

IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PENJUALAN PRODUK PADA ONLINE SHOP TOKO GIZI

IMPLEMENTATION OF K-MEANS CLUSTERING ALGORITHM FOR GROUPING PRODUCT SALES IN ONLINE SHOP TOKO GIZI

Harun Al Rasyid¹⁾, Brian Firmansyah Kartono Soebari²⁾, Dhian Satria Yudha Kartika³⁾
E-mail : ¹⁾19082010065@student.upnjatim.ac.id, ²⁾19082010073@student.upnjatim.ac.id,
³⁾dhian.satria@upnjatim.ac.id

^{1,2,3}Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa timur

Abstrak

Pengusaha bisnis terutama yang berjualan di *Online Shop* tentunya memiliki target penjualan dan transaksi dalam jumlah besar pada umumnya, maka dari itu Toko Gizi yang berjualan di salah satu *e-commerce* merupakan perusahaan yang berjualan di bidang Kebutuhan Susu, Obat, dan sebagainya, memiliki sebuah permasalahan dimana stok produk yang berputar sangat cepat mengakibatkan seringkali pesanan harus ditunda dikarenakan ketersediaan barang yang kerap kali kosong, Toko Gizi memiliki riwayat transaksi yang berisi banyaknya penjualan dan transaksi, dimana nantinya data tersebut dapat digunakan dalam *Data Mining* untuk melihat perputaran barang dari yang kurang diminati hingga sangat diminati, yang harapannya dapat digunakan untuk Toko Gizi dalam mempersiapkan stok barang yang disesuaikan dari penjualan dalam rentang waktu tertentu. Penelitian *Data Mining* yang kami gunakan yakni *clustering* yang merupakan metode digunakan untuk mengetahui tingkat penjualan berdasarkan terbentuknya *cluster*. Algoritma *K-Means* ini juga cocok digunakan untuk mengatur riwayat data pembelian dan penjualan karena implementasinya yang efisien, tidak perlu memakan banyak waktu serta penerapannya cukup mudah dibandingkan teknik *clustering* lainnya. Dari penelitian yang sudah dilakukan telah didapatkan hasil dari metode Clustering algoritma K-Means yang menunjukkan bahwa dari 41 jenis data penjualan dan transaksi yang dilakukan dari bulan April hingga Mei 2022 menghasilkan 3 cluster yaitu sangat diminati(3), cukup diminati(2) dan kurang diminati(1). Dimana dari beberapa cluster tersebut menunjukkan bahwa cluster pertama terdapat 5 barang, cluster kedua terdapat 5 barang dan cluster ketiga terdapat 1 barang.

Kata kunci: *Online Shop, Clustering, K-Means, Data Mining, e-Commerce.*

Abstract

Business entrepreneurs, especially those selling in Online Stores, of course have sales targets and transactions in large numbers in general, therefore the Nutrition Shop that sells in one of the e-commerce is a company that sells in the fields of Milk Needs, Medicines, and so on, has a problem product stock that rotates very quickly where orders often have to be postponed due to the availability of items that are often empty, Gizi Stores has a transaction history that contains many and transactions, later this data can be used in Data Mining to see the turnover of goods from less desirable to very in demand, it is hoped that it can be used for Nutrition Stores in preparing customized stock items from sales within a certain timeframe. The Data Mining research that we use is clustering which is a method used to determine the level of sales based on the formation of clusters. The K-Means algorithm is also suitable for managing historical purchase and sales data because of its efficient implementation, does not require much time and is quite easy to implement compared to other clustering techniques.

Keywords: *Online Shop, Clustering, K-Means, Data Mining, e-Commerce.*

1. PENDAHULUAN

Perkembangan teknologi informasi yang saat ini bisa dibilang sangat pesat dalam berbagai bidang saat ini menghasilkan data yang sangat berlimpah. Contohnya bidang pendidikan, industri, ekonomi, ilmu dan teknologi dan masih banyak lagi bidang kehidupan yang sekarang ini mampu diambil datanya dan pastinya data yang diambil sangat banyak. Apalagi di dunia bisnis, salah satu alasannya adalah karena pada saat ini banyak toko *online* atau *e-commerce* yang bermunculan dalam berbagai jenis barang yang dijual. Sebutan lainnya untuk perdagangan *online* atau toko *online* bisa juga disebut dengan *Online Shop*.

Semakin berkembang dan ramai toko *online* tersebut otomatis akan berpengaruh dengan banyaknya transaksi yang dilakukan, serta data yang didapatkan akan semakin banyak juga. Agar tahu informasi lebih lengkap dari data yang didapatkan maka data tersebut harus diolah dengan baik. Salah satu informasi yang bisa diambil adalah jumlah penjualan produk terbanyak pada suatu perusahaan, dimana data yang diambil dapat diukur atau dihitung dari banyaknya jumlah transaksi dan penjualan pada setiap produknya. Dalam data penjualan tersebut dapat dikelompokkan berdasarkan tingkatannya dan melakukan pengolahan data menggunakan *Data Mining*. Contoh metode yang ada pada *data mining* adalah *Clustering*, metode ini juga merupakan metode yang valid diimplementasikan yang berguna untuk memecahkan berbagai macam masalah kompleks pada ilmu komputer dan statistik.[1]

Clustering merupakan satu dari beberapa metode data mining yang memiliki tujuan dalam proses pengelompokan sebuah informasi data berdasarkan kemiripan karakteristik data tersebut dengan kelompok data lainnya [2] Salah satu algoritma clustering yang dapat digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik adalah *k-means clustering*. Algoritma *K-Means* memiliki algoritma yang secara berulang mengelompokkan dataset, membagi dataset menjadi beberapa cluster yang ditentukan pada awal proses dan dapat dikustomisasi dengan relatif cepat dan mudah [3]. Penelitian tentang algoritma *K-Means* telah dilakukan oleh banyak peneliti untuk mendukung kebutuhan berbagai disiplin ilmu. Salah satunya dalam bidang penjualan atau bisnis. Penelitian ini menggunakan algoritma *K-Means*. Hal ini karena implementasinya sangat sederhana, waktu yang dibutuhkan untuk menyelesaikan prosesnya tidak lama, dan mudah untuk disesuaikan dan dipelajari. Oleh karena itu, survei pengelompokan data penjualan Toko Gizi ini membantu dalam proses manajemen inventaris, mengidentifikasi peluang dari penjualan sebelumnya, menyesuaikan omset penjualan dan item masuk, mempertahankan inventaris produk yang tersedia, dan merespons dengan cepat, tetapi Toko Gizi cukup cepat.

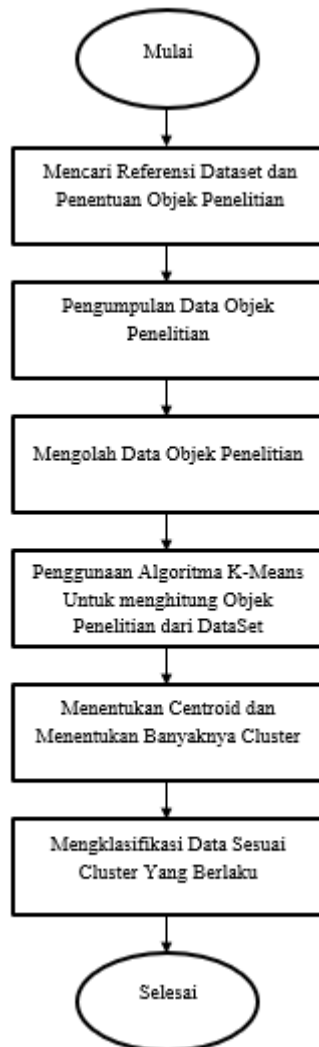
2. METODOLOGI

Penelitian ini menggunakan beberapa tahapan implementasi, mulai dari proses Mencari Referensi Dataset dan Penentuan Objek Penelitian, Pengumpulan Data Objek Penelitian, Mengolah Data Objek Penelitian, Penggunaan Algoritma *K-Means* Untuk menghitung Objek Penelitian dari *DataSet*, Menentukan Centroid dan Menentukan Banyaknya Cluster, Mengklasifikasi Data Sesuai Cluster Yang Berlaku. Alur dapat dilihat pada gambar 1.

Pengertian dari alur metodologi dan algoritma *K-Means* yang digunakan yaitu:

- a. Pengumpulan data objek penjualan adalah langkah yang harus dilakukan guna untuk pengumpulan data terhadap barang dari hasil penjualan toko Gizi dari Bulan April-Mei 2022.
- b. Menghitung data transaksi barang adalah langkah penghitungan riwayat transaksi yang sudah terjadi yang terlihat dari banyaknya Transaksi dan Penjualan suatu Barang.
- c. Penentuan seberapa banyaknya cluster (*k*) dari beberapa data yang sudah terkumpul lalu menentukan nilai centroid yang dilakukan secara acak.
- d. Penghitungan dari algoritma *K-Means* dari data transaksi atau penjualan dimana penghitungan yang dilakukan adalah menggunakan jarak Euclidean Distance terhadap data yang sudah ada dengan nilai Centroid yang sudah tersedia. Untuk menghitung jarak yaitu menggunakan rumus:

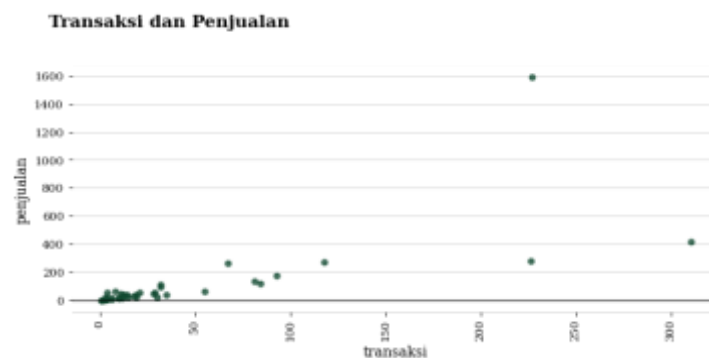
- $$d(x_i, u_j) = \sqrt{\sum (x_i, u_j)^2}$$
- e. Mengelompokkan objek. Anggota Cluster ditentukan oleh objek apa yang akan dijadikan sebagai patokan Clustering. Pada penelitian ini ditentukan 3 Cluster.



Gambar 1. Alur Metodologi

3. HASIL DAN PEMBAHASAN

3.1 Memperlihatkan Data Transaksi dan Penjualan Dalam Bentuk Grafik



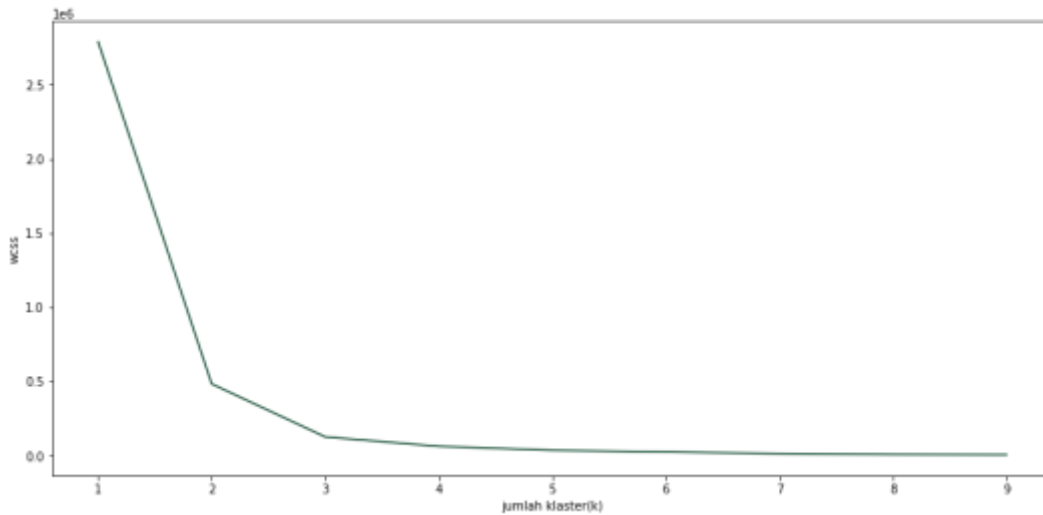
Gambar 2. Grafik Transaksi dan Penjualan

Sebelum menentukan cluster, kita harus mengetahui terlebih dahulu kumpulan data ke dalam bentuk Grafik agar mampu mendeskripsikan bagaimana data harus diproses.

3.2 Penentuan Jumlah Cluster yang Optimal

Dalam penentuan jumlah cluster yang paling optimal, digunakan google collabs dan diterapkan metode elbow agar mampu melihat statistik data, yang selanjutnya akan dapat diproses seberapa banyak cluster yang dapat diterapkan.

Kluster 2D: Transaksi dan Penjualan
Metode Elbow



Gambar 3. Grafik Metode Elbow

3.3 Menentukan Silhouette

Jumlah Kluster dan Silhouettenya, untuk yang terbesar adalah 2 Kluster, akan tetapi kami memutuskan untuk menggunakan 3 Kluster karena Banyaknya Data agar dapat memenuhi kesesuaian Objektif, yakni Sangat Laris, Cukup Laris, Kurang Laris.

```
Jumlah Kluster = 2 Nilai rata-rata silhouette = 0.9080493120923051  
Jumlah Kluster = 3 Nilai rata-rata silhouette = 0.7574199965373875  
Jumlah Kluster = 4 Nilai rata-rata silhouette = 0.7013162332112347  
Jumlah Kluster = 5 Nilai rata-rata silhouette = 0.6358526426306267
```

Gambar 4. Silhouette Kluster

3.4 Penerapan Teknik Cluster K-Means

A. Input Data

Penelitian Kali ini dimasukkan 41 Jenis Data Barang yang mempunyai masing-masing atribut seperti transaksi, penjualan, dan rata-rata diambil dari beberapa periode yang sudah ditentukan. Periode dari penelitian ini menggunakan periode penjualan yang terjadi pada Bulan Maret-Mei 2022

Tabel 1. Data Barang beserta Informasi Transaksi, Penjualan dan Rata-rata

No	Nama	Transaksi	Penjualan	Rata-Rata
1	DIABETASOL JUMBO 1 KG / 1000 g	30	25	27,5

No	Nama	Transaksi	Penjualan	Rata-Rata
2	DIABETASOL SWEETENER 100 Sachet	2	4	3
3	DIANERAL 180 G	6	18	12
4	ENTRAKID VANILA / COKELAT 185 G (Untuk anak Usia 1-5 Tahun)	32	109	70,5
5	ENTRAMIX 174g Vanila/Coklat	21	58	39,5
6	ENTRASOL GOLD 600 g (isi 2 x 300 g)	1	2	1,5
7	ENTRASOL PLATINUM 400 g	15	21	18
8	GARAM LOSOSA 250 Gr	14	41	27,5
9	GOLDSURE by ENSURE 900g	4	7	5,5
...
41	ZINC OPTIMA 60 ML	19	39	29

B. Penentuan Nilai Centroid dan Penentuan Jumlah Cluster

Tahap Kedua adalah penentuan nilai awal *Centroid* dan penentuan jumlah *cluster*, dimana hasil dari perumusan secara otomatis muncullah hasil data berikut ini.

Tabel 2. Centroid

Centroid	Jumlah Transaksi	Rata-Rata	Total Penjualan
1	18	26,75	35,5
2	162,8	222,3	281,8
3	227	909	1591

C. Penentuan Kelompok serta Penghitungan Kembali Nilai Centroid

Di tahap ketiga penghitungan jarak ada pada titik pusat cluster dan jarak pada titik tiap objek. Berikutnya adalah pengelompokkan yang merupakan awal dari perbandingan serta pemilihan melalui jarak yang paling dekat dengan pusat dari cluster, jarak inilah yang merupakan petunjuk dari data yang ada termasuk dalam kelompok yang sama dengan pusat Cluster yang paling dekat dengan perbandingan dari hasil cluster yang diambil dari yang paling kecil.

D. Hasil Pengelompokkan

Sesudah mendapatkan hasil proses metode perhitungan *K-Means Clustering* maka diperoleh hasil dari pengelompokkan per *Cluster* yakni:

Tabel 3. Tabel Hasil Pengelompokkan

No	Nama	Transaksi	Penjualan	Cluster
1	GULA TROPICANA SLIM CLASSIC 160'S	226	280	2

No	Nama	Transaksi	Penjualan	Cluster
2	GULA TROPICANA SLIM DIABTX Diabet isi 150 SACHET	310	418	2
3	NUTRINIDRINK 400g - VANILA (nutricia)	118	271	2
4	PEPTISOL 189g	93	175	2
5	SANTAN SUN KARA 200ML	67	265	2
6	NUTRINIDRINK 200 ml Cair	227	1591	1
7	DIABETASOL JUMBO 1 KG / 1000 g	30	25	0
8	DIABETASOL SWEETENER 100 Sachet	2	4	0
9	DIANERAL 180 G	6	18	0
10	ENTRAKID VANILA / COKELAT 185 G (Untuk anak Usia 1-5 Tahun)	32	109	0
...
41	ZINC OPTIMA 60 ML	19	39	0

Dengan keterangan tiap cluster dapat dilihat pada tabel 4.

Tabel 4. Cluster

Cluster	Kategori	Penjualan Minimal
Cluster 1	Kurang Diminati	35,5
Cluster 2	Cukup Diminati	281,8
Cluster 3	Sangat Diminati	1591

4. KESIMPULAN DAN SARAN

Dari hasil yang sudah didapat dengan mengimplementasikan algoritma K-Means terhadap 41 jenis data penjualan dan transaksi dari bulan April hingga bulan Mei 2022 memunculkan 3 cluster yang bisa dibilang saat ini sangat diminati, cukup diminati dan kurang diminati. Dari ketiga cluster yang disebutkan dan terletak pada titik centroid yaitu C1 (18, 35.5, 26.75), C2 (162.8, 281.8, 222.3) dan C3 (227, 1591, 909). Dan mendapatkan kesimpulan data yang ada di cluster pertama atau Kurang Diminati mempunyai 5 barang, cluster kedua atau Cukup diminati mempunyai 5 barang dan cluster ketiga atau sangat diminati mempunyai 1 barang.

5. DAFTAR RUJUKAN

- [1] H. Syahputra, L. Mayola and D. Guswand, "Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means," Jurnal KomtekInfo, vol. Vol 9 No 1, pp. 29-33, 2022.
- [2] M. R. Praand tama R. Supardi, "DECISION SUPPORT SYSTEM METODE K-MEANS CLUSTERING DALAM MEMREDIKSI TINGKAT PENJUALAN PADA STORE RYAN MART," Jurnal Media Infotama, vol. Vol.18 No.1, pp. 81-86, 2022.
- [3] A. F. Sallaby, R. T. Alinse, V. N. Sari and T. Ramadani, "PENGELOMPOKAN BARANG MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN HASIL PENJUALAN DI TOKO WIDYA BENGKULU," Jurnal Media Infotama, vol. Vol.18 No.1, pp. 99-104, 2022.
- [4] H. Prastiwi, J. Pricilia and E. Raswir, "Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering," Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM), Vols. Volume 1, Nomor 2, pp. 141-148, 2022.

-
- [5] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah Asuransi," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, Vols. Volume 5, Nomor 3, pp. 1187-1194, 2021.
- [6] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means (Clustering Sales Data at Outdoor Equipment Stores Using K-Means Method)," *JUITA: Jurnal Informatika*, Vols. Volume 7, Nomor 2, pp. 109-114, 2019.
- [7] M. A. K. Lutfi and A. Nilogiri, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN MINAT KONSUMEN PADA PRODUK ONLINE SHOP," Publisher Muhammadiyah University Press Country, Jember, 2013.
- [8] N. Mirantika, A. T. Ain and F. D. Agnia, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat," *JURNAL NUANSA INFORMATIKA*, vol. Volume 15 Nomor 2, pp. 92-98, 2021.
- [9] Y. D. Putra, M. Sudarma and I. B. A. Swamardika, "Clustering History Data Penjualan Menggunakan Algoritma K-Means," *Majalah Ilmiah Teknologi Elektro*, Vols. Vol. 20, No. 2, pp. 195-202, 2021.
- [10] A. H. Al-atho'illah, E. D. Wahyuni and A. A. Arifiyanti, "MODEL KLASIFIKASI TEKS PRODUK TERLARANG MENGGUNAKAN ALGORITMA CAMPURAN (HYBRID) SVM DAN LEKSIKON," *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, Vols. Vol. 1, No. 2, pp. 563-572, 2020.