

ANALISIS PERBANDINGAN KINERJA METODE K-MEANS DAN AGGLOMERATIVE HIERARCHICAL CLUSTERING DALAM PENGELOMPOKAN DATA PENJUALAN BERDASARKAN EVALUASI DAVIES-BOULDIN INDEX

Fiki¹, Nining Rahaningsih², Irfan Ali³, Faturrohman⁴.

Program Studi Teknik Informatika¹
Program Studi Komputerisasi Akuntansi²
Program Studi Rekayasa Perangkat Lunak^{3,4}

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
fikiiki240@gmail.com

(*) Corresponding Author : fikiiki240@gmail.com
Published : 20 Juli 2026

Abstract—This study aims to analyze product segmentation based on sales data and compare the performance of the K-Means and Agglomerative Hierarchical Clustering (AHC) methods in producing optimal clusters. The data used are 2,555 sales transaction records with attributes of price, quantity sold (qty), and total sales as the main variables in the clustering process. The results show that the best number of clusters obtained is $k = 3$ for both methods based on evaluation using the Davies-Bouldin Index (DBI). The DBI value in the K-Means method is 0.671688, while in the AHC method it is 0.673826. This shows that the K-Means method has slightly better performance than AHC in producing more compact clusters and has clearer separation, although the difference is relatively small. The clustering results obtained are able to divide sales data into three main groups, namely groups with high, medium, and low sales levels. The group with high sales has an average number of sales and total sales that are greater than the other groups, while the low sales group has the smallest value. In addition, Visualization using Principal Component Analysis (PCA) shows that there is a fairly clear separation between clusters, thus strengthening the results of the grouping that has been carried out. Thus, the results of this study indicate that the clustering method can be used to identify sales patterns in a more structured manner and can be used as a basis for decision making, especially in determining marketing strategies and managing product stock.
Keywords: Clustering, K-Means, Agglomerative Hierarchical Clustering, Data Mining, Product Segmentation

Abstrak—Penelitian ini bertujuan untuk menganalisis segmentasi produk berdasarkan data penjualan serta membandingkan kinerja metode K-Means dan Agglomerative Hierarchical Clustering (AHC, dalam menghasilkan cluster yang optimal. Data yang digunakan merupakan data transaksi penjualan sebanyak 2.555 record dengan atribut harga, jumlah terjual (qty), dan total penjualan sebagai variabel utama dalam proses clustering. Hasil penelitian menunjukkan bahwa jumlah cluster terbaik yang diperoleh adalah $k = 3$ untuk kedua metode berdasarkan evaluasi menggunakan Davies-Bouldin Index (DBI). Nilai DBI pada metode K-Means sebesar 0,671688, sedangkan pada metode AHC sebesar 0,673826. Hal ini menunjukkan bahwa metode K-Means memiliki performa yang sedikit lebih baik dibandingkan AHC dalam menghasilkan cluster yang lebih kompak dan memiliki pemisahan yang lebih jelas, meskipun perbedaannya relatif kecil. Hasil clustering yang diperoleh mampu membagi data penjualan ke dalam tiga kelompok utama, yaitu kelompok dengan tingkat penjualan tinggi, sedang, dan rendah. Kelompok dengan penjualan tinggi memiliki rata-rata jumlah terjual dan total penjualan yang lebih besar dibandingkan kelompok lainnya, sedangkan kelompok penjualan rendah memiliki nilai yang paling kecil. Selain itu, visualisasi menggunakan Principal Component Analysis (PCA) menunjukkan bahwa antar cluster memiliki pemisahan yang cukup jelas, sehingga memperkuat hasil pengelompokan yang telah dilakukan. Dengan demikian, hasil penelitian ini menunjukkan bahwa metode clustering dapat digunakan untuk mengidentifikasi pola penjualan secara lebih terstruktur dan dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan, khususnya dalam menentukan strategi pemasaran dan pengelolaan stok produk.

Kata Kunci: Clustering, K-Means, Agglomerative Hierarchical Clustering, Data Mining, Segmentasi Produk

INTRODUCTION

Perkembangan teknologi informasi yang semakin pesat telah mendorong peningkatan volume data di berbagai sektor, khususnya dalam bidang bisnis dan perdagangan. Data penjualan yang dihasilkan secara terus-menerus memiliki potensi besar untuk diolah menjadi informasi yang bernilai dalam mendukung pengambilan keputusan berbasis data (*data-driven decision making*). Namun demikian, pada praktiknya masih banyak pelaku usaha, terutama pada sektor Usaha Mikro, Kecil, dan Menengah (UMKM), yang belum mampu memanfaatkan data penjualan secara optimal akibat keterbatasan dalam pengolahan dan analisis data [1]

Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah teknik *data mining*, khususnya metode *clustering*. Clustering merupakan teknik *unsupervised learning* yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan tanpa memerlukan label sebelumnya. Dalam konteks data penjualan, metode ini mampu mengidentifikasi segmentasi produk, pola pembelian, serta kategori performa penjualan yang dapat digunakan untuk mendukung strategi pemasaran dan pengelolaan stok secara lebih efektif [2]. Variabel jumlah penjualan (*quantity/qty*) juga telah diakui sebagai indikator utama dalam menentukan tingkat permintaan produk dan menjadi dasar dalam analisis segmentasi penjualan [2]

Berbagai metode clustering telah banyak digunakan dalam penelitian sebelumnya, di antaranya adalah K-Means dan Agglomerative Hierarchical Clustering (AHC). K-Means dikenal memiliki keunggulan dalam hal efisiensi dan kemudahan implementasi, terutama untuk data numerik dalam jumlah besar, namun memiliki keterbatasan seperti ketergantungan pada jumlah cluster yang harus ditentukan di awal serta sensitivitas terhadap inisialisasi centroid. Sementara itu, AHC menawarkan keunggulan dalam membangun struktur hierarki data melalui dendrogram sehingga memberikan fleksibilitas dalam menentukan jumlah cluster dan interpretasi hubungan antar data, meskipun memiliki kompleksitas komputasi yang lebih tinggi [3]

Sejumlah penelitian terdahulu menunjukkan bahwa kedua metode tersebut memiliki performa yang bervariasi tergantung

pada karakteristik data dan tujuan analisis. K-Means cenderung unggul dalam kecepatan dan stabilitas, sedangkan AHC lebih unggul dalam aspek interpretabilitas struktur data. Selain itu, evaluasi kualitas cluster umumnya dilakukan menggunakan metrik internal seperti *Silhouette Score* dan *Davies-Bouldin Index (DBI)*, di mana nilai *Silhouette* yang tinggi dan DBI yang rendah menunjukkan kualitas clustering yang baik [4]

Meskipun demikian, masih terdapat kesenjangan penelitian (*research gap*) dalam studi-studi sebelumnya. Sebagian besar penelitian hanya berfokus pada penerapan satu metode clustering atau menggunakan satu metrik evaluasi saja, sehingga belum memberikan perbandingan yang komprehensif antara metode K-Means dan AHC dalam konteks data penjualan, khususnya pada lingkungan UMKM. Selain itu, masih terbatas penelitian yang secara khusus menggabungkan dua metrik evaluasi sekaligus, yaitu *Silhouette Score* dan *Davies-Bouldin Index*, untuk menilai kualitas cluster secara lebih menyeluruh [4]

Berdasarkan permasalahan tersebut, diperlukan suatu penelitian yang secara sistematis membandingkan kinerja metode K-Means dan Agglomerative Hierarchical Clustering pada data penjualan dengan menggunakan atribut utama jumlah penjualan (*qty*) sebagai representasi tingkat permintaan produk. Penelitian ini menjadi penting terutama bagi pelaku UMKM, seperti Warbil, yang memiliki data penjualan namun belum dimanfaatkan secara optimal untuk pengambilan keputusan strategis.

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja metode K-Means dan Agglomerative Hierarchical Clustering dalam mengelompokkan data penjualan berdasarkan evaluasi *Silhouette Score* dan *Davies-Bouldin Index*. Hasil yang diharapkan adalah terbentuknya segmentasi produk berdasarkan tingkat penjualan, seperti kategori produk dengan penjualan tinggi, sedang, dan rendah, yang dapat digunakan sebagai dasar dalam pengambilan keputusan terkait strategi pemasaran, pengelolaan stok, serta peningkatan efisiensi operasional. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi praktis bagi pengembangan usaha berbasis data serta kontribusi akademis dalam pengembangan metode clustering pada analisis data penjualan.

MATERIALS AND METHODS

Penelitian ini dilakukan melalui beberapa

tahapan yang terstruktur untuk menghasilkan proses analisis data yang sistematis dan akurat. Tahapan dimulai dari pengumpulan data penjualan Warbil sebagai sumber utama penelitian, kemudian dilanjutkan dengan preprocessing untuk membersihkan dan memilih atribut data yang relevan. Setelah itu, dilakukan transformasi data menggunakan teknik normalisasi dan reduksi dimensi agar data lebih optimal untuk dianalisis. Tahap berikutnya adalah pemodelan menggunakan metode clustering, yaitu K-Means dan Agglomerative Hierarchical Clustering (AHC). Seluruh proses kemudian dievaluasi menggunakan Davies-Bouldin Index (DBI) untuk menentukan kualitas hasil pengelompokan data yang diperoleh.



Gambar 1. Alur Penelitian

Gambar tersebut menunjukkan alur tahapan penelitian yang digunakan dalam proses analisis data penjualan Warbil berbasis metode clustering. Tahapan pertama adalah akuisisi data, yaitu proses pengumpulan data penjualan yang mencakup atribut harga, jumlah terjual (qty), dan total transaksi. Tahap ini sangat penting karena kualitas data yang dikumpulkan akan memengaruhi hasil analisis pada tahap berikutnya. Data yang diperoleh menjadi dasar utama dalam proses pengolahan dan pengelompokan data.

Tahap kedua adalah preprocessing, yang bertujuan untuk membersihkan data dari kesalahan, data kosong, maupun inkonsistensi

yang dapat mengganggu proses analisis. Pada tahap ini juga dilakukan seleksi atribut sehingga hanya variabel yang relevan, seperti harga, qty, dan total, yang digunakan dalam penelitian. Proses preprocessing membantu meningkatkan kualitas dataset agar lebih siap digunakan dalam pemodelan.

Tahap ketiga merupakan transformasi data. Pada tahap ini digunakan metode normalisasi Min-Max Scaling untuk menyamakan rentang nilai antar atribut sehingga tidak terjadi dominasi data tertentu dalam proses clustering. Selain itu, diterapkan reduksi dimensi menggunakan Principal Component Analysis (PCA) untuk menyederhanakan data tanpa menghilangkan informasi penting yang terkandung di dalamnya. Transformasi data berperan penting dalam meningkatkan efisiensi dan performa model clustering.

Tahap keempat adalah pemodelan menggunakan dua metode clustering, yaitu K-Means dan Agglomerative Hierarchical Clustering (AHC). K-Means digunakan untuk membagi data ke dalam beberapa kelompok berdasarkan kedekatan nilai centroid, sedangkan AHC digunakan untuk membentuk hierarki cluster berdasarkan tingkat kemiripan antar data. Penggunaan dua metode ini bertujuan untuk membandingkan performa pengelompokan sehingga dapat diperoleh metode yang paling optimal dalam mengelompokkan data penjualan Warbil.

Tahap terakhir adalah evaluasi model menggunakan Davies-Bouldin Index (DBI). Metode evaluasi ini digunakan untuk mengukur kualitas cluster berdasarkan tingkat kesamaan data dalam satu cluster dan perbedaan antar cluster. Nilai DBI yang semakin kecil menunjukkan hasil clustering yang semakin baik karena cluster yang terbentuk lebih kompak dan memiliki pemisahan yang jelas. Dengan adanya evaluasi ini, penelitian dapat menentukan model clustering yang paling efektif dan sesuai untuk analisis data penjualan.

RESULTS AND DISCUSSION

Pemodelan

Menentukan cluster k mean dan ahc

Tabel 1 clusther k mean dan ahc

N o	Jumla h Terju al	Harg a	Total Terjua l	Cluste r K- Mean s	Cluste r AHC
1	59	5.000	295.00 0	1	0

2	20	7.000	140.00	1	0
			0		
3	18	6.000	108.00	1	0
			0		
4	30	8.000	240.00	1	0
			0		
5	26	10.00	260.00	0	1
		0	0		

Berdasarkan tabel hasil clustering, data penjualan telah dikelompokkan menggunakan dua metode, yaitu K-Means dan Agglomerative Hierarchical Clustering (AHC). Setiap data memiliki label cluster yang menunjukkan kelompok tempat data tersebut berada berdasarkan tingkat kemiripan karakteristiknya.

Pada metode K-Means, terlihat bahwa sebagian besar data pada sampel, yaitu data ke-1 hingga ke-4, masuk ke dalam cluster 1, sedangkan data ke-5 masuk ke dalam cluster 0. Hal ini menunjukkan bahwa K-Means mengelompokkan data ke dalam dua kategori utama berdasarkan kedekatan terhadap centroid, di mana data dengan pola tertentu cenderung berada dalam kelompok yang sama. Sementara itu, pada metode AHC, terlihat adanya perbedaan hasil pengelompokan. Data ke-1 hingga ke-4 berada dalam cluster 0, sedangkan data ke-5 berada dalam cluster 1. Perbedaan ini menunjukkan bahwa AHC memiliki pendekatan yang berbeda dalam membentuk cluster, yaitu berdasarkan proses penggabungan bertahap yang mempertimbangkan struktur hierarki antar data.

Perbedaan hasil clustering antara K-Means dan AHC pada beberapa data menunjukkan bahwa masing-masing metode memiliki karakteristik dalam menentukan kelompok. K-Means cenderung mengelompokkan berdasarkan pusat cluster, sedangkan AHC mempertimbangkan hubungan antar data secara bertahap. Oleh karena itu, hasil ini perlu dievaluasi lebih lanjut menggunakan metrik evaluasi seperti Davies-Bouldin Index untuk menentukan metode yang menghasilkan cluster terbaik.

Nilai DBI

Tabel 1 nilai DBI

No	Metode	Metrik	K	DBI
1	K-Means	Euclidean	2	0.904403
2	K-Means	Euclidean	3	0.671688
3	K-Means	Euclidean	4	0.754105
4	K-Means	Euclidean	5	0.710389
5	AHC	Euclidean	2	0.904403
6	AHC	Euclidean	3	0.673826

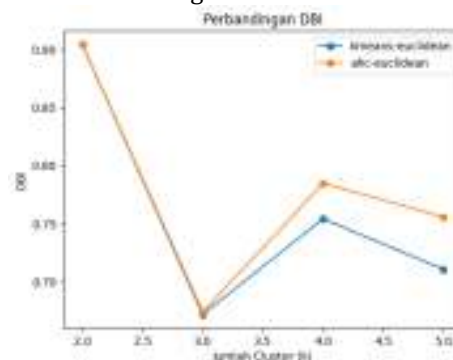
7	AHC	Euclidean	4	0.784635
8	AHC	Euclidean	5	0.755545

Berdasarkan hasil evaluasi menggunakan *Davies-Bouldin Index (DBI)*, dilakukan pengujian terhadap dua metode clustering, yaitu K-Means dan Agglomerative Hierarchical Clustering (AHC), dengan menggunakan metrik jarak Euclidean serta variasi jumlah cluster ($K = 2$ hingga 5).

Pada metode K-Means, nilai DBI mengalami penurunan dari $K = 2$ ke $K = 3$, yaitu dari 0.904403 menjadi 0.671688. Nilai ini merupakan yang paling kecil dibandingkan variasi K lainnya, sehingga menunjukkan bahwa jumlah cluster terbaik pada metode K-Means adalah $K = 3$. Setelah itu, nilai DBI kembali meningkat pada $K = 4$ dan sedikit menurun pada $K = 5$, namun tetap lebih besar dibandingkan $K = 3$.

Pada metode AHC, pola yang serupa juga terjadi, di mana nilai DBI terendah diperoleh pada $K = 3$ sebesar 0.673826. Nilai ini menunjukkan bahwa $K = 3$ merupakan jumlah cluster terbaik untuk metode AHC. Jika dibandingkan antara kedua metode, nilai DBI pada K-Means (0.671688) sedikit lebih kecil dibandingkan AHC (0.673826) pada K yang sama, yaitu $K = 3$. Hal ini menunjukkan bahwa metode K-Means memiliki performa yang sedikit lebih baik dalam mengelompokkan data dibandingkan dengan AHC pada penelitian ini.

Visualisasi Perbandingan DBI



Gambar 2 Perbandingan DBI

Grafik perbandingan Davies-Bouldin Index (DBI) menunjukkan kinerja metode K-Means dan Agglomerative Hierarchical Clustering (AHC) terhadap variasi jumlah cluster ($K = 2$ hingga 5). Terlihat bahwa kedua metode memiliki pola yang hampir sama, di mana nilai DBI mengalami penurunan signifikan dari $K = 2$ ke $K = 3$, kemudian kembali meningkat pada $K = 4$ dan sedikit menurun pada $K = 5$. Nilai DBI terendah diperoleh pada $K = 3$ untuk kedua metode, yaitu sekitar 0,67, yang menunjukkan bahwa jumlah cluster tersebut merupakan yang paling optimal karena menghasilkan cluster yang lebih kompak dan

memiliki pemisahan yang lebih baik. Selain itu, jika dibandingkan antara kedua metode, K-Means secara konsisten menghasilkan nilai DBI yang sedikit lebih rendah dibandingkan AHC pada jumlah cluster yang sama. Hal ini menunjukkan bahwa metode K-Means memiliki performa yang lebih baik dalam mengelompokkan data pada penelitian ini. Secara keseluruhan, berdasarkan grafik tersebut dapat disimpulkan bahwa jumlah cluster terbaik adalah tiga cluster dengan metode K-Means sebagai metode yang paling optimal.

b. karakteristik

Tabel 2 Tabel Karakteristik

K Mean	
Cluster	Rata-rata Jumlah Terjual
0	23.51
1	29.18
AHC	
Cluster	Rata-rata Jumlah Terjual
0	29.18
1	23.51

Berdasarkan hasil analisis karakteristik cluster menggunakan metode K-Means, terlihat bahwa terdapat dua kelompok utama yang dibedakan berdasarkan rata-rata jumlah terjual. Cluster 1 memiliki rata-rata jumlah terjual sebesar 29.18, yang menunjukkan bahwa cluster ini merepresentasikan kelompok produk dengan tingkat penjualan yang lebih tinggi. Sementara itu, cluster 0 memiliki rata-rata sebesar 23.51, yang menunjukkan kelompok produk dengan tingkat penjualan yang relatif lebih rendah.

Pada metode *Agglomerative Hierarchical Clustering* (AHC), diperoleh hasil yang serupa namun dengan penomoran cluster yang berbeda. Cluster 0 pada AHC memiliki rata-rata jumlah terjual sebesar 29.18, sedangkan cluster 1 sebesar 23.51. Hal ini menunjukkan bahwa secara substansi, kedua metode menghasilkan pola pengelompokan yang sama, yaitu membagi data menjadi dua kelompok utama: produk dengan penjualan tinggi dan produk dengan penjualan lebih rendah.

Perbedaan hanya terletak pada label atau penamaan cluster, yang memang bersifat arbitrer dalam metode clustering. Oleh karena itu, interpretasi cluster tidak didasarkan pada nomor cluster, melainkan pada karakteristik nilai yang dimilikinya.

Secara keseluruhan, hasil ini menunjukkan bahwa baik K-Means maupun AHC mampu mengidentifikasi pola segmentasi produk berdasarkan tingkat penjualan dengan konsisten, sehingga dapat digunakan sebagai

dasar dalam pengambilan keputusan, seperti penentuan prioritas stok atau strategi pemasaran.

CONCLUSION

Berdasarkan hasil penelitian, metode K-Means dan Agglomerative Hierarchical Clustering (AHC) mampu mengelompokkan data penjualan berdasarkan kemiripan atribut, yaitu harga, jumlah terjual, dan total penjualan.

Hasil evaluasi menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa jumlah cluster terbaik adalah $K = 3$ untuk kedua metode. Nilai DBI pada K-Means sebesar 0,671688, sedangkan pada AHC sebesar 0,673826.

Perbedaan nilai DBI antara K-Means dan AHC relatif kecil, sehingga kedua metode memiliki performa yang hampir sama, meskipun K-Means sedikit lebih unggul dalam penelitian ini.

Hasil clustering menunjukkan bahwa data penjualan dapat dikelompokkan ke dalam beberapa kategori berdasarkan tingkat penjualan, seperti kelompok dengan penjualan tinggi, sedang, dan rendah.

Jika dibandingkan dengan penelitian terdahulu, hasil penelitian ini menunjukkan kesesuaian di mana metode K-Means sering menghasilkan performa yang baik dalam pengelompokan data penjualan. Namun demikian, pada beberapa penelitian lain, metode AHC juga menunjukkan hasil yang kompetitif tergantung pada karakteristik data yang digunakan. Hal ini menunjukkan bahwa tidak ada satu metode yang selalu paling unggul, melainkan bergantung pada struktur data dan atribut yang digunakan.

REFERENCE

- [1] F. D. Wahyuningtyas, A. Arafat, A. Stiawan, and D. Rolliawati, "Komparasi Algoritma Hierarchical, K-Means, Dan DBSCAN Pada Analisis Data Penjualan Melalui Facebook," *Explore Jurnal Sistem Informasi Dan Telematika*, vol. 14, no. 1, p. 7, 2023, doi: 10.36448/jsit.v14i1.2931.
- [2] R. P. Sari and A. M. Bachtar, "Penerapan Data Mining Pada Data Penjualan Sepatu Untuk Membentuk Segmentasi Distributor Di Cv. Xyz Menggunakan Metode Clustering," *Komputa Jurnal Ilmiah Komputer Dan Informatika*, vol. 5, no. 2, pp.

- 85-90, 2016, doi:
10.34010/komputa.v5i2.2459.
- [3] P. Astaman, "Analisis Klaster Sapi Bali Di Kabupaten Barru, Provinsi Sulawesi Selatan," *Jurnal Ilmu Dan Industri Peternakan*, vol. 8, no. 2, pp. 130-141, 2022, doi: 10.24252/jiip.v8i2.25269.
- [4] J. Faran and R. T. Aldisa, "Analisis Data Mining Dalam Komparasi Average Linkage AHC Dan K-Means Clustering Untuk Dataset Facebook Live Sellers," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, p. 2041, 2023, doi: 10.30865/mib.v7i4.6892.