunaan Euclidean distance diterapkan pada atribut numerik yang telah dinormalisasi sehingga perbedaan skala tidak mendominasi perhitungan jarak. Jarak yang sama digunakan pada K-Medoids agar perbandingan

kedua algoritma dilakukan dengan dasar pengukuran jarak yang konsisten.

$$d(x,y)=\sqrt{\sum(xi-yi)^2}$$

3.8. Algoritma K-Medoids

K-Medoids diterapkan menggunakan implementasi yang tersedia pada RapidMiner [14]. Tahapannya meliputi penentuan K, inisialisasi medoid, penghitungan jarak objek terhadap medoid, penempatan objek ke medoid terdekat, evaluasi biaya, penggantian medoid apabila menghasilkan biaya yang lebih rendah, dan penghentian ketika tidak terdapat perubahan yang memenuhi kriteria optimasi. Untuk reproduktibilitas, parameter implementasi yang digunakan perlu dicantumkan: metode inisialisasi [LENGKAPI], distance measure [LENGKAPI], maximum iterations [LENGKAPI], dan stopping criterion [LENGKAPI].

a. Menentukan jumlah klaster (K).
b. Memilih medoid awal sesuai strategi inisialisasi yang digunakan, kemudian mengevaluasi setiap objek terhadap medoid yang tersedia.
c. Mengelompokkan data berdasarkan jarak terdekat.
d. Menggunakan metode medoid untuk menentukan biaya akhir.
e. Iterasi dilakukan hingga tidak ada perubahan medoid.

3.9. Evaluasi Klaster Menggunakan Davies-Bouldin Index (DBI)

Kualitas klaster dievaluasi menggunakan Davies-Bouldin Index (DBI). DBI menggabungkan ukuran kekompakan dalam klaster dan pemisahan antar-klaster, sehingga nilai yang lebih rendah menunjukkan hasil yang lebih baik [12]. DBI digunakan sebagai metrik utama karena dapat diterapkan secara konsisten pada K-Means dan K-Medoids dalam eksperimen yang sama. Namun, hasil tidak dimaksudkan sebagai bukti universal bahwa satu algoritma selalu lebih unggul, karena kualitas clustering dapat berubah menurut karakteristik dataset dan metrik evaluasi.

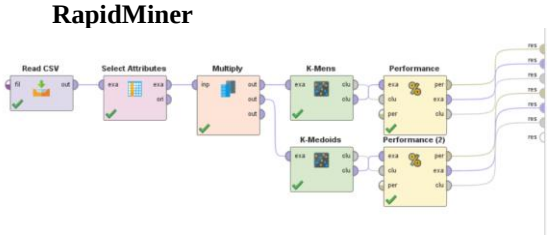
K-Means dan K-Medoids diuji pada K=2 sampai K=7. Rentang tersebut mengikuti konfigurasi eksperimen yang digunakan pada penelitian ini. Karena dokumen tidak mencantumkan dasar domain atau eksperimen awal yang menetapkan batas K=2 dan K=7, alasan pemilihan rentang tersebut perlu ditambahkan dari prosedur penelitian asli. Nilai DBI terendah pada setiap algoritma digunakan untuk menentukan K terbaik masing-masing algoritma.

3.10. Tahapan Penelitian

Tahapan penelitian yang dilakukan adalah sebagai berikut:

a. Pengumpulan data penjualan produk
b. Pra-pemrosesan data
c. Normalisasi data
d. Penggunaan teknik K-Means untuk klasterisasi
e. Penggunaan algoritma K-Medoids untuk klasterisasi
f. Penggunaan DBI untuk mengevaluasi hasil klasterisasi
g. Perbandingan hasil klasterisasi
h. Kesimpulan

3.11. Workflow Penelitian Menggunakan RapidMiner



Gambar 1. Alur Proses Klasterisasi Data Penjualan Menggunakan RapidMiner

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil preprocessing, hasil clustering K-Means dan K-Medoids, perbandingan nilai DBI, serta interpretasi karakteristik klaster. Interpretasi dilakukan berdasarkan atribut yang benar-benar digunakan dalam clustering, yaitu jumlah order, harga, dan total penjualan.

4.1. Dataset Penelitian

Dataset penelitian memuat tanggal, jenis produk, jumlah order, harga, dan total penjualan. Tiga atribut numerik, yaitu jumlah order, harga, dan total penjualan, digunakan sebagai variabel clustering. Berdasarkan struktur data yang ditampilkan, baris data merupakan catatan transaksi sehingga hasil pengelompokan lebih tepat disebut segmentasi transaksi berdasarkan karakteristik penjualan. Jumlah transaksi, periode pengamatan, sumber data, dan jumlah produk unik perlu dilengkapi dari dataset asli.

4.2. Hasil Klasterisasi K-Means

Tabel 5. Hasil Evaluasi K-Means Berdasarkan Davies-Bouldin Index

K	DBI
2	0,709
3	0,572
4	0,429
5	0,456
6	0,56
7	0,456

Hasil K-Means menunjukkan nilai DBI sebesar 0,709 pada K=2, 0,572 pada K=3, 0,429 pada K=4, 0,456 pada K=5, 0,560 pada K=6, dan 0,456 pada K=7. Nilai terendah adalah 0,429 pada K=4. Dengan demikian, berdasarkan DBI, K=4 merupakan jumlah klaster dengan nilai evaluasi terbaik untuk K-Means pada dataset penelitian ini.

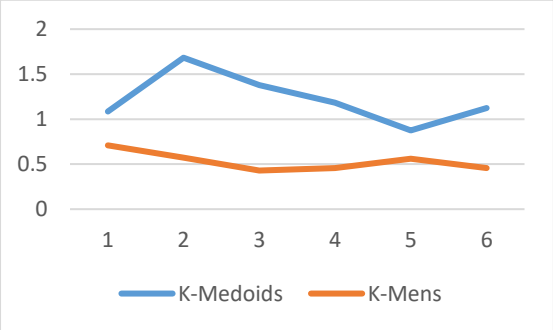
4.3. Hasil Klasterisasi K-Medoids

Tabel 6. Hasil Evaluasi K-Medoids Berdasarkan Davies-Bouldin Index

K	DBI
2	1,084
3	1,683
4	1,379
5	1,183
6	0,875
7	1,125

Hasil K-Medoids menunjukkan nilai DBI sebesar 1,084 pada K=2, 1,683 pada K=3, 1,379 pada K=4, 1,183 pada K=5, 0,875 pada K=6, dan 1,125 pada K=7. Nilai terendah adalah 0,875 pada K=6. Dengan demikian, berdasarkan DBI, K=6 merupakan jumlah klaster dengan nilai evaluasi terbaik untuk K-Medoids pada dataset penelitian ini.

4.4. Perbandingan K-Means dan K-Medoids



Gambar 2. Perbandingan Nilai Davies-Bouldin Index pada Algoritma K-Means dan K-Medoids

Hasil pengujian menunjukkan bahwa K-Means dan K-Medoids menghasilkan nilai DBI terbaik pada jumlah klaster yang berbeda. K-Means memperoleh DBI terendah 0,429 pada K=4, sedangkan K-Medoids memperoleh DBI terendah 0,875 pada K=6. Karena kedua algoritma dievaluasi menggunakan atribut, preprocessing, dan metrik DBI yang sama, nilai tersebut dapat dibandingkan secara langsung pada dataset penelitian.

DBI yang lebih rendah menunjukkan kombinasi kekompakan klaster dan pemisahan antar-klaster yang lebih baik []. Dengan demikian, berdasarkan metrik DBI saja, K-Means menghasilkan nilai yang lebih rendah daripada K-Medoids pada dataset ini. Temuan tersebut merupakan hasil empiris untuk dataset penelitian dan tidak dapat digeneralisasikan sebagai keunggulan universal K-Means. Validasi menggunakan metrik internal tambahan dapat digunakan untuk memperkuat kesimpulan pada penelitian lanjutan.

Hasil clustering membentuk kelompok transaksi dengan karakteristik jumlah order, harga, dan total penjualan yang berbeda. Kelompok tersebut dapat digunakan sebagai informasi deskriptif untuk mendukung analisis persediaan, promosi, dan evaluasi kinerja penjualan. Pemanfaatan hasil klaster untuk keputusan bisnis tetap memerlukan pertimbangan tambahan dari konteks organisasi dan informasi produk yang tidak digunakan dalam clustering.

4.5. Analisis Centroid Klaster

Tabel 7. Nilai Centroid Akhir K-Means pada K=4

Atribut	C0	C1	C2	C3
Jumlah Order	0.031659	0.005251	0.091162	0.545398
Harga	0.015341	1.000000	0.015027	0.001502
Total	0.078185	1.000000	0.233261	0.253242

Tabel 8. Interpretasi Karakteristik Klaster K-Means

Kluster	Karakteristik
C0	Karakteristik centroid rendah
C1	Harga dan total penjualan relatif tinggi
C2	Karakteristik penjualan menengah
C3	Volume order relatif tinggi

Berdasarkan nilai centroid pada Tabel 3, setiap klaster memiliki karakteristik numerik yang berbeda. C1 memiliki nilai centroid harga dan total penjualan yang relatif tinggi, sehingga secara deskriptif dapat disebut sebagai kelompok transaksi dengan harga dan total penjualan relatif tinggi. C3 memiliki nilai centroid jumlah order paling tinggi sehingga dapat disebut sebagai kelompok transaksi dengan volume order relatif tinggi. C2 memiliki karakteristik penjualan pada tingkat menengah berdasarkan nilai centroid yang tersedia, sedangkan C0 memiliki nilai centroid yang relatif rendah pada atribut yang digunakan. Label tersebut hanya menggambarkan karakteristik variabel clustering dan tidak dimaksudkan sebagai penilaian kualitas produk atau kepuasan konsumen.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, K-Means dan K-Medoids dapat digunakan untuk mengelompokkan transaksi berdasarkan jumlah order, harga, dan total penjualan. Berdasarkan Davies-Bouldin Index, K-Means memperoleh nilai terendah sebesar 0,429 pada K=4, sedangkan K-Medoids memperoleh nilai terendah sebesar 0,875 pada K=6. Dengan demikian, pada dataset dan konfigurasi penelitian ini, K-Means menghasilkan nilai DBI yang lebih rendah daripada K-Medoids. K=4 hanya merupakan jumlah klaster terbaik untuk K-Means, sedangkan K=6 merupakan jumlah klaster terbaik untuk K-Medoids; keduanya tidak dapat disebut sebagai satu nilai K optimal untuk kedua algoritma. Hasil centroid menunjukkan perbedaan karakteristik transaksi berdasarkan jumlah order, harga, dan total penjualan. Interpretasi tersebut dapat menjadi informasi pendukung untuk analisis persediaan dan pemasaran, dengan tetap mempertimbangkan keterbatasan atribut dan penggunaan satu metrik evaluasi.

Penelitian selanjutnya disarankan untuk melengkapi informasi provenance dataset, memperluas periode pengamatan dan jumlah data,

serta menjelaskan parameter implementasi kedua algoritma agar penelitian dapat direplikasi. Evaluasi juga dapat menggunakan Silhouette Coefficient dan Calinski-Harabasz Index sebagai metrik tambahan. Hubungan antara jumlah order, harga, dan total penjualan perlu dianalisis, termasuk kemungkinan redundansi fitur. Penelitian berikutnya juga dapat membandingkan K-Means dan K-Medoids dengan Hierarchical Clustering, DBSCAN, atau Fuzzy C-Means serta menambahkan atribut seperti kategori produk, keuntungan, dan frekuensi pembelian apabila tersedia. Jika unit analisis tetap berupa transaksi, penelitian lanjutan dapat mempertimbangkan agregasi per produk apabila tujuan utamanya adalah segmentasi produk unik.

DAFTAR PUSTAKA

[1] I. Gede Iwan Sudipa *et al.*, *Data Mining*. 2022. [Online]. Available: www.globaleksekutifteknologi.co.id

[2] K. Rinci *et al.*, *Buku Ajar Data Mining CV. Luminary Press Indonesia*. 2024. [Online]. Available: www.luminarypress.id

[3] A. R. Firdawanti, H. F. Ahmad, and N. Agustiani, "Analisis Perbandingan Kinerja Algoritma K-Means dan K-Medoids dengan Reduksi Dimensi PCA pada Indikator Kesehatan dan Sosial," *Bulletin of Computer Science Research*, vol. 5, no. 5, pp. 1235–1247, Aug. 2025, doi: 10.47065/bulletincsr.v5i5.742.

[4] Dr. Jonathan Kiwasi Wororomi, Felix Reba, and dkk Samuel Aleksander Mandowen, "Buku Dasar Data Mining," 2021.

[5] A. Fikri Sallaby, R. Tri Alinse, V. Novita Sari, and T. Ramadani, "Pengelompokan Barang Menggunakan Metode K-Means Clustering Berdasarkan Hasil PenjualanDi Toko Widya Bengkulu," *Jurnal Media Infotama*, vol. 18, no. 1, p. 2022, 2022.

[6] M. Yustika, A. Perdana Windarto, Y. Pranavarna Purba, S. Tunas Bangsa, and P. - Indonesia, "Analisis Metode K-Medoids Pada Penjualan Produk Smartphone Vivo Di Kota Pematangsiantar," *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, vol. 3, no. 1, pp. 27–33, 2021.

[7] B. Julio Niceforus *et al.*, "Klasterisasi Data Penjualan Barang Menggunakan Algoritma K-Medoids Untuk Mengoptimalkan Penjualan Sepatu," *Jurnal Teknik Informatika Unis*, vol. 13, no. 2, 2025.

[8] A. Lesmana and W. Gunawan, "Perbandingan Algoritma K-Means dan K-Medoids Dalam Penclusteran Data Penjualan PT. United Teknologi Integrasi Comparison between K-Means and K-Medoids Algorithm in the Collection of Sales Data of PT. United Teknologi Integrasi."

[9] N. A. Rahman and W. Yustanti, "Perbandingan Algoritma Klastering dalam Pengelompokan Penjualan Produk Komputer," 2023.

[10] B. Jejen and Z. Abidin, "Perbandingan Klasterisasi Data Penjualan dengan Menggunakan K-Mean, K-Medoids, dan Fuzzy C-Mean di PT.XYZ".

[11] M. Fithriyah, M. A. Yaqin, and S. Zaman, "K-Means Clustering Untuk Segmentasi Produk Berdasarkan Analisis Recency, Frequency, Monetary (RFM) Pada Data Transaksi Penjualan," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 3, no. 2, pp. 151–164, Aug. 2021, doi: 10.28926/ilkomnika.v3i2.284.

[12] S. Fulvi Intan, W. Elvira, S. Rahayu, and N. Nurfadilla, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Pengeluaran Mahasiswa," *Seminar Nasional Penelitian dan Pengabdian Masyarakat*, pp. 35–40, Aug. 2023, [Online]. Available: https://journal.irpi.or.id/index.php/sentimas

[13] K. Ditha *et al.*, "Penerapan Data Mining Diberbagai Bidang," 2024.

[14] D. Transaksi Bongkar Muat di Provinsi Riau, I. Kamila, U. Khairunnisa, P. Studi Sistem Informasi, and F. Sains dan Teknologi UIN Sultan Syarif Kasim Riau, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019.