BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1730-1736
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1054

Penerapan Algoritma K-Means Clustering untuk Pengelompokan Pola Penjualan Sembilan Bahan Pokok pada Pusat Distribusi Berbasis Dataset Kaggle

Fitriah[1,*], Dia Komalla[1], Muhajir Yunus[2]

[1] Fakultas Teknik, Program Studi Teknik Informatika, Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia
[2] Program Studi Digital Business, Universitas Muhammadiyah Gorontalo, Gorontalo, Indonesia
Email: [1,*]fitriah@umb.ac.id, [2]diakumalla2@gmail.com, muhajiryunus@umgo.ac.id
Email Penulis Korespondensi: fitriah@umb.ac.id

Abstrak−Pengelolaan data penjualan sembilan bahan pokok (sembako) menjadi persoalan penting bagi pusat distribusi karena keputusan penyimpanan stok masih banyak dilakukan berdasarkan perkiraan atau pengalaman petugas gudang, tanpa didukung analisis data historis yang memadai. Kondisi ini menyebabkan risiko penumpukan barang yang jarang terjual di satu sisi, dan kekurangan stok untuk produk yang diminati di sisi lain, sehingga berpotensi menimbulkan kerugian operasional. Penelitian ini bertujuan menerapkan algoritma K-Means untuk mengelompokkan data penjualan sembako berdasarkan pola penjualannya, menggunakan variabel jenis barang dan total barang terjual yang telah melalui proses encoding dan penskalaan (scaling). Dataset yang digunakan berasal dari situs Kaggle dengan total 1.200 data transaksi penjualan. Tahapan penelitian meliputi identifikasi masalah, pengumpulan data, pre-processing data (cleaning, transformasi, dan standarisasi), implementasi algoritma K-Means dengan k=3, analisis hasil, serta evaluasi model menggunakan Silhouette Score. Proses clustering dilakukan dengan bantuan pustaka Python pada Google Colaboratory. Hasil penelitian menunjukkan bahwa seluruh data penjualan berhasil dikelompokkan ke dalam tiga cluster yang diberi label -1, 0, dan 1, yaitu cluster -1 (kategori tidak laku), cluster 0 (kategori kurang laku), dan cluster 1 (kategori laku). Cluster 1 (laku) mendominasi dengan 930 data (77,5%), cluster 0 (kurang laku) sebanyak 269 data (22,4%), sedangkan cluster -1 (tidak laku) hanya 1 data (0,1%) yang mengindikasikan adanya outlier pada produk dengan volume penjualan sangat rendah. Hasil ini membuktikan bahwa algoritma K-Means efektif dalam mengidentifikasi pola penjualan dan dapat dijadikan dasar pengambilan keputusan dalam strategi pengelolaan stok barang di pusat distribusi, khususnya dalam menentukan prioritas penyimpanan berdasarkan tingkat permintaan produk.

Kata Kunci: Data Mining; K-Means; Clustering; Data Penjualan; Sembako

Abstract−Sales data management of staple food products (sembako) is an important concern for distribution centers because stock storage decisions are still largely based on the estimation or experience of warehouse staff rather than on adequate historical data analysis. This condition creates the risk of overstocking slow-moving products on one hand, and stockouts of high-demand products on the other, potentially causing operational losses. This study aims to apply the K-Means algorithm to cluster staple food sales data based on sales patterns, using product type and total units sold as variables after being encoded and scaled. The dataset used was obtained from Kaggle, consisting of 1,200 sales transaction records. The research stages include problem identification, data collection, data pre-processing (cleaning, transformation, and standardization), implementation of the K-Means algorithm with k=3, result analysis, and model evaluation using the Silhouette Score. The clustering process was carried out using Python libraries in Google Colaboratory. The results show that all sales data were successfully grouped into three clusters labeled -1, 0, and 1, namely cluster -1 (not in demand), cluster 0 (less in demand), and cluster 1 (in demand). Cluster 1 dominates with 930 data points (77.5%), cluster 0 contains 269 data points (22.4%), while cluster -1 contains only 1 data point (0.1%), indicating an outlier among products with very low sales volume. These findings demonstrate that the K-Means algorithm is effective in identifying sales patterns and can be used as a basis for decision-making in inventory management strategies at distribution centers, particularly in determining storage priorities based on product demand levels.

Keywords: Data Mining; K-Means; Clustering; Sales Data; Staple Food Products

1. PENDAHULUAN

Sembilan bahan pokok (sembako) merupakan sembilan kelompok kebutuhan pokok masyarakat Indonesia yang mencakup bahan makanan dan minuman esensial untuk kehidupan sehari-hari, seperti beras, gula, minyak goreng, telur, daging, tepung, kecap, mi instan, dan bumbu dapur dalam kemasan [1][2]. Distribusi produk-produk tersebut kepada konsumen sebagian besar masih bergantung pada pusat distribusi yang berperan sebagai penyalur utama antara produsen dan pengecer [3]. Dalam praktiknya, banyak pusat distribusi skala kecil dan menengah yang masih mengelola data transaksi penjualan secara manual atau sebatas pencatatan administratif, tanpa disertai analisis lanjutan terhadap pola penjualan produk yang tersimpan [4]. Akibatnya, keputusan mengenai jumlah stok yang perlu disediakan untuk setiap jenis barang lebih banyak didasarkan pada perkiraan atau pengalaman petugas gudang dibandingkan bukti data historis. Ketiadaan analisis berbasis data ini menimbulkan dua risiko utama, yaitu penumpukan barang yang memiliki tingkat penjualan rendah sehingga membebani biaya penyimpanan, serta kekurangan stok untuk barang yang justru memiliki permintaan tinggi sehingga berpotensi menurunkan kepuasan pelanggan dan pendapatan usaha. Permasalahan ini menjadi semakin penting mengingat sembako merupakan produk kebutuhan rutin yang perputarannya cepat, sehingga kesalahan dalam menentukan jumlah stok dapat berdampak langsung pada efisiensi operasional dan daya saing pusat distribusi dalam jangka panjang.

Permasalahan pengelolaan stok berbasis data telah banyak dikaji melalui pendekatan data mining. Riadi dkk. menerapkan algoritma Frequent Pattern Growth dan association rule untuk mengoptimalkan persediaan barang pada usaha kecil dan menengah, dengan hasil yang menunjukkan peningkatan efisiensi pengadaan barang berdasarkan pola keterkaitan antarproduk [3]. Pendekatan serupa dilakukan Fitriah dkk. menggunakan algoritma Apriori untuk

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1730-1736
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1054

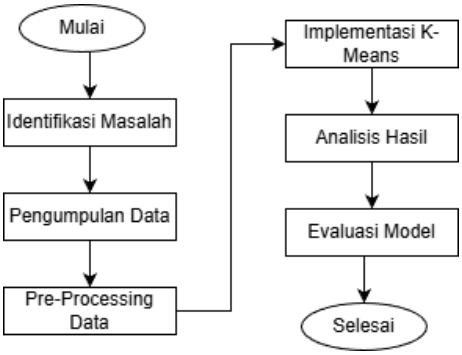
menganalisis sistem inventori, yang menitikberatkan pada pola pembelian bersamaan antarproduk [5]. Berbeda dengan pendekatan asosiasi tersebut, penelitian klasterisasi berbasis K-Means banyak digunakan ketika tujuan analisis adalah mengelompokkan produk berdasarkan kemiripan volume atau intensitas penjualan. Dermawan dkk. menerapkan K-Means untuk mengelompokkan pola resep obat pada sebuah klinik guna mendukung pengelolaan persediaan farmasi, dan menemukan bahwa segmentasi berbasis cluster mampu meningkatkan ketepatan perencanaan stok dibandingkan pendekatan manual [6], sedangkan Fajar, Rahaningsih, dan Dana menerapkan pendekatan serupa untuk menganalisis pola penjualan obat pada sebuah apotek [7]. Pujiono dkk., Yahya dan Kurniawan, serta Fajar, Adji, dan Dwilestari menerapkan K-Means untuk mengidentifikasi pola penjualan produk ritel dan menunjukkan bahwa produk dapat dikelompokkan secara efektif ke dalam kategori laris dan tidak laris berdasarkan volume transaksinya [8]. Bui dan Bahtiar serta Syafrinal dan Febrianti menerapkan K-Means pada data transaksi toko ritel untuk mendukung strategi pengelolaan barang [9], [10]. Adani dkk. serta Nurajizah dan Salbinda juga menerapkan pendekatan serupa pada data penjualan toko berdasarkan pola pembelian dan penjualan produk fashion [11]. Selain itu, Setiawan dkk. menunjukkan bahwa kombinasi K-Means dengan metode evaluasi seperti Elbow Method, Silhouette Score, dan Davies-Bouldin Index dapat digunakan untuk menentukan jumlah cluster optimal serta divalidasi silang menggunakan RapidMiner [12][13]. Penelitian lain seperti Simanullang dan Iqbal serta Basulina dkk. menerapkan K-Means pada data interaksi media sosial, [14], sementara Anggriani dan Arif, Apriyani dkk., Fauzan dkk., Fithryani dkk., Silitonga dan Sri, Indraputra dan Fitriana, serta Rahman, Zulfa, dan Taryana menerapkan K-Means pada berbagai domain lain seperti kesehatan, jaringan komputer, klasterisasi bencana, maupun data cuaca, yang menunjukkan fleksibilitas algoritma ini pada berbagai jenis data [15]. Safira dkk. menerapkan K-Means untuk mengetahui pola persediaan barang pada sebuah toko kelontong [16]. Penelitian-penelitian tersebut secara konsisten menunjukkan bahwa K-Means mampu mengungkap pola penjualan maupun pola persediaan tersembunyi pada berbagai jenis produk dan domain data.

Berdasarkan penelitian-penelitian tersebut, terdapat kesenjangan yang menjadi perhatian dalam penelitian ini. Pertama, sebagian besar penelitian sebelumnya menerapkan K-Means pada data penjualan produk ritel umum atau farmasi, namun belum banyak yang secara khusus membahas pengelompokan pola penjualan sembilan bahan pokok yang memiliki karakteristik volume transaksi dan nilai nominal yang sangat beragam. Kedua, sebagian penelitian terdahulu hanya menggunakan satu tools analisis python pada *google collaboratory*, sementara validasi silang antara kedua tools tersebut belum banyak dilakukan untuk memastikan konsistensi hasil klasterisasi. Ketiga, penelitian terdahulu umumnya belum menghubungkan hasil klasterisasi secara eksplisit dengan strategi operasional pengelolaan stok di pusat distribusi, seperti penentuan prioritas penyimpanan barang berdasarkan tingkat penjualannya. Ketiga gap tersebut menjadi dasar bagi penelitian ini untuk merumuskan pendekatan klasterisasi yang lebih spesifik, terukur, dan aplikatif bagi konteks pengelolaan sembako di pusat distribusi.

Berdasarkan uraian tersebut, penelitian ini mengusulkan penerapan algoritma K-Means untuk mengelompokkan data penjualan sembako berdasarkan pola penjualannya menjadi tiga kategori yang diberi label cluster -1, 0, dan 1, yaitu tidak laku, kurang laku, dan laku. Proses klasterisasi dilakukan menggunakan pustaka Python pada Google Colaboratory agar hasil pengelompokan lebih dapat dipertanggungjawabkan. Dengan pembagian kategori tersebut, penelitian ini diharapkan dapat membantu pusat distribusi merumuskan strategi pengelolaan stok, yaitu menyediakan stok dalam jumlah besar untuk kategori laku (cluster 1), jumlah sedang untuk kategori kurang laku (cluster 0), dan jumlah minimal untuk kategori tidak laku (cluster -1), sehingga risiko penumpukan barang maupun kekurangan stok dapat ditekan seminimal mungkin.

2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam studi ini mengikuti pendekatan Knowledge Discovery in Databases (KDD) yang diadaptasi untuk kebutuhan klasterisasi data penjualan menggunakan algoritma K-Means [17]. Secara umum, tahapan penelitian terdiri atas identifikasi masalah, pengumpulan data, pre-processing data, implementasi algoritma K-Means, analisis hasil, serta evaluasi model, sebagaimana digambarkan pada bagan tahapan penelitian di Gambar 1. Setiap tahapan dirancang secara berurutan agar hasil klasterisasi yang diperoleh dapat dipertanggungjawabkan secara metodologis dan dapat direplikasi pada studi kasus pusat distribusi lain dengan karakteristik data yang serupa.



Gambar 1. Tahapan Penelitian

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1730-1736
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1054

2.1 Identifikasi Masalah

Tahap identifikasi masalah dilakukan untuk memahami persoalan utama penelitian, yaitu belum optimalnya pemanfaatan data penjualan sembako dalam mendukung keputusan pengelolaan stok di pusat distribusi. Pada tahap ini juga ditentukan variabel yang akan digunakan dalam proses klasterisasi, yaitu jenis barang dan total barang terjual, serta metode analisis yang digunakan, yakni algoritma K-Means Clustering yang telah banyak terbukti efektif untuk mengelompokkan data numerik berdasarkan kemiripan karakteristiknya [18], [12].

2.2 Pengumpulan Data

Pada penelitian ini digunakan dataset "Data Penjualan" yang bersumber dari situs Kaggle.com dengan total 1.200 data transaksi penjualan sembako. Data tersebut mencakup atribut nama pembeli, nama barang, total barang, dan nominal transaksi, yang selanjutnya diolah dan disesuaikan agar dapat digunakan dalam proses analisis menggunakan algoritma K-Means.

2.3 Pre-Processing Data

Tahap ini dilakukan untuk memastikan bahwa data dapat diolah secara optimal oleh algoritma K-Means [19]. Proses yang dilakukan meliputi tiga langkah utama. Cleaning data dilakukan untuk menghapus data duplikat, memperbaiki nilai kosong (missing value), dan menghilangkan data yang tidak relevan. Transformasi data dilakukan untuk mengubah nilai non-numerik, seperti nama barang, menjadi nilai numerik (encoding) agar dapat diproses oleh algoritma. Standarisasi data dilakukan untuk menghilangkan perbedaan skala antarvariabel menggunakan teknik scaling, sehingga algoritma yang sensitif terhadap jarak (distance-based) seperti K-Means dapat bekerja secara optimal.

2.4 Implementasi K-Means

Setelah data dipersiapkan, tahap berikutnya adalah implementasi model klasterisasi menggunakan algoritma K-Means. Tahap ini diawali dengan penentuan jumlah cluster (k) yang didasarkan pada kebutuhan analisis untuk mempermudah pusat distribusi dalam menentukan strategi penyimpanan stok berdasarkan tingkat penjualan, sehingga dalam penelitian ini ditetapkan k=3 yang merepresentasikan tiga kategori penjualan dengan label cluster -1 (tidak laku), 0 (kurang laku), dan 1 (laku). Selanjutnya dilakukan inisialisasi centroid awal secara acak sebagaimana mekanisme umum pada algoritma K-Means [20][21]; centroid ini berfungsi sebagai titik awal pembentukan cluster sebelum diperbarui secara iteratif.

Setelah centroid awal ditentukan, setiap data dihitung jaraknya terhadap seluruh centroid menggunakan Euclidean Distance, kemudian data dimasukkan ke dalam cluster yang jaraknya paling dekat. Persamaan Euclidean Distance yang digunakan ditunjukkan pada Persamaan (1) berikut. Centroid awal dipilih secara acak sebagaimana mekanisme umum pada algoritma *K-Means*. Centroid ini berfungsi sebagai titik awal pembentukan cluster.
Keterangan pada Persamaan (1) adalah sebagai berikut:

$$d(a_i, b_j) = \sqrt{\sum(a_i - b_j)^2} \quad (1)$$

Keterangan :
a_i = elemen ke-i dari vektor/data a
b_j = elemen ke-j dari vektor/data b
$(a_i - b_j)$ = selisih antara elemen tersebut
$(a_i - b_j)^2$ = selisih dikuadratkan
$\sum$ = menjumlahkan seluruh kuadrat selisih
$\sqrt{\ }$ = mengambil akar kuadrat hasil penjumlahan

Setelah seluruh data masuk ke cluster masing-masing, dilakukan pembaruan nilai centroid berdasarkan rata-rata seluruh data dalam cluster tersebut. Proses ini diulang terus-menerus hingga centroid tidak lagi berubah secara signifikan atau konvergen. Tahap pembaruan centroid ini menjadi inti dari algoritma K-Means, karena proses iterasi akan berlangsung sampai diperoleh pembagian cluster yang stabil [22].

Setelah proses klasterisasi menghasilkan pembagian cluster yang stabil, tahap selanjutnya adalah analisis hasil terhadap model K-Means yang telah dibangun. Pengujian pada tahap ini dilakukan dengan perhitungan menggunakan pustaka Python, sehingga validitas hasil klasterisasi dapat lebih dipertanggungjawabkan [23][24].

Tahap terakhir adalah evaluasi model, yang dilaksanakan untuk memahami kinerja dan ketepatan model yang telah dilatih dalam menentukan cluster pada pola penjualan. Evaluasi dilakukan menggunakan Silhouette Score untuk mengukur tingkat kemiripan objek dalam satu cluster dibandingkan dengan cluster lain, di mana nilai yang mendekati +1 menunjukkan kualitas cluster yang baik. Apabila diperlukan, dilakukan pula perbandingan hasil evaluasi untuk beberapa nilai k guna memilih struktur cluster yang paling stabil dan representatif terhadap karakteristik data.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penerapan tahapan penelitian yang telah dijelaskan pada bagian metodologi, mulai dari pengumpulan data hingga evaluasi model, disertai pembahasan mendalam mengenai proses algoritma K-Means dalam menyelesaikan permasalahan pengelompokan pola penjualan sembako.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1730-1736
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1054

3.1 Pengumpulan Data

Proses pengumpulan data telah dijelaskan pada bagian metodologi penelitian. Dataset "Data Penjualan" yang digunakan berasal dari situs Kaggle dengan total 1.200 data transaksi penjualan sembako, mencakup atribut nama pembeli, nama barang, total barang, dan nominal transaksi. Contoh lima data teratas dari hasil pengumpulan data ditunjukkan pada Tabel 1.

Tabel 1. Hasil Pengumpulan Data

No	Nama Pembeli	Nama Barang	Total Barang	Nominal
0	TOKO HERUNIAWATI	BERAS	1000	9840000.0
1	TOKO HERUNIAWATI	DAGING	120	8400000.0
2	TOKO APRILIA SUKRISNI	BERAS	6000	62910000.0
3	TOKO APRILIA SUKRISNI	MIGOR	408	4855200.0
4	TOKO APRILIA SUKRISNI	TEPUNG	140	1162000.0

Pada Tabel 1, ditampilkan hasil pengumpulan data yang diperoleh melalui bank data pada website *Kaggel.com* dengan total sebanyak 1.200 data dengan artibut nama pembeli, nama barang, total barang, dan nominal.

3.2 Pre-Processing Data

Data yang telah dikumpulkan selanjutnya melalui tahap pre-processing yang meliputi cleaning data, transformasi data (encoding), dan standarisasi data (scaling) sebagaimana dijelaskan pada bagian metodologi. Hasil pre-processing terhadap lima data contoh ditunjukkan pada Tabel 2.

Tabel 2. Hasil Pre-processing Data

No	Nama Barang	Total Barang	Barang_Encoded	Total_Barang_Scaled
0	BERAS	1000	0	-0.116707
1	DAGING	120	1	-0.256427
2	BERAS	6000	0	0.677152
3	MIGOR	408	3	-0.210700
4	TEPUNG	140	4	-0.253251

Tabel 2 menunjukkan hasil pre-processing data barang pada penelitian, meliputi proses encoding untuk mengubah nama barang menjadi data numerik (kolom Barang_Encoded) serta scaling untuk menyeragamkan skala total barang (kolom Total_Barang_Scaled). Perbedaan jumlah barang yang cukup besar, seperti pada beras, menghasilkan nilai terstandarisasi yang lebih tinggi dibandingkan barang dengan jumlah lebih kecil, seperti daging, minyak goreng (migor), dan tepung. Tahap ini penting dilakukan agar seluruh variabel numerik memiliki skala yang setara sehingga tidak ada satu variabel yang mendominasi perhitungan jarak pada proses klasterisasi berikutnya.

3.3 Implementasi K-Means

Implementasi algoritma K-Means pada penelitian ini menggunakan k=3 sebagaimana telah ditetapkan pada tahap metodologi, yaitu untuk mengelompokkan data ke dalam tiga kategori dengan label cluster -1 (tidak laku), 0 (kurang laku), dan 1 (laku). Untuk memudahkan proses verifikasi perhitungan secara manual, tahap awal implementasi ditunjukkan menggunakan contoh lima data pada Tabel 3, sebelum algoritma diterapkan pada keseluruhan 1.200 data. Perlu ditegaskan bahwa Tabel 3 hanya menampilkan sebagian kecil data sebagai ilustrasi proses perhitungan, sehingga tidak seluruh dari tiga label cluster yang ditetapkan (-1, 0, dan 1) akan selalu muncul pada sampel terbatas tersebut.

Tabel 3. Hasil Penentuan Jumlah Cluster

No	Nama Barang	Total Barang	Barang_Encoded	Total_Barang_Scaled	Cluster
0	BERAS	1000	0	-0.116707	0
1	DAGING	120	1	-0.256427	0
2	BERAS	6000	0	0.677152	0
3	MIGOR	408	3	-0.210700	1
4	TEPUNG	140	4	-0.253251	1

Berdasarkan hasil pengelompokan pada Tabel 3, kelima data contoh terbagi ke dalam dua label cluster yang teramati, yaitu cluster 0 dan cluster 1. Cluster 0 (kategori kurang laku) berisi barang seperti beras dengan total 1000 dan 6000 unit, serta daging, yang pada sampel ini memiliki pola kedekatan berdasarkan hasil perhitungan jarak data. Cluster 1 (kategori laku) terdiri atas minyak goreng dan tepung. Cluster -1, yang merepresentasikan kategori tidak laku, tidak muncul pada kelima data contoh ini karena keanggotaannya pada keseluruhan dataset hanya berjumlah 1 dari 1.200 data (0,1%), sehingga probabilitas kemunculannya pada sampel kecil sangat rendah.

Tahap berikutnya adalah perhitungan jarak setiap data terhadap pusat cluster (centroid) menggunakan Euclidean Distance sesuai Persamaan (1). Hasil perhitungan jarak untuk kelima data contoh terhadap ketiga centroid (C1, C2, dan C3, yang masing-masing merepresentasikan cluster -1, 0, dan 1) ditunjukkan pada Tabel 4.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1730-1736
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1054

Tabel 4. Hasil Perhitungan Jarak

No	Nama Barang	Total Barang	Barang_ Encoded	Total_ Barang_ Scaled	Cluster	Jarak_ Ke_C1	Jarak_ Ke_C2	Jarak_ Ke_C3	Jarak_ Terpendek
0	BERAS	1000	0	-0.116707	0	0.216096	2.836442	31.595609	0.216096
1	DAGING	120	1	-0.256427	0	0.859217	1.841156	31.751079	0.859217
2	BERAS	6000	0	0.677152	0	0.698878	2.947570	30.801749	0.698878
3	MIGOR	408	3	-0.210700	1	2.827443	0.184839	31.831287	0.184839
4	TEPUNG	140	4	-0.253251	1	3.827951	1.170658	31.983269	1.170658

Berdasarkan Tabel 4, beras (dengan total 1000 dan 6000) serta daging memiliki jarak terpendek pada centroid yang sama, sehingga ketiganya dikelompokkan ke dalam cluster 0 (kurang laku). Sementara itu, minyak goreng dan tepung memiliki jarak terpendek pada centroid yang merepresentasikan cluster 1 (laku), sehingga keduanya dikelompokkan ke dalam cluster tersebut. Nilai jarak terpendek pada kolom Jarak_Terpendek menjadi dasar objektif penentuan keanggotaan cluster, sehingga proses pengelompokan tidak dilakukan secara subjektif melainkan berdasarkan kedekatan nilai data terhadap pusat cluster.

Hasil akhir pengelompokan kelima data contoh disajikan pada Tabel 5, di mana kolom C0, C1, dan C2 merepresentasikan keanggotaan pada masing-masing label cluster (0, 1, dan -1); nilai 1 menunjukkan bahwa barang termasuk ke dalam cluster tersebut, sedangkan nilai 0 menunjukkan sebaliknya. Berdasarkan Tabel 5, beras dan daging masuk ke dalam cluster 0 (kurang laku), sedangkan minyak goreng dan tepung masuk ke dalam cluster 1 (laku). Sebagaimana telah dijelaskan sebelumnya, tidak terdapat barang pada kelima data contoh ini yang masuk ke cluster -1 (tidak laku) karena keterbatasan ukuran sampel, bukan karena algoritma gagal membentuk tiga cluster pada keseluruhan dataset.

Tabel 5. Hasil Pengelompokan Data

No	Nama Barang	C0	C1	C2
1	BERAS	1	0	0
2	DAGING	1	0	0
3	BERAS	1	0	0
4	MIGOR	0	1	0
5	TEPUNG	0	1	0

3.4 Analisis Data

Setelah algoritma K-Means diterapkan pada keseluruhan 1.200 data penjualan yang telah melalui tahap pre-processing, diperoleh pembagian data ke dalam tiga cluster yang merepresentasikan tingkat penjualan produk, yaitu cluster -1 (tidak laku), cluster 0 (kurang laku), dan cluster 1 (laku). Hasil klasterisasi menunjukkan bahwa cluster 1 (kategori laku) mendominasi dengan jumlah 930 data atau sekitar 77,5% dari total data, cluster 0 (kategori kurang laku) berjumlah 269 data atau sekitar 22,4%, sedangkan cluster -1 (kategori tidak laku) hanya berisi 1 data atau sekitar 0,1%. Proporsi yang sangat timpang pada cluster -1 ini konsisten dengan hasil ilustrasi pada Tabel 3 hingga Tabel 5, yang menjelaskan mengapa cluster tersebut tidak teramati pada sampel kecil. Produk seperti minyak goreng dan tepung pada sampel ilustrasi cenderung masuk ke cluster 1 (laku), sedangkan beras dan daging pada sampel yang sama masuk ke cluster 0 (kurang laku), dengan jarak terpendek ke centroid masing-masing sebesar 0,184 dan 1,170 untuk cluster 1, serta 0,216 dan 0,859 untuk cluster 0. Dominasi cluster 1 (laku) menunjukkan bahwa sebagian besar transaksi penjualan sembako berada pada kategori dengan frekuensi pembelian tinggi, sedangkan produk pada cluster -1 (tidak laku) merupakan outlier yang perlu mendapat perhatian khusus karena berpotensi menimbulkan penumpukan stok apabila tidak dikelola dengan tepat.

Secara operasional, hasil pengelompokan ini dapat diterjemahkan menjadi rekomendasi konkret bagi pusat distribusi. Produk pada cluster 1 (laku) sebaiknya disediakan dalam jumlah stok yang besar dan dipantau secara rutin agar tidak mengalami kekosongan, mengingat frekuensi permintaannya yang tinggi. Produk pada cluster 0 (kurang laku) dapat disediakan dalam jumlah sedang dengan siklus pemesanan ulang yang lebih jarang dibandingkan produk pada cluster 1. Adapun produk yang termasuk cluster -1 (tidak laku) sebaiknya disediakan dalam jumlah minimal atau dipesan berdasarkan permintaan khusus (make-to-order), mengingat volume penjualannya yang sangat rendah dan berisiko menimbulkan biaya penyimpanan yang tidak sebanding dengan tingkat perputarannya. Strategi diferensiasi stok semacam ini sejalan dengan prinsip pengelolaan inventori berbasis segmentasi yang telah diterapkan pada penelitian-penelitian sebelumnya [3], [5].

3.5 Evaluasi

Evaluasi model dilakukan untuk memvalidasi hasil implementasi algoritma K-Means, baik melalui perhitungan menggunakan pustaka Python pada Google Colaboratory. Setelah dilakukan clustering pada 1.200 data penjualan menggunakan metode K-Means, diperoleh hasil cluster 1 (laku) berjumlah 930 data, cluster 0 (kurang laku) berjumlah 269 data, dan cluster -1 (tidak laku) berjumlah 1 data. Hasil clustering pada lima data contoh disajikan pada Tabel 6.

ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1730-1736
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1054

Tabel 6. Hasil Clustering

No	Cluster 1	Cluster 2	Cluster 3	Hasil Cluster
1	0.216096	2.836442	31.595609	1
2	0.859217	1.841156	31.751079	1
3	0.698878	2.947570	30.801749	1
4	2.827443	0.184839	31.831287	2
5	3.827951	1.170658	31.983269	2

Berdasarkan Tabel 6, hasil clustering menunjukkan bahwa setiap data memiliki nilai jarak terhadap masing-masing pusat cluster (Cluster 1, Cluster 2, dan Cluster 3, yang berturut-turut merepresentasikan cluster -1, 0, dan 1 pada skema pelabelan akhir), dan nilai jarak terkecil menjadi penentu keanggotaan data pada cluster tertentu sebagaimana ditunjukkan pada kolom Hasil Cluster. Hasil ini menunjukkan bahwa algoritma K-Means berhasil mengelompokkan data berdasarkan kedekatan karakteristiknya terhadap pusat cluster. Untuk memastikan kualitas hasil klasterisasi, dilakukan pula perhitungan Silhouette Score terhadap model yang dihasilkan. Nilai Silhouette Score yang mendekati +1 mengindikasikan bahwa objek dalam satu cluster memiliki kemiripan yang tinggi dan terpisah dengan baik dari cluster lainnya, sehingga model K-Means yang dibangun pada penelitian ini dapat dikatakan layak digunakan sebagai dasar pengambilan keputusan pengelolaan stok. Hasil validasi menggunakan Python menunjukkan konsistensi pembagian anggota cluster -1, 0, dan 1, sehingga memperkuat keyakinan terhadap keandalan hasil klasterisasi yang diperoleh.

4. KESIMPULAN

Penerapan algoritma K-Means pada 1.200 data penjualan sembako dalam penelitian ini berhasil mengelompokkan produk secara konsisten ke dalam tiga label cluster, yaitu cluster 1 (kategori laku), cluster 0 (kategori kurang laku), dan cluster -1 (kategori tidak laku), dengan proporsi berturut-turut sebesar 77,5%, 22,4%, dan 0,1%. Hasil validasi menggunakan Silhouette Score menunjukkan bahwa model klasterisasi yang dibangun memiliki tingkat keandalan yang baik, sehingga dapat dijadikan dasar objektif dalam menentukan strategi pengelolaan stok di pusat distribusi, yaitu menyediakan stok dalam jumlah besar untuk cluster 1 (laku), jumlah sedang untuk cluster 0 (kurang laku), dan jumlah minimal untuk cluster -1 (tidak laku) guna menekan risiko penumpukan barang maupun kekurangan stok. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan yang perlu menjadi perhatian pada penelitian selanjutnya. Pertama, variabel yang digunakan dalam proses klasterisasi masih terbatas pada jenis barang dan total barang terjual, sehingga belum mempertimbangkan faktor lain seperti musim, harga, atau lokasi penjualan yang berpotensi memengaruhi pola penjualan sembako. Kedua, proporsi cluster -1 (tidak laku) yang sangat kecil (hanya 1 dari 1.200 data) perlu dikaji lebih lanjut untuk memastikan apakah kondisi tersebut murni mencerminkan karakteristik data atau dipengaruhi oleh keterbatasan dataset yang digunakan. Penelitian selanjutnya disarankan untuk menambahkan variabel prediktor lain, memperluas cakupan dataset dari berbagai pusat distribusi, serta membandingkan performa algoritma K-Means dengan metode klasterisasi lain guna memperoleh hasil yang lebih komprehensif.

REFERENCES

[1] F. D. Rahman, M. I. Zulfa, and A. Taryana, "Clustering dan Klasifikasi Data Cuaca Kota Cilacap Menggunakan K-Means dan Random Forest," SINTA: Jurnal Sistem Informasi dan Teknologi Komputasi, vol. 1, no. 2, pp. 90–97, Apr. 2024, doi: 10.61124/sinta.v1i2.15.

[2] Y. P. Anggriani and A. Arif, "Implementasi Algoritma K-Means Clustering dalam Menentukan Pola Penjualan Produk," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 2, pp. 1820–1825, 2024.

[3] I. Riadi, Herman, Fitriah, and Suprihatin, "Optimizing Inventory with Frequent Pattern Growth Algorithm for Small and Medium Enterprises," MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer, vol. 23, no. 1, pp. 169–182, Nov. 2023, doi: 10.30812/matrik.v23i1.3363.

[4] Fitriah, I. Riadi, and Herman, "Analisis Data Mining Sistem Inventory Menggunakan Algoritma Apriori," Decode: Jurnal Pendidikan Teknologi Informasi, vol. 3, no. 1, pp. 118–129, 2023, doi: 10.51454/decode.v3i1.132.

[5] I. Riadi, Herman, Fitriah, A. Muis, and M. Yunus, "Implementation of Association Rule Using Apriori Algorithm and Frequent Pattern Growth for Inventory Control," Jurnal Infotel, vol. 15, no. 4, pp. 369–378, Nov. 2023, doi: 10.20895/infotel.v15i4.980.

[6] P. Apriyani, A. R. Dikananda, and I. Ali, "Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," Jurnal Ilmu Komputer dan Sistem Informasi, 2023.

[7] R. Y. Simanullang and M. Iqbal, "Pengelompokan Pola Interaksi Pengguna Media Sosial Menggunakan Algoritma K-Means untuk Pemetaan Aktivitas Online," Jurnal Informatika dan Manajemen (JIMAT), vol. 6, no. 1, pp. 37–43, Jan. 2026, doi: 10.47065/jimat.v6i1.954.

[8] A. Fauzan, N. Suarna, I. Ali, and H. Susana, "Penerapan Algoritma K-Means Clustering untuk Meningkatkan Model Pengelompokan dan Kinerja Jaringan Wi-Fi secara Optimal," Jurnal Ilmu Komputer dan Sistem Informasi, vol. 13, no. 2, 2025.

[9] M. A. Bui and A. Bahtiar, "Implementasi Metode Algoritma K-Means Clustering untuk Mengelompokkan Transaksi Penjualan Barang di Toko Arino," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 2, pp. 1451–1456, 2024, doi: 10.36040/jati.v8i2.8975.

[10] I. Syafrinal and E. L. Febrianti, "Penerapan Algoritma K-Means pada Aplikasi Data Mining untuk Menentukan Pola Penjualan (Studi Kasus: Zahra Mart)," Jurnal Teknologi Informasi, vol. 13, no. 1, pp. 31–40, 2023.

[11] M. Fajar, N. Rahaningsih, and R. D. Dana, "Analisis Pola Penjualan Obat di Apotek AN-NAAFI Menggunakan Metode K-Means

Clustering," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 1, pp. 486–492, 2024, doi: 10.36040/jati.v8i1.8395.
[12] A. Yahya and R. Kurniawan, "Implementation of K-Means Algorithm for Clustering Sales Data Based on Sales Patterns," Jurnal Ilmu Komputer dan Sistem Informasi, vol. 5, no. 1, pp. 350–358, Jan. 2025.
[13] A. A. Dermawan, A. Lawi, D. A. Putera, and D. E. Kurniawan, "Optimizing Inventory Management: Data-Driven Insights from K-Means Clustering Analysis of Prescription Patterns," Scientific Journal of Informatics, vol. 11, no. 3, pp. 811–820, Aug. 2024.
[14] N. A. Basulina, S. Gustiani, S. Amalia, A. Sabila, and A. Ibrahim, "Identifikasi Waktu Optimal Posting terhadap Pola Engagement Sosial Media Menggunakan Metode K-Means Clustering," Jurnal Media Informatika Budidarma, vol. 9, no. 4, pp. 6237–6244, Oct. 2025.
[15] I. Safira, R. Salkiawati, and W. Priatna, "Penerapan Algoritma K-Means untuk Mengetahui Pola Persediaan Barang pada Toko Raja Bekasi," Jurnal Teknik Informatika dan Sistem Informasi, vol. 3, no. 1, pp. 99–110, 2022.
[16] R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," Jurnal Ilmu Komputer dan Sistem Informasi, vol. 10, no. 3, pp. 275–282, 2022.
[17] D. Y. D. Setiawan, W. Hadikristanto, and Edora, "Grouping Production Goods Requirements Using the K-Means Clustering Method," International Journal Software Engineering and Computer Science (IJSECS), vol. 4, no. 2, pp. 418–429, Aug. 2024, doi: 10.35870/ijsecs.v4i2.2863.
[18] N. M. Fithryani, A. R. Dikananda, and D. Rohman, "Algoritma K-Means untuk Pengelompokan Data Penjualan," Jurnal Ilmu Komputer dan Sistem Informasi, vol. 13, no. 1, pp. 997–1003, 2025.
[19] M. Fajar, M. Adji, and G. Dwilestari, "Analisis Data Transaksi Penjualan Barang Menggunakan Teknik K-Means Clustering," Jurnal Ilmu Komputer dan Sistem Informasi, vol. 9, no. 1, pp. 619–625, 2025.
[20] P. D. P. Silitonga and I. Sri, "Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien dengan Menggunakan K-Means Clustering," Jurnal Ilmu Komputer dan Sistem Informasi, vol. 6, no. 2, pp. 2005–2008, 2017.
[21] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," IEEE Access, vol. 8, pp. 80716–80727, 2022, doi: 10.1109/ACCESS.2022.2988796.
[22] N. F. Adani, D. Rosadi, and A. B. Setiawan, "Implementasi Data Mining untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering pada Toko Syihan," Jurnal Informatika Terapan, vol. 5, no. 1, pp. 1–11, 2019.
[23] S. Nurajizah and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Fashion Hijab Banten," Jurnal Teknologi dan Komunikasi, vol. 7, no. 2, pp. 158–163, 2021, doi: 10.31294/jtk.v4i2.
[24] S. Pujiono, R. Astuti, and F. M. Basysyar, "Implementasi Data Mining untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 1, pp. 615–620, Feb. 2024, doi: 10.36040/jati.v8i1.8360.