

Pengelompokkan Sembako Terlaris Menggunakan Algoritma K-Means

Julian Firmandanu¹, Abrar Hiswara¹, Fata Nidaul Khasanah^{1*}

¹Informatika, Universitas Bhayangkara Jakarta Raya, Indonesia

Dikirimkan: 27-10-2025

Diterbitkan: 30-11-2025

Keywords:

Clustering;
Data Mining;
K-Means

E-mail Penulis

korespondensi:

fatanidaul@gmail.com

Abstrak. Perkembangan teknologi informasi telah mendorong digitalisasi dalam berbagai sektor, termasuk sektor penjualan ritel. Toko Biru Agen Sembako Bekasi merupakan usaha yang bergerak di bidang distribusi kebutuhan pokok, yang mengalami kendala dalam manajemen stok karena belum adanya sistem informasi yang memadai. Skripsi ini bertujuan untuk mengembangkan sistem informasi berbasis web yang mampu mengelompokkan produk sembako terlaris dengan menggunakan algoritma K-Means Clustering. Metode penelitian yang digunakan meliputi observasi, wawancara, dan studi pustaka, serta pengembangan sistem menggunakan metode prototipe dan framework CodeIgniter. Sistem ini mengelompokkan data penjualan produk berdasarkan tingkat permintaan selama dua tahun (2023–2024) untuk membantu pemilik toko dalam pengambilan keputusan stok dan strategi penjualan. Hasil implementasi menunjukkan bahwa sistem mampu mengelompokkan produk secara akurat ke dalam kategori berdasarkan pola penjualan, serta menampilkan data secara interaktif melalui antarmuka web. Dengan demikian, sistem ini dapat meningkatkan efisiensi operasional dan mengurangi risiko kekosongan atau penumpukan stok barang.

Abstract. The rapid advancement of information technology has driven digitalization across various sectors, including the retail industry. Toko Biru, a basic goods distributor in Bekasi, faces challenges in stock management due to the absence of a proper information system. This thesis aims to develop a web-based information system capable of grouping best-selling products using the K-Means Clustering algorithm. The research methods employed include observation, interviews, and literature study, along with system development using the prototype method demand levels over a two-year period (2023–2024) to assist store owners in making stocking and sales strategy decisions. The implementation results demonstrate that the system can accurately categorize products based on sales patterns and present the data interactively through a web interface. Thus, this system can enhance operational efficiency and reduce the risk of stockouts or overstocking, and the CodeIgniter framework. The system groups product sales data based

1. Pendahuluan

Perkembangan teknologi informasi semakin cepat, baik di bidang kesehatan, pendidikan, maupun usaha. Sistem informasi ini meliputi proses mengumpulkan, dan menganalisis informasi untuk melakukan aktivitasnya supaya mencapai tujuan tertentu. Suatu sistem informasi terdiri dari data (*input*) dan laporan yang dihasilkan (*output*). Salah satu bidang usaha yang dapat memanfaatkan teknologi informasi yaitu bidang penjualan, penjualan merupakan proses dimana penjual dapat memenuhi segala kebutuhan dan keinginan pembeli, sehingga penjual dan pembeli dapat menguntungkan kedua belah pihak [1].

Beragam sembako yang dijual seringkali membuat pemilik toko kesulitan dalam memenuhi permintaan pelanggan dalam beberapa bulan terakhir, sering terjadinya kekosongan stok sembako dan penumpukan stok sembako. Dikarenakan masih belum adanya sistem pencatatan transaksi dan sistem persediaan stok sembako yang masih belum terkomputerisasi sehingga kurang efisien dan akurat dalam persediaan stok sembako yang terlaris dan kurang laris.

Pencatatan data transaksi yang masih menggunakan cara pembukuan sering kali membuat data yang sudah dicatat hilang ataupun rusak dikarenakan umur dari kertas yang singkat sehingga data seringkali tidak sesuai dan membuat penumpukan stok sembako dan kekurangan stok sembako di gudang penyimpanan dikarenakan belum memiliki pola pembelian pelanggan terhadap sembako yang laris dan kurang laris. Sehingga membuat proses perputaran keuangan tidak berjalan dengan baik.

Dalam menghadapi persaingan pasar untuk meningkatkan pendapatan Agen, diperlukan taktik pemasaran sembako yang dijual. Diperlukan adanya analisa untuk mengelompokkan data sembako terlaris dan sembako yang kurang laris. Hasil analisa pengelompokkan data tersebut diharapkan membantu dan memudahkan pemilik toko untuk mengetahui sembako mana yang paling banyak diminati dan yang kurang diminati oleh pelanggan. Analisa tersebut memerlukan data penjualan, penulis menggunakan data penjualan selama dua tahun (2022 – 2023) dengan menggunakan *data mining* dan menerapkan algoritma *K-Means Clustering*.

Data mining merupakan suatu proses menemukan sebuah data informasi penelitian yang berguna, pola dan kesamaan dalam memilih pada sekumpulan data penelitian, memakai teknik algoritma *naïve bayes* dengan pola lainnya misalnya teknik statistika menggunakan aplikasi bantuan seperti *rapidminer* atau *weka* [2][3]. *Algoritma K-means* merupakan perhitungan yang sangat dikenal dalam mencari *itemset* dengan *clustering*. Dengan adanya elemen ini, dapat mempermudah untuk meningkatkan penawaran dengan menunjukkan item yang berkaitan dengan item yang diinginkan pelanggan.

Oleh karena itu, untuk lebih meningkatkan minat beli pelanggan toko, pentingnya mengkaji riwayat transaksi penjualan untuk mengamati pola penjualan sembako yang sering laris dan kurang laris. Tujuan dari penelitian ini untuk mengetahui hasil implementasi algoritma *K-Means* dalam pengelompokkan data penjualan sembako terlaris.

Metode Penelitian

Algoritma *K-Means* merupakan salah satu metode pengelompokkan data non-hierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih 13 kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam kelompok yang lain. Adapun tujuan pengelompokkan data ini adalah untuk meminimalkan fungsi objektif yang diset dalam proses pengelompokkan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok [4].

Ide dasar algoritma *k-means* sangatlah sederhana, yaitu meminimalkan *Sum of Squared Error* (SSE) antara objek – objek data dengan sejumlah *k-centroid*. Algoritma *k-means* bekerja dengan empat langkah, yang di ilustrasikan dalam *pseudocode* dibawah ini. Pertama, dari himpunan data yang akan di klusterisasi, dipilih sejumlah *k*-objek secara acak sebagai *centroid* awal. Kedua, setiap objek yang bukan *centroid* dimasukkan ke klaster terdekat berdasarkan ukuran jarak tertentu. Ketiga, setiap *centroid* diperbarui berdasarkan rata – rata dari objek yang ada di dalam setiap klaster. Keempat, langkah kedua dan ketiga tersebut di ulang-ulang (di iterasi) sampai semua *centroid* stabil atau konvergen dalam arti *centroid* yang dihasilkan dalam iterasi saat ini sama dengan semua *centroid* yang dihasilkan pada iterasi sebelumnya [5].

Karakter *K-means* dapat diringkas menjadi seperti berikut [6] *K-means* merupakan metode pengelompokkan yang sederhana dan dapat digunakan dengan mudah, Pada jenis set data tertentu, *k-means* tidak dapat melakukan segmentasi data dengan baik dimana hasil segmentasinya tidak dapat memberikan pola kelompok yang mewakili karakteristik bentuk alami data, *K-means* bisa mengalami masalah ketika mengelompokkan data yang mengandung *outlier*.

Secara umum, cara kerja dari algoritma *K-Means clustering* adalah sebagai berikut [7]:

- Menentukan *k* sebagai jumlah kluster yang diinginkan.
- Menentukan nilai random untuk pusat *centroid* (*cluster* awal) sebanyak *k*.
- Menghitung jarak setiap data input terhadap masing – masing *centroid* menggunakan rumus jarak *Euclidean* (*Euclidean Distance*) sampai menemukan jarak terdekat dari setiap data dengan *centroid*, berikut adalah persamaan pada rumus *Euclidean Distance*:
$$d(a_i, b_j) = \sqrt{\sum (a_i - b_j)^2}$$

dimana:

ai: data kriteria

bj: *centroid* pada *cluster* ke-j

d. Mengklasifikasikan data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).

e. Memperbaharui nilai *centroid*, nilai *centroid* baru diperoleh dari rata – rata cluster yang bersangkutan dengan menggunakan Rumus 1:

$$bj(t + 1) = \frac{1}{NSj} \sum_{j \in S_j} aj \quad (1)$$

Dimana:

bj(t + 1) : *centroid* baru pada iterasi ke (t+1)

3. Hasil dan Pembahasan

Berikut adalah perhitungan K-Means clustering dalam menentukan pengelompokkan sembako terlaris dari data penjualan tahun 2023 dan 2024. Tabel 1 menyajikan data sampel dalam penelitian.

Tabel 1. Data Sampel

Jenis Sembako	QTY	
	01/01/2023	01/01/2024
Minyak	5	4
Beras	12	10
Gula	9	7
Tepung terigu	8	10

Sumber: Hasil Pengolahan Data (2025)

Inisialisasi *Centroid*:

1. *Centroid* 1 = Minyak pada 01/01/2023 (QTY = 5)

2. *Centroid* 2 = Beras pada 01/01/2023 (QTY = 12)

Langkah 1 : Menghitung Jarak Euclidean untuk setiap data ke *centroid*

Menghitung jarak *Euclidean* antara setiap titik data dengan dua *centroid* (*Centroid* 1 dan *Centroid* 2) untuk kedua tahun (2023 dan 2024).

Rumus jarak *Euclidean* adalah:

$$\text{Jarak} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Dimana x_1, y_1 adalah nilai dari 01/01/2023 (Qty) dan 01/01/2024 (Qty) untuk setiap data, dan x_1, y_1 adalah *Centroid* 1 dan *Centroid* 2.

Jarak ke *Centroid* 1 (Minyak) dan *Centroid* 2 (Beras):

Centroid 1 (Minyak) = (5, 4)

Centroid 2 (Beras) = (12, 10)

Perhitungan jarak pada setiap data:

1. Minyak

Jarak ke *Centroid* 1

$$= \sqrt{(5 - 5)^2 + (4 - 4)^2} = \sqrt{(0^2 + 0^2)} = 0$$

Jarak ke *Centroid* 2

$$= \sqrt{(5 - 12)^2 + (4 - 10)^2} = \sqrt{(7^2 + 6^2)} = \sqrt{(49 + 36)} = \sqrt{85} = 9.22$$

2. Beras

Jarak ke *Centroid* 1

$$= \sqrt{(12 - 5)^2 + (10 - 4)^2} = \sqrt{(7^2 + 6^2)} = \sqrt{(49 + 36)} = \sqrt{85} = 9.22$$

Jarak ke *Centroid* 2

$$= \sqrt{(12 - 12)^2 + (10 - 10)^2} = \sqrt{(0^2 + 0^2)} = 0$$

3. Gula

Jarak ke *Centroid* 1

$$= \sqrt{(9 - 5)^2 + (7 - 4)^2} = \sqrt{(4^2 + 3^2)} = \sqrt{(16 + 9)} = \sqrt{25} = 5$$

Jarak ke *Centroid* 2

$$= \sqrt{(9 - 12)^2 + (7 - 10)^2} = \sqrt{(3^2 + 3^2)} = \sqrt{(9 + 9)} = \sqrt{18} = 4.24$$

4. Tepung

Jarak ke *Centroid* 1

$$= \sqrt{(8 - 5)^2 + (10 - 4)^2} = \sqrt{(3^2 + 6^2)} = \sqrt{(9 + 36)} = \sqrt{45} = 6.71$$

Jarak ke *Centroid* 2

$$= \sqrt{(8 - 12)^2 + (10 - 10)^2} = \sqrt{(4^2 + 0^2)} = \sqrt{16} = 4$$

Tabel 2 menyajikan data hasil perhitungan jarak untuk masing-masing jenis sembako melalui proses perhitungan yang telah dilakukan.

Tabel 2. Hasil Perhitungan Jarak

Jenis Sembako	QTY		Jarak ke <i>Centroid</i> 1	Jarak ke <i>Centroid</i> 2	Cluster Terdekat
	01/01/2023	01/01/2024			
Minyak	5	4	0	9.22	<i>Centroid</i> 1
Beras	12	10	9.22	0	<i>Centroid</i> 2
Gula	9	7	5	4.24	<i>Centroid</i> 2
Tepung terigu	8	10	6.71	4	<i>Centroid</i> 2

Sumber: Hasil Pengolahan Data (2025)

Langkah 2: Penetapan Cluster

Berdasarkan jarak yang dihitung, data akan dikelompokkan ke centroid terdekat:

Cluster 1 (*Centroid* 1): Minyak

Cluster 2 (*Centroid* 2): Beras, Gula, Tepung

Langkah 3: Update Centroid Baru

Tahap selanjutnya, menghitung *centroid* baru berdasarkan rata-rata dari qty pada setiap *cluster*.

Centroid Baru untuk Tahun 2023 dan 2024:

Cluster 1 (*Centroid* 1 - Minyak): Hanya ada 1 data (Minyak), jadi *centroid* tetap = (5, 4).

Cluster 2 (*Centroid* 2 - Beras, Gula, Tepung):

$$\text{Rata-rata untuk 2023} = \frac{12+9+8}{3} = \frac{29}{3} = 9.67$$

$$\text{Rata-rata untuk 2024} = \frac{10+7+10}{3} = \frac{27}{3} = 9$$

Jadi *Centroid* 2 yang baru adalah (9.67, 9).

Langkah 4: Iterasi Jika Centroid Berubah

Jika centroid berubah, maka perlu menghitung jarak lagi dengan centroid yang baru. Tetapi karena *Centroid* 1 tetap sama (Minyak) dan *Centroid* 2 berubah sedikit (9.67, 9), maka dapat melanjutkan dengan pembagian cluster yang sudah ada.

4. Kesimpulan

Dalam Pengelompokan data penjualan sembako terlaris pada toko sembako dengan mengimplementasikan data mining menggunakan algoritma K-means. Hasil langkah perhitungan K-Means menunjukkan centeroid 1 minyak (5, 4) dan centeroid 2 (beras, gula, tepung) dengan nilai (9.67, 9). Saran untuk selanjutnya dapat menggunakan model algoritma pengelompokan yang lainnya untuk dapat mengetahui perbandingan hasil.

Daftar Rujukan

- [1] Y. Firmansyah, R. Maulana, and D. O. Hutagalung, "Implementasi Model Prototipe Dalam Pembuatan Sistem Informasi Penjualan Sparepart," *J. Sist. Inf. Akunt.*, vol. 2, no. 1, pp. 63–71, 2021, doi: 10.31294/justian.v2i01.366.
- [2] F. N. Khasanah, "Klasifikasi Proses Penjurusan Siswa Tingkat SMA Menggunakan Data Mining," *Informatics Educ. Prof.*, vol. 1, no. 1, pp. 65–69, 2016.
- [3] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, 2019.
- [4] F. A. Bramasta and R. Halilintar, "Penerapan Data Mining Untuk Menentukan Strategi Penjualan Toko Sepatu," *Pros. SEMNAS INOTEK ...*, pp. 236–241, 2021.
- [5] R. Gustrianda and D. I. Mulyana, "Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 27, 2022, doi: 10.30865/mib.v6i1.3294.
- [6] N. Sirait, "Implementasi K-Means Clustering Pada Pengelompokan Mutu Biji Sawit," *J. Pelita Inform.*, vol. 16, no. 4, pp. 368–372, 2017.
- [7] M. S. Nawawi, F. Sembiring, and A. Erfina, "Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris," *Semin. Nas. Teknol. Inf. dan Komunikasi-2021*, pp. 789–797, 2021.