

Pengelompokan Data Penjualan Produk Cetakan Menggunakan Algoritma K-Means Dengan Bantuan Tool Orange

Susliansyah¹, Muhammad Ridho Caroko², Heny Sumarno³, Hendro Priyono⁴, Linda Maulida^{5,*}

^{1,2,3,4,5} Teknik dan Informatika, Teknologi Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: [1susliansyah.slx@bsi.ac.id](mailto:susliansyah.slx@bsi.ac.id), [2ridhocaroko2@gmail.com](mailto:ridhocaroko2@gmail.com), [3heny.hnm@bsi.ac.id](mailto:heny.hnm@bsi.ac.id),

[3hendro.hop@bsi.ac.id](mailto:hendro.hop@bsi.ac.id), [4,linda.lma@bsi.ac.id](mailto:linda.lma@bsi.ac.id)

(* : coressponding author)

Abstrak- Tujuan penelitian ini adalah mengelompokkan data penjualan produk cetakan menggunakan algoritma K-Means dengan bantuan tool Orange. Permasalahan utama yang dihadapi adalah banyaknya data penjualan yang tidak terstruktur sehingga sulit dianalisis secara manual. Penelitian dilakukan melalui tahapan identifikasi masalah, pengumpulan data, preprocessing, penerapan algoritma K-Means, evaluasi hasil clustering, serta analisis dan interpretasi. Hasil iterasi menunjukkan kluster C1 berisi data penjualan laris, sedangkan kluster C2 mencakup mayoritas data penjualan tidak laris. Evaluasi menggunakan Davies-Bouldin Index (DBI) menghasilkan nilai 0,2818 yang menunjukkan kualitas kluster cukup baik. Sementara itu, Silhouette Plot memberikan nilai 0,082 pada C1 dan 0,276 pada C2, menandakan kluster C2 lebih stabil. Visualisasi Scatter Plot memperlihatkan distribusi data membentuk pola garis miring dari C1 ke C2. Penelitian ini membuktikan bahwa algoritma K-Means efektif dalam mengelompokkan data penjualan, sehingga dapat dijadikan dasar pengambilan keputusan bisnis terkait strategi pemasaran, pengendalian stok, dan evaluasi performa produk.

Kata Kunci: Algoritma K-Means, Pengelompokan, Penjualan, Produk Cetakan, Tool Orange

Abstract- The purpose of this study is to classify sales data of printed products using the K-Means algorithm with the help of the Orange tool. The main problem encountered is the large amount of unstructured sales data, making it difficult to analyze manually. The study was conducted through stages of problem identification, data collection, preprocessing, application of the K-Means algorithm, evaluation of clustering results, as well as analysis and interpretation. Iteration results show that cluster C1 contains data on best-selling products, while cluster C2 encompasses the majority of unsold product data. Evaluation using the Davies-Bouldin Index (DBI) yields a value of 0.2818, indicating fairly good cluster quality. Meanwhile, the Silhouette Plot provides values of 0.082 for C1 and 0.276 for C2, indicating that cluster C2 is more stable. Scatter Plot visualization shows the data distribution forming a diagonal line pattern from C1 to C2. This research proves that the K-Means algorithm is effective in clustering sales data, so it can be used as a basis for business decision-making related to marketing strategies, stock control, and product performance evaluation.

Keywords: K-Means Algorithm, Clustering, Sales, Printed Products, Orange Tool

1. PENDAHULUAN

Data mining merupakan bidang ilmu interdisipliner yang menyatukan teknik pembelajaran dari mesin (machine learning), pengenalan pola (pattern recognition), statistik, basis data, dan visualisasi untuk mengatasi masalah ekstraksi informasi dari basis data yang besar [1], yang merupakan proses untuk mengekstrak informasi yang berguna untuk menentukan pola, hubungan atau tren dalam data yang dapat digunakan untuk mengelompokkan data penjualan yang dapat menentukan produk yang sesuai dengan target pasar dan membuat keputusan bisnis atau meningkatkan penjualan [2]. Dalam data penjualan tersebut dapat dikelompokkan berdasarkan tingkatannya dan melakukan pengolahan data menggunakan Data Mining [3]. Dalam konteks industri percetakan, penjualan berkaitan erat dengan produk cetakan seperti brosur, pamflet, buku, majalah, kalender, hingga kemasan produk, yang berperan penting dalam mendukung promosi, dokumentasi, dan kebutuhan administratif perusahaan.

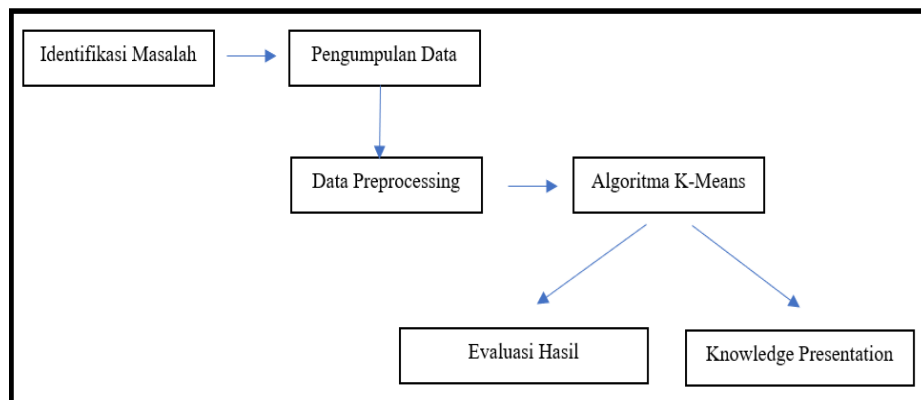
Seiring berkembangnya teknologi, variasi produk cetakan semakin beragam, sehingga menuntut perusahaan untuk mengelola data penjualan secara lebih sistematis. Namun, kenyataannya banyak perusahaan percetakan masih menggunakan metode manual dalam pencatatan penjualan, misalnya dengan buku besar atau lembar kerja sederhana. Cara ini sering menimbulkan kendala berupa duplikasi data, kesalahan pencatatan, serta kesulitan dalam menganalisis tren penjualan. Akibatnya, pengambilan keputusan strategis menjadi kurang akurat dan memakan waktu lebih lama. Oleh karena itu, perusahaan membutuhkan kemampuan untuk memperkirakan volume penjualan dari setiap produk yang dijual [4][5].

Untuk mengatasi permasalahan tersebut, teknik clustering dalam data mining dapat diterapkan. Clustering berfungsi mengelompokkan objek data ke dalam kelas atau cluster berdasarkan informasi yang tersedia, sehingga pola alami dalam data dapat diidentifikasi meskipun sebelumnya tidak memiliki label [6][7][8][9]. Salah satu metode clustering yang populer adalah algoritma K-Means, yang mampu mempartisi data ke dalam beberapa cluster dengan cepat dan efisien. Metode ini bekerja secara iteratif untuk menentukan pusat cluster dan batas-batasnya, dengan tujuan meminimalkan variasi dalam kelompok dan memaksimalkan perbedaan antar kelompok [10][11][12][13]. Menemukan bahwa algoritma K-Means Clustering memiliki kemampuan untuk membedakan wilayah penjualan produk yang rendah, sedang, dan tinggi. Wilayah penjualan rendah dapat diatasi dengan memberikan promosi penjualan, sementara wilayah penjualan tinggi tidak perlu dipromosikan [14].

Penelitian ini difokuskan pada penerapan algoritma K-Means untuk mengelompokkan data penjualan produk cetakan yang tidak terstruktur. Pendekatan ini dipilih karena masih jarang digunakan dalam konteks industri percetakan, padahal data penjualan yang kompleks membutuhkan analisis yang lebih sistematis. Selain itu, penelitian ini menggunakan tool Orange yang memudahkan visualisasi hasil clustering serta dilengkapi evaluasi kuantitatif dengan Davies-Bouldin Index (DBI) dan Silhouette Plot. Dengan demikian, penelitian diharapkan memberikan informasi yang lebih terstruktur mengenai kelompok produk (laris, sedang, tidak laris), pelanggan aktif, serta tren penjualan yang berkembang. Hasilnya dapat mendukung perusahaan dalam menyusun strategi pemasaran yang tepat, meningkatkan efisiensi operasional, dan memperkuat daya saing di industri percetakan [15][16].

2. METODOLOGI PENELITIAN

Adapun tahapan penelitian yang dilakukan dalam menyelesaikan masalah pada pengelompokan data penjualan, dapat dilihat pada gambar 1 dibawah ini:



Gambar 1. Tahapan Penelitian

2.1 Identifikasi Masalah

Penelitian ini berfokus pada tantangan dalam memahami pola penjualan produk cetakan yang beragam dan tidak terstruktur. Banyaknya data transaksi yang tercatat setiap periode membuat analisis manual tidak efisien dan rentan terhadap kesalahan interpretasi. Selain itu, belum adanya sistem pengelompokan yang mampu memisahkan produk berdasarkan tingkat penjualan, harga, atau frekuensi pemesanan menyebabkan pengambilan keputusan bisnis menjadi kurang tepat sasaran. Penelitian ini menetapkan jumlah kluster sebanyak dua ($k=2$), yaitu kluster produk dengan penjualan tinggi (laris) dan kluster produk dengan penjualan rendah (tidak laris). Pemilihan dua kluster dianggap paling relevan karena sesuai dengan tujuan penelitian yang ingin menyoroti perbedaan signifikan antara produk yang berkontribusi besar terhadap penjualan dan produk yang kurang diminati.

2.2 Pengumpulan Data

Dataset diambil dari Kaggle.com, berupa data penjualan produk cetakan dalam format CSV. Data mencakup atribut jumlah order, harga produk, dan total penjualan. Pemilihan atribut ini didasarkan pada relevansinya terhadap analisis penjualan, karena ketiganya merepresentasikan volume transaksi, nilai ekonomi, dan performa produk. Data kemudian diekstrak ke dalam folder kerja yang terstruktur agar mudah digunakan pada tahap preprocessing.

2.3 Data Preprocessing

Dalam tahap preprocessing data untuk analisis clustering, langkah pertama yang dilakukan adalah menghapus data duplikat, memperbaiki kesalahan input, serta mengisi atau menghapus nilai kosong (missing values). Hal ini bertujuan agar data yang digunakan konsisten dan tidak menimbulkan bias. Setelah itu, ditentukan variabel yang relevan untuk clustering, seperti jumlah order, harga produk, dan total penjualan, sementara atribut yang tidak relevan atau berlebihan dihilangkan agar analisis lebih fokus. Selanjutnya, format data diubah ke bentuk yang sesuai, misalnya dari file Excel atau CSV ke format yang dapat dibaca oleh Tool Orange, sekaligus menstandarkan tipe data agar seragam antara angka, teks, maupun kategori. Tahap terakhir adalah melakukan pengecekan ulang untuk memastikan data sudah sesuai dengan kebutuhan analisis, termasuk memastikan tidak ada outlier ekstrem yang dapat mengganggu hasil clustering.

2.4 Algoritma K-means

Algoritma K-means pada dasarnya melakukan dua porses, yakni proses pendeteksian lokasi pusat tiap cluster dan proses pencarian anggota dari tiap-tiap cluster. Cara kerja algoritma K-means [17] [18].

Pada tahap pertama, dilakukan penentuan jumlah kluster (k) yang akan digunakan dalam proses pengelompokan data. Jumlah kluster ditentukan berdasarkan kebutuhan analisis dan tujuan penelitian. Sebagai contoh, jika data penjualan ingin dikelompokkan menjadi dua kategori, yaitu penjualan tinggi dan penjualan rendah, maka nilai k yang digunakan adalah 2. Penentuan jumlah kluster yang tepat sangat penting karena akan memengaruhi hasil pengelompokan yang diperoleh. Tahap berikutnya adalah menentukan centroid awal. Centroid merupakan titik pusat yang menjadi representasi dari setiap kluster. Pemilihan centroid awal dapat dilakukan secara acak dari data yang tersedia atau menggunakan metode tertentu untuk mendapatkan titik awal yang lebih optimal. Centroid awal ini berfungsi sebagai acuan dalam proses pengelompokan data pada iterasi pertama algoritma K-Means.

Setelah centroid awal ditentukan, langkah selanjutnya adalah menghitung jarak antara setiap data dengan masing-masing centroid. Perhitungan jarak umumnya menggunakan metode Euclidean Distance karena sederhana dan banyak digunakan dalam algoritma K-Means. Setiap data akan dihitung jaraknya terhadap seluruh centroid yang ada, kemudian data tersebut akan ditempatkan pada kluster yang memiliki centroid dengan jarak paling dekat. Proses ini bertujuan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik yang dimilikinya.

$$D_{11} = \sqrt{(m1x - c1x)^2 + (m1y - c1y)^2} \quad (1)$$

Pada tahap pengelompokan data ke kluster terdekat, setiap data akan ditempatkan ke dalam kluster yang memiliki centroid dengan jarak paling kecil. Penentuan kluster dilakukan berdasarkan hasil perhitungan jarak antara data dan seluruh centroid yang tersedia. Dengan demikian, data yang memiliki karakteristik serupa akan tergabung dalam kluster yang sama. Hasil dari proses ini adalah terbentuknya sejumlah kluster sesuai dengan nilai k yang telah ditentukan sebelumnya. Setelah seluruh data berhasil dikelompokkan, langkah berikutnya adalah menghitung centroid baru untuk masing-masing kluster. Centroid baru diperoleh dengan menghitung nilai rata-rata dari setiap atribut data yang berada dalam kluster tersebut. Nilai rata-rata ini kemudian digunakan sebagai titik pusat baru yang akan menggantikan centroid sebelumnya. Pembaruan centroid bertujuan untuk mendapatkan representasi kluster yang lebih akurat berdasarkan anggota kluster saat ini. Selanjutnya dilakukan proses iterasi atau pengulangan. Pada tahap ini, perhitungan jarak data ke centroid, pengelompokan data ke kluster terdekat, dan pembaruan centroid dilakukan secara berulang. Iterasi akan terus berlangsung hingga kondisi konvergen tercapai, yaitu ketika tidak terjadi perubahan yang signifikan pada anggota kluster atau posisi centroid telah stabil. Proses pengulangan ini memastikan bahwa hasil pengelompokan yang diperoleh merupakan hasil yang optimal sesuai dengan karakteristik data.

Tahap terakhir adalah memperoleh hasil akhir clustering. Kluster yang terbentuk dianggap final ketika komposisi anggota kluster tidak lagi berubah pada iterasi berikutnya. Hasil pengelompokan ini dapat dimanfaatkan untuk berbagai kebutuhan analisis, seperti mengidentifikasi pola data, melakukan segmentasi pelanggan, mengelompokkan tingkat penjualan, menentukan strategi pemasaran, serta mendukung proses pengambilan keputusan berdasarkan karakteristik masing-masing kluster yang telah terbentuk. produk dengan penjualan laris dan tidak laris

2.5 Evaluasi Hasil Kluster

Dalam evaluasi kualitas hasil clustering menggunakan Davies-Bouldin Index (DBI), terdapat beberapa tahapan penting yang harus dilakukan. Pertama, menghitung rata-rata jarak intra-kluster (within-cluster distance) untuk mengetahui seberapa rapat data dalam satu kluster terhadap centroid-nya. Semakin kecil jarak rata-rata, semakin kompak kluster tersebut. Kedua, menghitung jarak antar kluster (between-cluster distance) dengan cara mengukur jarak centroid satu kluster terhadap centroid kluster lainnya. Semakin besar jarak antar kluster, semakin baik pemisahan antar kelompok. Selanjutnya, dilakukan perhitungan rasio Davies-Bouldin untuk setiap pasangan kluster dengan membandingkan tingkat kemiripan antar kluster. Dari hasil tersebut, ditentukan nilai DBI dengan cara mengambil rasio terbesar dari setiap kluster terhadap kluster lain, kemudian menghitung rata-rata dari semua nilai maksimum rasio tersebut. Interpretasi nilai DBI menjadi tahap akhir: jika nilai DBI rendah, maka kualitas kluster dianggap baik karena kluster lebih kompak dan terpisah jelas; sebaliknya, jika nilai DBI tinggi, kualitas kluster kurang baik karena cenderung tumpang tindih atau tidak terpisah dengan jelas [19] [20]. Adapun rumus mencari DBI sebagai berikut:

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2)$$

2.6 Knowledge Presentation

Dalam tahap visualisasi hasil clustering, langkah pertama adalah menyiapkan data yang sudah terbagi ke dalam kluster setelah proses K-Means selesai. Setiap data akan memiliki label kluster, misalnya C1 dan C2, yang kemudian digunakan sebagai input untuk visualisasi. Selanjutnya, dilakukan visualisasi menggunakan Scatter Plot untuk menampilkan distribusi data berdasarkan atribut tertentu, seperti jumlah order dan total penjualan. Titik-titik data diberi warna berbeda sesuai kluster, sehingga pola pengelompokan terlihat jelas, sementara centroid ditampilkan dengan tanda khusus untuk

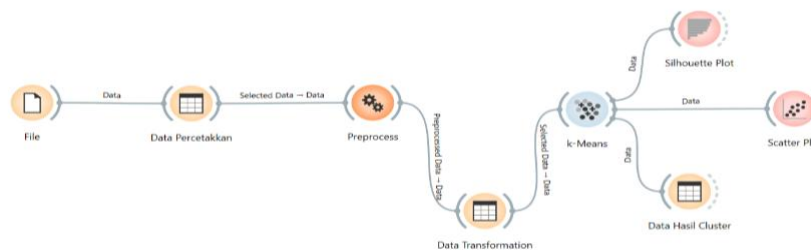
menunjukkan pusat kluster. Setelah itu, kualitas kluster dievaluasi menggunakan Silhouette Plot, yang menunjukkan seberapa rapat data dalam satu kluster dan seberapa jauh dari kluster lain. Nilai Silhouette mendekati 1 menandakan data terkelompok dengan baik, sedangkan nilai mendekati 0 atau negatif menunjukkan data kurang tepat dalam kluster tersebut. Dari Scatter Plot dapat dilihat apakah kluster sudah terpisah dengan jelas atau masih tumpang tindih, sedangkan dari Silhouette Plot dapat dianalisis kluster mana yang lebih stabil, misalnya C2 dengan nilai lebih tinggi dibanding C1.

3. HASIL DAN PEMBAHASAN

Pada tahapan ini, penelitian akan difokuskan pada bagaimana hasil analisis ditampilkan agar mudah dipahami dan digunakan. *Tools Orange* sangat mendukung dalam tahapan ini, karena fungsinya untuk dirancang sebagai *visual programming tool* pada analisis data. Adapun tahapan yang akan dilakukan dalam proses menggunakan tools orange sebagai berikut:

3.1 Alur Kerja Orange

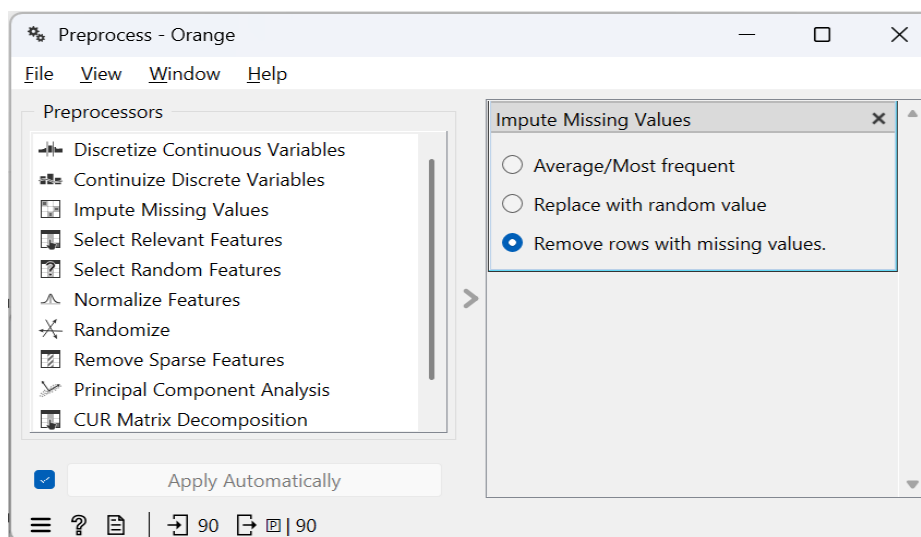
Pada tahapan ini alur kerja orange dilakukan mulai memasukkan data → pra-pemrosesan → memilih metode → menjalankan model → mengevaluasi hasil → menyajikan visualisasi. Adapun alur kerja orange pada penelitian ini dapat dilihat pada gambar 2 dibawah ini:



Gambar 2. Alur Kerja Orange

3.2 Data Processing

Data *processing* dalam Orange adalah tahap pra-pemrosesan data sebelum dilakukan analisis atau pemodelan. Tujuannya agar data lebih bersih, terstruktur, dan siap digunakan sehingga hasil analisis lebih akurat. Adapun hasilnya dapat dilihat pada gambar 3 dibawah ini:

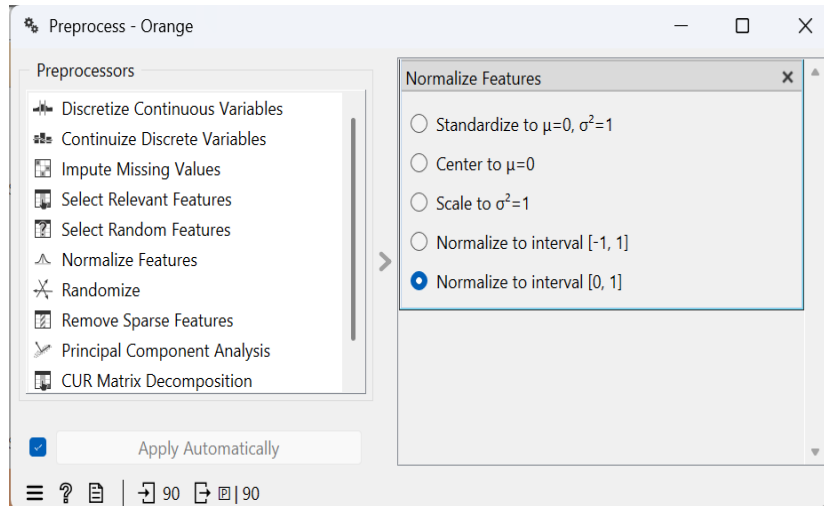


Gambar 3. Data *Processing*

Pada tahapan ini, dimana data penelitian yang digunakan tidak ada data yang bersifat *missing value* atau dataset tidak memiliki nilai yang kosong.

3.3 Data Transformation

Data *transformation* adalah proses mengubah bentuk, skala, atau tipe data agar lebih sesuai untuk analisis. Dalam Orange, tahap ini penting karena banyak algoritma machine learning membutuhkan data dalam format tertentu (misalnya numerik, terstandarisasi, atau bebas dari skala yang berbeda). Adapun hasilnya dapat dilihat pada gambar 4 dibawah ini:



Gambar 4. Data Transformation

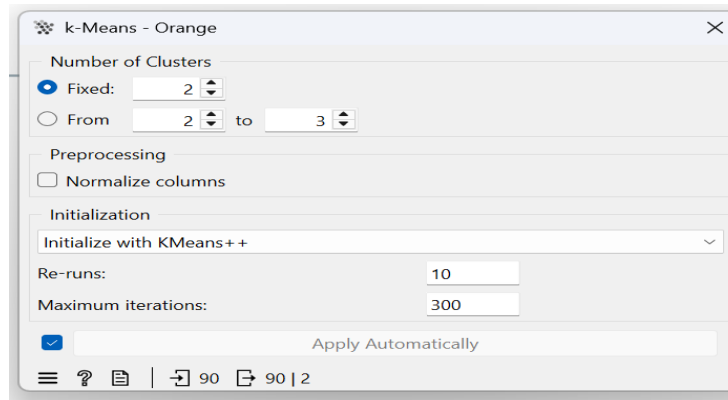
Pada tahapan ini, ketika data diubah ke dalam bentuk normalize to interval (0,1) yang terlihat pada gambar 4, dimana data tersebut berubah. Adapun hasilnya dapat dilihat pada gambar 5 dibawah ini.

	Jumlah Order	Harga	Total	Jenis Produk
1	0.0	0.0	0.0	245
2	0.0	0.0	0.0	260Glossy
3	0.0	0.0	0.0	Bowl650mlCIS
4	0.0	0.0	0.0	BOWL800ML
5	0.0	0.0	0.0	CEMERLANGIN...
6	0.0	0.0	0.0	Craft270
7	0.0	0.0	0.0	Craft275
8	0.0	0.0	0.0	Craft275Biasa
9	0.0	0.0	0.0	Craft290
10	0.0	0.0	0.0	Craft290Foodpak
11	0.0	0.0	0.0	Craft290Laminasi
12	0.0	0.0	0.0	Craft290Lamina...
13	0.0	0.0	0.0	Craft300
14	0.0	0.0	0.0	Craft310
15	0.0	0.0	0.0	CraftBiasa310
16	0.0	0.0	0.0	CraftFoodpak
17	0.0	0.0	0.0	CraftFoodpak290
18	0.0	0.0	0.0	CraftFoodpak2...
19	0.0	0.0	0.0	CraftFoodpak2...
20	0.0	0.0	0.0	CraftLaminasi
21	0.5	1.0	1.0	CraftLaminasi290
22	0.0	0.0	0.0	CraftLaminasi8
23	0.0	0.0	0.0	CraftPackaging8
24	0.0	0.0	0.0	Cupstock
25	0.0	0.0	0.0	Dupleks270
26	0.0	0.0	0.0	Dupleks270Eplex
27	0.0	0.0	0.0	Dupleks270Lam...
28	1.0	0.750	1.0	Dupleks310

Gambar 5. Hasil Data Setelah Transformation

3.4 Algoritma K-Means

Dalam Orange, K-Means tersedia sebagai widget yang memudahkan pengguna melakukan pengelompokan data secara visual tanpa harus menulis kode. Adapun hasilnya dapat dilihat pada gambar 6 dibawah ini:



Gambar 6. Pengelompokan Kluster

Pada gambar diatas, penelitian menggunakan fixed dan klusternya ditentukan 2 karena hanya ada 2 pengelompokan, yaitu Diterima dan Tidak Diterima. Selanjutnya hasil pengelompokan berdasarkan kluster dapat dilihat pada gambar 7 dibawah ini:

	Cluster	Silhouette	Jumlah Order	Harga	Total	Jenis Produk
1	C2	0.555696	0.0	0.0	0.0	245
2	C2	0.555696	0.0	0.0	0.0	260Glossy
3	C2	0.555696	0.0	0.0	0.0	Bowl650mlCIS
4	C2	0.555696	0.0	0.0	0.0	BOWL800ML
5	C2	0.555696	0.0	0.0	0.0	CEMERLANGIN...
6	C2	0.555696	0.0	0.0	0.0	Craft270
7	C2	0.555696	0.0	0.0	0.0	Craft275
8	C2	0.555696	0.0	0.0	0.0	Craft275Biasa
9	C2	0.555696	0.0	0.0	0.0	Craft290
10	C2	0.555696	0.0	0.0	0.0	Craft290Foodpak
11	C2	0.555696	0.0	0.0	0.0	Craft290Laminasi
12	C2	0.555696	0.0	0.0	0.0	Craft290Lamina...
13	C2	0.555696	0.0	0.0	0.0	Craft300
14	C2	0.555696	0.0	0.0	0.0	Craft310
15	C2	0.555696	0.0	0.0	0.0	CraftBiasa310
16	C2	0.555696	0.0	0.0	0.0	CraftFoodpak
17	C2	0.555696	0.0	0.0	0.0	CraftFoodpak290
18	C2	0.555696	0.0	0.0	0.0	CraftFoodpak2...
19	C2	0.555696	0.0	0.0	0.0	CraftFoodpak2...
20	C2	0.555696	0.0	0.0	0.0	CraftLaminasi
21	C1	0.558292	0.5	1.0	1.0	CraftLaminasi290
22	C2	0.555696	0.0	0.0	0.0	CraftLaminasiB
23	C2	0.555696	0.0	0.0	0.0	CraftPackagingB
24	C2	0.555696	0.0	0.0	0.0	Cupstock
25	C2	0.555696	0.0	0.0	0.0	Dupleks270
26	C2	0.555696	0.0	0.0	0.0	Dupleks270Eplex
27	C2	0.555696	0.0	0.0	0.0	Dupleks270Lam...
28	C1	0.557944	1.0	0.750	1.0	Dupleks310
29	C2	0.555696	0.0	0.0	0.0	Dupleks310Eplex

Gambar 7. Hasil Kluster

Pada gambar diatas, penerapan yang dilakukan menggunakan hitungan manual dan hitungan menggunakan tools orange, hasilnya tetap sama, dimana C1 (LARIS) terdiri dari 9 data dan C2 (TIDAK LARIS) terdiri dari 81 data. Artinya kedua perhitungan tersebut hasilnya tidak menunjukkan perbedaan.

3.5 Mencari *Davies Bouldin Index (DBI)*

Perhitungan yang dilakukan untuk mencari DBI dengan menggunakan hasil pada tahapan perhitungan rasio sebagai berikut:

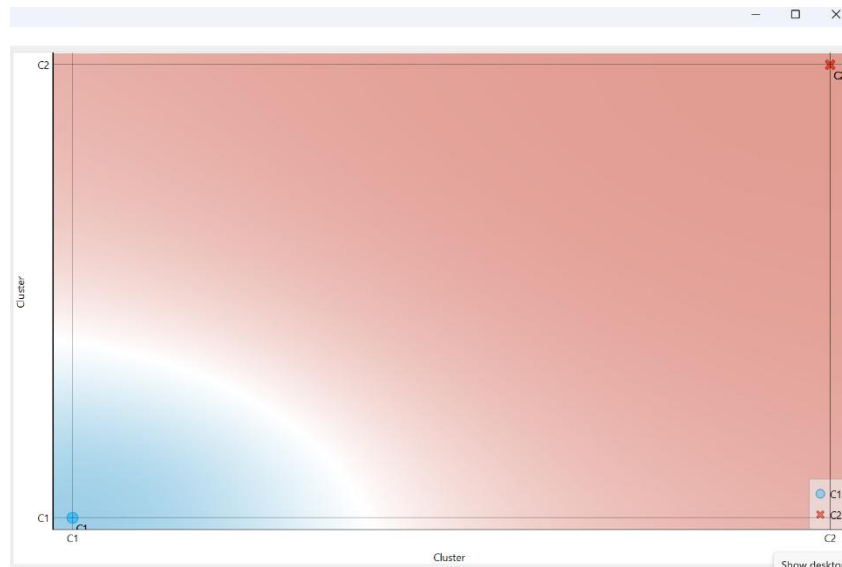
$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

$$DBI = \frac{1}{2} (0.563609805) = 0.281804903$$

Hasil dari perhitungan *Davies-Bouldin Index (DBI)* 0.281804903, menunjukkan tingkat kualitas kluster yang terbentuk dalam proses *clustering* menggunakan Algoritma K-Means. Dengan nilai 0.281804903, dapat diartikan bahwa hasil pengelompokan data memiliki kualitas yang baik, di mana kluster yang terbentuk cukup jelas terpisah satu sama lain dan data dalam masing-masing kluster memiliki kesamaan yang tinggi. Nilai ini menunjukkan bahwa model *clustering* yang digunakan sudah efektif dalam menggambarkan pola data penjualan produk cetakan sesuai dengan tujuan analisis.

3.6 Mencari Scatter Plot dan Shilhouette Plot

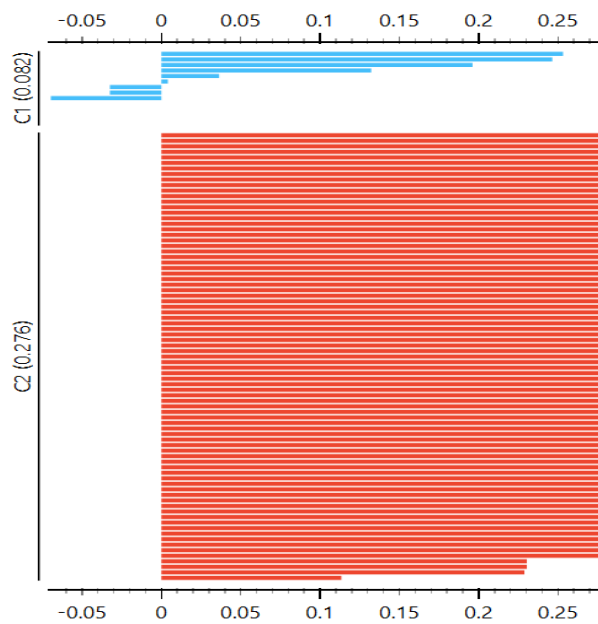
Scatter Plot Penelitian menggunakan tahapan ini dengan tujuan untuk menunjukkan hubungan antara dua variabel dengan cara menempatkan data dalam bentuk titik-titik pada bidang koordinat [21]. Adapun hasilnya dapat dilihat pada gambar 8 dibawah ini:



Gambar 8. Visualisasi Scatter Plot

Jika pada scatter plot titik-titik data membentuk garis miring dari C1 di bagian bawah menuju C2 di bagian kanan atas, maka hal ini menunjukkan adanya hubungan positif antara variabel C1 dan C2. Artinya, setiap kali nilai C1 meningkat, nilai C2 juga ikut meningkat. Pola ini menandakan bahwa kedua variabel memiliki korelasi searah: semakin besar nilai pada sumbu C1, semakin besar pula nilai pada sumbu C2.

Shilhouette Plot Penelitian menggunakan tahapan ini bertujuan untuk mengevaluasi kualitas hasil clustering dengan cara melihat seberapa baik data dalam suatu cluster terkelompok dan seberapa jauh jaraknya dari cluster lain. Adapun hasil dapat dilihat pada gambar 9 dibawah ini:



Gambar 9. Silhouette plot

Hasil Silhouette plot dengan nilai 0,082 pada C1 dan 0,276 pada C2 memberikan gambaran mengenai kualitas pengelompokan data dalam masing-masing kluster. Nilai Silhouette berkisar antara -1 hingga 1, di mana semakin mendekati 1 menunjukkan bahwa data dalam kluster lebih kompak dan terpisah jelas dari kluster lain. Nilai 0,082 pada C1 menunjukkan bahwa data dalam kluster tersebut memiliki tingkat kesamaan yang rendah dan cenderung kurang terpisah dengan baik dari kluster lain, sehingga kualitas pengelompokan pada C1 relatif lemah. Sebaliknya, nilai 0,276 pada C2 lebih tinggi, yang berarti data dalam kluster C2 lebih homogen dan memiliki pemisahan yang lebih baik dibandingkan C1.

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma K-Means efektif dalam mengelompokkan data penjualan produk cetakan menjadi dua kluster, yaitu C1 (produk laris, 9 data) dan C2 (produk tidak laris, 81 data). Evaluasi menggunakan Davies-Bouldin Index (DBI) menghasilkan nilai 0,2818, yang menunjukkan kualitas kluster cukup baik. Silhouette Plot memberikan nilai 0,082 pada C1 dan 0,276 pada C2, menandakan bahwa kluster C2 lebih stabil dibandingkan C1. Visualisasi Scatter Plot memperlihatkan distribusi data membentuk pola garis miring dari C1 ke C2, yang menunjukkan perbedaan karakteristik antara produk laris dan tidak laris. Hasil ini menegaskan bahwa pengelompokan data penjualan dapat digunakan sebagai dasar pengambilan keputusan bisnis, khususnya dalam strategi pemasaran, pengendalian stok, dan evaluasi performa produk. Penelitian selanjutnya disarankan menambahkan variabel lain seperti kategori produk, wilayah distribusi, atau periode waktu penjualan agar hasil analisis lebih relevan dengan kondisi nyata.

REFERENCES

- [1] S. W. Harjono, N. W. Utami, and I. G. A. P. D. Putri, "Klusterisasi Tingkat Penjualan pada Startup Panak.id dengan Algoritma K-Means," *J. Ilm. Teknol. Inf. Asia*, vol. 17, no. 1, pp. 55–66, 2023, doi: 10.32815/jitika.v17i1.888.
- [2] N. P. Gantara and I. Ali, "Penerapan Metode K-Means Clustering Pada Penjualan Barang Di Sports Station," *E-Link J. Tek. Elektro dan Inform.*, vol. 18, no. 1, pp. 28–39, 2023, doi: 10.30587/e-link.v18i1.5339.
- [3] H. Al Rasyid, B. F. K. Soebari, and D. S. Y. Kartika, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penjualan Produk Pada Online Shop Toko Gizi," in *Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI)*, 2022, pp. 242–248. [Online]. Available: <http://sitasi.upnjatim.ac.id/%7C242>
- [4] D. R. Sari, A. P. Windarto, D. Hartama, and Solikhun, "Sistem Pendukung Keputusan untuk Rekomendasi Kelulusan Sidang Skripsi Menggunakan Metode AHP-TOPSIS," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 1, pp. 1–6, 2018, doi: 10.14710/jtsiskom.6.1.2018.1-6.
- [5] Y. D. Putra, M. Sudarma, and I. B. A. Swamardika, "Clustering History Data Penjualan Menggunakan Algoritma K-Means," *Maj. Ilm. Teknol. Elektro*, vol. 20, no. 2, pp. 195–202, 2021.

- [6] S. Setiyanto *et al.*, *Penerapan Teknologi Untuk Penelitian Dan Penyampaian Informasi*. Bandung: Widina Media Utama, 2023.
- [7] F. Juliawati, R. Buaton, and R. Saragih, “Pengelompokan Data Mining Penerimaan Bantuan Pangan Non Tunai (BPNT) Menggunakan Metode Clustering (Studi Kasus : Kantor Desa Payabakung Hampan Perak),” *Explorer (Hayward)*, vol. 3, no. 2, pp. 69–76, 2023, doi: 10.47065/explorer.v3i2.793.
- [8] R. Ariansyah, “Sistem Pendukung Keputusan Penentuan Penerima Bantuan Langsung Tunai (BLT) dengan Kombinasi Metode Clustering K-Means dan Visekriterijumsko Kompromisno Rangiranje (VIKOR),” *ADA J. Inf. Syst. Res.*, vol. 1, no. 3, pp. 129–140, 2024, doi: 10.64366/adajisr.v1i3.51.
- [9] Z. Setiawan, Muhammad Fajar, A. M. Priyatno, A. Y. P. Putri, M. A. Anggi Yhurinda Perdana Putri, and S. Yuliyanti, *Buku Ajar Data Mining*. Jambi: PT. Sonpedia Publishing Indonesia, 2023.
- [10] M. A. M. Hamu and A. C. Talakua, “Optimalisasi Manajemen Bantuan Sosial Dengan Implementasi Data Mining Menggunakan Algoritma K-Means Clustering,” *J. Ilmu Komput. Univ. Katolik Widya Mandir*, vol. 1, no. 1, pp. 7–12, 2024.
- [11] N. Siti Paridah and M. Martanto, “Klasterisasi Penerima Dana Bantuan Program Keluarga Harapan Menggunakan Metode K-Means Pada Desa Gereba,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 1036–1043, 2024, doi: 10.36040/jati.v8i1.8873.
- [12] A. S. Nugroho and Sulyono, “Penerapan Metode K-Means Clustering Untuk Penentuan Eligibilitas Bantuan Langsung Tunai Berbasis Mobile,” *J. Inf. Technol. Ampera*, vol. 5, no. 1, pp. 2774–2121, 2024, doi: 10.51519/journalita.v5i1.510.
- [13] V. Hutagaol, D. Anggraeni, and A. Nata, “PENERAPAN DATA MINING UNTUK CLUSTERING MASYARAKAT KURANG MAMPU PADA DESA RAWANG PASAR VI DENGAN ALGORITMA K-MEANS,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 1, pp. 304–312, 2025.
- [14] T. T. Alifa, R. Astuti, and F. M. Basysyar, “Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Dalam Analisis Penjualan Produk,” *J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 602–607, 2024.
- [15] D. Hartama, Wanayumini, and I. S. Damanik, *Optimasi Algoritma Clustering Studi Kasus Lokasi Bencana Alam Di Indonesia*. Indramayu: PT. Adab Indonesia, 2024.
- [16] Mustika *et al.*, *Data Mining Dan Aplikasinya*. Bandung: Widina Bhakti Persada Bandung, 2021.
- [17] R. M. Sari, A. Rizka, N. A. Putri, and A. Efriana, *Perhitungan Metode Clustering*. Payakumbuh: Serasi Media Teknologi, 2024.
- [18] S. Ufriani, Jasmir, and Y. Arvita, “Penerapan Algoritma Clustering K-Means Untuk Menentukan Prioritas Penerima Bantuan Dana Sosial PKH Di Kelurahan Kampung Singkep,” *J. Inform. Dan Rekayasa Komput. (JAKAKOM)*, vol. 3, no. 1, pp. 342–350, 2023.
- [19] D. Jollyta, M. Siddik, H. Mawengkang, and S. Efendi, *Teknik Evaluasi Cluster Solusi Menggunakan Phytion Dan Rapidminer*. Yogyakarta: deepublish, 2021.
- [20] S. R. Agustin, I. Purnamasari, and B. N. Sari, “Implementasi K-Means Untuk Pengelompokan Kategori Penjualan Barang Berbasis Web,” vol. 5, no. 3, pp. 167–176, 2025, doi: 10.47065/jimat.v5i3.610.
- [21] M. Setiyo and B. Waluyo, *Metodologi Penelitian dan Perancangan Eksperimen*. Magelang: UNIMMA PRESS, 2025.