

Penerapan Algoritma *K-Means* untuk Mengetahui Pola Persediaan Barang pada Toko Raja Bekasi

Intan Safira¹, Ratna Salkiawati^{1,*}, Wowon Priatna¹

¹ Informatika; Universitas Bhayangkara Jakarta Raya; Jl. Perjuangan No.81, Marga Mulya, Kec. Bekasi Utara, Kota Bekasi, Jawa Barat 17143, Telp. 021-88966882; e-mail: intan.safira18@mhs.ubharajaya.ac.id, ratna_tind@dsn.ubharajaya.ac.id, wowon.priatna@dsn.ubharajaya.ac.id

* Korespondensi: e-mail: ratna_tind@dsn.ubharajaya.ac.id

Diterima: 23 Juni 2022 ; Review: 30 Juli 2022; Disetujui: 31 Juli 2022; Diterbitkan: 31 Juli 2022

Abstract

This study aims to determine how much the results of grouping goods affect the needs of consumers. Excess inventory will greatly fill the warehouse and be inefficient because of the expiration date on food products, beverages, etc. Currently Toko Raja still manages goods manually so it is not time efficient. To solve this problem, a technique is needed, namely data mining. The data mining technique that will be used in this research is the K-Means Clustering method. K-Means is one of the most popular algorithms because it is easy and simple to implement. However, the results of the clustering of K-Means are very dependent on the selection of the initial cluster center point. Calculation of accuracy in this study using the test results of the K-Means clustering method using the Davies-Bouldin Index (DBI) is 1.856 where the DBI value close to zero cluster is good enough. From the resulting accuracy, it can be concluded that the K-Means Clustering method can support the system well.

Keywords: *Clustering, Data Mining, DBI, Store, Suply of Goods*

Abstrak

Penelitian bertujuan untuk mengetahui seberapa besar hasil dari pengelompokan barang berpengaruh terhadap kebutuhan dari konsumen. Kelebihan persediaan barang akan sangat memenuhi gudang dan tidak efisien karena adanya tanggal *expired* pada produk makanan, minuman dan lain-lain. Saat ini Toko Raja masih mengelola barang secara manual sehingga tidak efisiennya waktu. Untuk menyelesaikan masalah tersebut maka dibutuhkan suatu teknik yaitu data *mining*. Teknik data *mining* yang akan digunakan pada penelitian ini adalah metode *K-Means Clustering*. *K-Means* merupakan salah satu algoritma yang paling populer karena mudah dan sederhana ketika diimplementasikan. Namun hasil *clustering* dari *K-Means* sangat bergantung terhadap pemilihan titik pusat *cluster* awalnya. Perhitungan akurasi dalam penelitian ini menggunakan Hasil pengujian metode *clustering K-Means* menggunakan *Davies-Bouldin Index* (DBI) adalah sebesar 1,856 dimana nilai DBI mendekati nol *cluster* sudah cukup baik. Dari akurasi yang dihasilkan dapat disimpulkan bahwa metode *Clustering K-Means* dapat mendukung sistem dengan baik.

Kata Kunci: *Clustering, Data Mining, DBI, Persediaan Barang, Toko*

1. Pendahuluan

Toko Raja merupakan toko yang menjual berbagai macam sembako, diantaranya makanan pokok, peralatan mandi, keperluan dapur, obat-obatan dan lain-lain. Saat ini Toko

Raja masih melakukan pemenuhan persediaan barang atau produk dan melakukan pencatatan transaksi secara manual sehingga sering terjadi kesalahan dan pencatatan data-data yang ada dan juga kurangnya efisiensi waktu yang di perlukan. Jumlah permintaan dari konsumen yang fluktuatif mengakibatkan persediaan yang harus disiapkan menjadi tidak stabil. Selain itu persediaan barang yang beragam dan banyak jenisnya menjadikan manajemen persediaan yang dilakukan menjadi tidak akurat, biaya penyimpanan yang tinggi, waktu produksi yang lama, tidak efektif dan seringkali mengecewakan konsumen karena kekosongan suatu produk tertentu. Disamping itu Toko Raja tidak dapat mengelompokkan produk yang laris dan yang tidak laris terjual. Sehingga kesulitan yang di alami adalah kurangnya persediaan produk yang laku karena penjualan tinggi, dan menumpuknya produk yang tidak laku di gudang karena penjualannya rendah.

Permasalahan yang terjadi pada Toko Raja tersebut di sebabkan karena mengalami kesulitan dalam menentukan persediaan minimum tiap barang yang harus di penuhi berdasarkan minat konsumen. Untuk dapat mengatasi permasalahan yang terjadi, maka Toko Raja membutuhkan suatu metode dan sistem perancangan persediaan barang yang lebih baik sehingga dapat menentukan produk mana yang harus di produksi secara banyak, sedang ataupun sedikit agar tidak lagi mengalami kekurangan atau bahkan kelebihan dalam pemenuhan persediaan produk tertentu.

Untuk menyelesaikan permasalahan dengan pemanfaatan algoritma *K-Means Clustering*. Algoritma *K-Means* adalah algoritma *clustering* yang paling sederhana dibanding algoritma *clustering* yang lain. Algoritma ini mempunyai kelebihan mudah diterapkan dan dijalankan, relatif cepat, mudah untuk diadaptasi, dan paling banyak dipraktikkan dalam tugas data mining.

Dalam penelitian ini dirumuskan suatu masalah yaitu bagaimana menggunakan algoritma *K-Means* dapat membantu *clustering* produk untuk penentuan prediksi persediaan barang? Dengan tujuan membantu *clustering* produk untuk penentuan produksi persediaan barang, dan mengelompokkan barang yang ada di Toko Raja.

State of The Art

Penelitian Ahmad Asif Qolbi di Universitas Negeri Semarang pada tahun 2016 dengan judul "Penerapan Metode *Clustering K-Means* Terhadap Dosen Berdasarkan Publikasi Jurnal Nasional dan Internasional" memiliki kelebihan Sistem untuk menerapkan metode *clustering K-Means* dalam menentukan pengelompokkan dosen terkait hasil publikasi jurnal nasional dan jurnal internasional dapat diciptakan dengan memanfaatkan *software* Matlab. Dan kekurangan Penelitian ini diharapkan adanya *variable* input yang lebih rinci untuk digunakan sebagai data yang diimplementasikan kedalam sistem. Di mana terdapat perhitungan untuk mendapatkan nilai bobot antara jurnal nasional dan jurnal internasional (Qolbi, 2016).

Penelitian Bayu Resi Indrawan di Universitas Islam Negeri Sunan Kalijaga pada tahun 2018 dengan judul "Penerapan Algoritma *K-Means* Untuk Menentukan Strategi Promosi

Universitas Islam Negeri Sunan Kalijaga Yogyakarta” memiliki kelebihan Penelitian ini menerapkan metode algoritma *K-Means clustering* menggunakan dua aplikasi yaitu *Microsoft excel/2010* dan sistem aplikasi *K-Means clustering* yang dirancang dan dibuat oleh peneliti. Dan memiliki kekurangan Kurangnya kecepatan proses dalam sistem *K-Means clustering* dan kurangnya sumber data dalam penelitian ini dari hanya mencakup data pada prodi Teknik Informatika menjadi data Fakultas ataupun mencakup semua data di Universitas itu sendiri (Indrawan, 2018).

Penelitian Mikael Aditya Wahyu Krisna Murti di Universitas Sanata Dharma Yogyakarta pada tahun 2017 dengan judul “Penerapan Metode *K-Means Clustering* Untuk Mengelompokkan Potensi Produksi Buah-buahan di Provinsi Daerah Istimewa Yogyakarta” memiliki kelebihan Dari penelitian ini algoritma *K-Means* dapat melakukan pengelompokkan dokumen dalam jumlah yang banyak akan tetapi belum efisien dalam mengelompokkan dokumen secara tepat. Dan memiliki kekurangan Hasil *clustering* yang terbentuk tidak dapat dikembangkan menjadi basis pengetahuan untuk sistem pendukung keputusan maupun sistem rekomendasi daerah yang cocok untuk ditanami buah oleh dinas pertanian (Murti, 2017)

2. Metode Penelitian

Data mining adalah rangkaian kegiatan untuk menemukan pola yang menarik dari data dalam jumlah besar, kemudian data-data tersebut dapat disimpan dalam *database*, data *warehouse* atau penyimpanan informasi (Prilianti & Wijaya, 2014). Data mining bagian internal dari *knowledge discovery in database* (KDD). Sebuah langkah dalam proses mencari pola-pola yang terdapat dalam setiap informasi. *Clustering* merupakan salah satu bagian dari teknik data mining yaitu sekumpulan objek yang mempunyai kesamaan diantara anggotanya dan memiliki ketidaksamaan dengan objek lain pada *cluster* lainnya, dengan kata lain sebuah *cluster* adalah sekumpulan objek yang digabung bersama karena persamaan atau kedekatannya (Safriaty, 2017).

2.1. K-Means

Algoritma *K-Means* merupakan algoritma pengelompokkan iterative yang melakukan partisi set data kedalam sejumlah *cluster* yang sudah ditetapkan di awal. Algoritma *K-Means* sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, *K-Means* menjadi satu-satunya algoritma yang paling penting dalam bidang data *mining* (Fahmi, 2013).

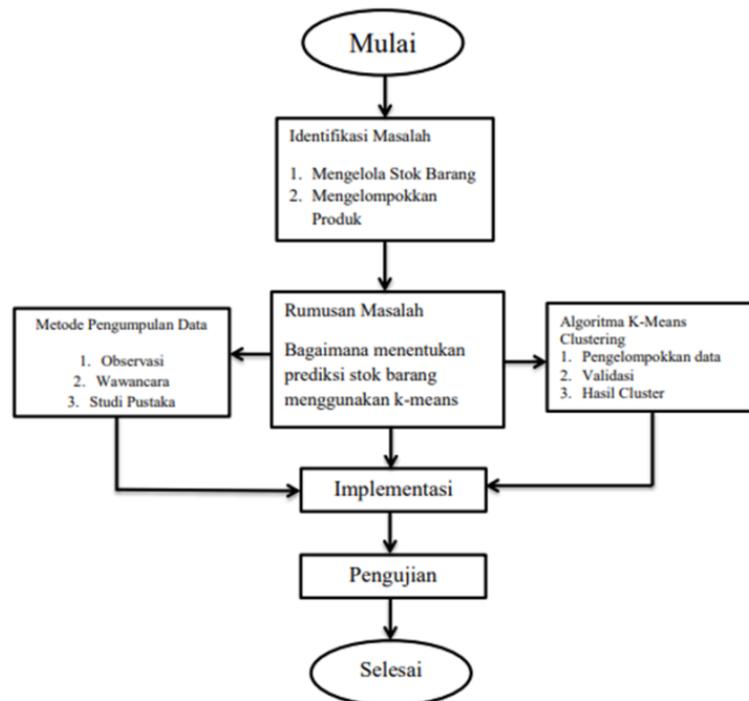
K-Means adalah metode *clustering* berbasis jarak yang membagi data kedalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut *numeric*. Algoritma *K-Means* termasuk *partitioning clustering* yang memisahkan data ke daerah bagian yang terpisah. Dalam algoritma *K-Means*, setiap data harus masuk ke-*cluster* tertentu dan bisa dimungkinkan bagi setiap data yang termasuk *cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke-*cluster* lainnya (Nasari & Sianturi, 2016).

2.2. Davies-Bouldin Index (DBI)

David L. Davies dan Donald W. Bouldin memperkenalkan sebuah metode yang diberi nama dengan nama mereka berdua, yaitu *Davies-Bouldin Index* (DBI) yang digunakan untuk mengevaluasi cluster (Nawrin et al., 2017). Evaluasi menggunakan *Davies-Bouldin Index* (DBI) memiliki skema evaluasi internal *cluster* (Bates & Kalita, 2016). *Davies-Bouldin Index* merupakan metode yang di gunakan untuk mengukur validitas *cluster* pada suatu metode pengelompokkan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari *cluster* yang di ikuti. Sedangkan separasi didasarkan pada jarak antar titik pusat *cluster* terhadap *clusternya*. Pengukuran dengan *Davies-Bouldin Index* ini memaksimalkan jarak inter-*cluster* antara *cluster* C_i dan C_j dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah *cluster*. Jika jarak inter-cluster maksimal, berarti kesamaan karakteristik antar masing-masing *cluster* sedikit sehingga perbedaan antar-*cluster* terlihat jelas. Jika jarak *intra-cluster* minimal berarti masing-masing objek dalam *cluster* tersebut memiliki tingkat kesamaan karakteristik yang tinggi (Wani & Riyaz, 2017).

2.3. Desain Penelitian

Dalam melakukan suatu penelitian di butuhkan suatu prosedur penelitian agar penelitian tersebut dapat berjalan dengan baik. Prosedur penelitian pada Gambar 1:



Sumber: Hasil Penelitian (2022)

Gambar 1. Prosedur Penelitian

2.4. Metode Pengumpulan Data

- a. Wawancara, adalah metode pengumpulan data melalui kegiatan tanya jawab langsung kepada pihak pemilik Toko Raja.

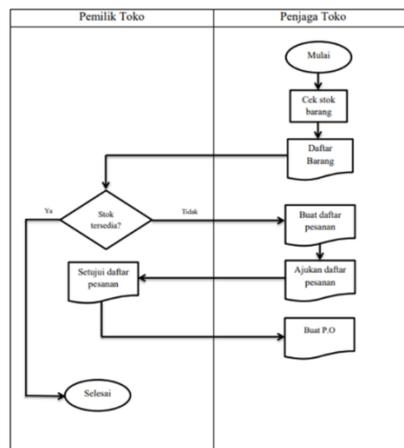
- b. Dokumentasi, adalah metode pengumpulan data dengan mengumpulkan catatan-catatan atau dokumen-dokumen. Data yang telah didapatkan dari metode wawancara, dikumpulkan menjadi satu sehingga menjadi dokumen yang siap digunakan untuk kepentingan penelitian.
- c. Studi pustaka, adalah mempelajari karya ilmiah, buku ilmiah, dan sumber ilmiah lainnya yang sesuai dan memiliki hubungan dengan masalah yang di teliti. Referensi ilmiah yang digunakan adalah sumber-sumber yang terdapat dalam daftar kepustakaan.

2.5. Metode Analisis

Analisis sistem merupakan tahap kritis dan penting karena kesalahan ditahap ini menyebabkan kesalahan lain pada tahap selanjutnya. Analisa sistem dapat di definisikan sebagai penguraian dari suatu sistem informasi yang utuh kedalam bagian komponen-komponennya dengan maksud mengidentifikasi dan mengevaluasi permasalahan-permasalahan, hambatan-hambatan, yang terjadi dan kebutuhan-kebutuhan yang diharapkan sehingga dapat diusulkan perbaikan-perbaikan. Adapun analisis yang akan di bahas ialah analisis sistem yang berjalan, analisis kelemahan sistem yang berjalan.

a. Analisis Sistem Berjalan

Analisis sistem yang sedang berjalan untuk mengetahui tahapan-tahapan proses yang dikerjakan. Berikut Gambar 2 *flowmap* sistem berjalan:



Sumber: Hasil Penelitian (2022)

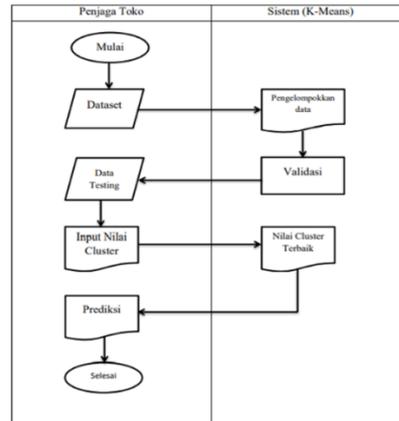
Gambar 2. Analisa Sistem Berjalan

b. Analisis Permasalahan

Setelah melihat *flowmap* sistem yang berjalan penulis mendapatkan beberapa kekurangan diantaranya: Penjaga toko harus mengecek persediaan barang secara manual, penjaga toko harus memberikan daftar barang kepada pemilik toko.

c. Analisis Sistem yang Di usulkan

Dari analisis permasalahan dibutuhkan sistem informasi yang dapat membantu memudahkan dalam mengelola pembelian agar lebih efektif kedepannya. Membuat sebuah sistem untuk membantu mengelompokkan penyediaan barang dari yang paling diminati hingga jarang diminati dan sistem dapat memberikan informasi hasil prediksi dengan jelas dan akurat. Gambar 3 merupakan *flowmap* proses sistem usulan:



Sumber: Hasil Penelitian (2022)

Gambar 3. Sistem Usulan

3. Hasil dan Pembahasan

3.1. Implementasi *K-Means*

Dilakukan proses utama yaitu segmentasi atau pengelompokan data penjualan barang yang di akses dari database, menggunakan metode *clustering K-Means*. Diagram *flowchart* dari algoritma *K-Means* dengan parameter input adalah jumlah set data sebanyak n data dan jumlah inialisasi *centroid* $K=3$ sesuai dengan penelitian. Dari banyak data penjualan yang diperoleh, diambil 10 jenis barang untuk dijadikan sampel untuk penerapan algoritma *K-Means*. Percobaan dilakukan dengan menggunakan parameter jumlah cluster = 3 dan jumlah data = 10.

Iterasi ke-1

a. Penentuan pusat awal *cluster*

Pusat awal *cluster* atau *centroid* didapatkan secara random, untuk penentuan awal *cluster* adalah :

Pusat *cluster* 1 = (300,900)

Pusat *cluster* 2 = (250,750)

Pusat *cluster* 3 = (150, 450)

b. Perhitungan jarak pusat cluster

Untuk mengukur jarak antara data dengan pusat *cluster* digunakan *Euclidian Distance*, kemudian akan di dapatkan matrik jarak perhitungan *Euclidian Distance* rumus 1:

$$d = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$$\sqrt{(X_i - X_{avg})^2 + (Y_i - Y_{avg})^2}$$

Dari 10 data yang dijadikan sampel telah dipilih pusat awal *cluster* yaitu **C1(300,900)**, **C2(250,750)**, **C3(150,450)**. Lalu dilakukan perhitungan jarak dari sisa sampel data dengan pusat cluster yang dimisalkan dengan M (a,b) dimana a merupakan penjualan, dan b merupakan persediaan barang dalam 3 tahun.

Tabel 1. Perhitungan Jarak

No	Nama Barang	Stok	Penjualan
1	M1	200	600
2	M2	150	450
3	M3	300	900
4	M4	90	270
5	M5	300	900
6	M6	90	270
7	M7	90	270
8	M8	250	750
9	M9	150	450
10	M10	80	240

Sumber: Hasil Penelitian (2022)

Hitung *Euclidian Distance* dari semua data ketiap titik pusat pertama:

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(200 - 300)^2 + (600 - 900)^2} = 316,227$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(150 - 300)^2 + (450 - 900)^2} = 474,341$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(300 - 300)^2 + (900 - 900)^2} = 0$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(90 - 300)^2 + (270 - 900)^2} = 664,078$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(300 - 300)^2 + (900 - 900)^2} = 0$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(90 - 300)^2 + (270 - 900)^2} = 664,078$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(90 - 300)^2 + (270 - 900)^2} = 664,078$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(250 - 300)^2 + (750 - 900)^2} = 158,113$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(150 - 300)^2 + (450 - 900)^2} = 474,341$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(80 - 300)^2 + (240 - 900)^2} = 695,701$$

Hitung *Euclidian Distance* dari semua data ketiap titik pusat kedua:

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(200 - 250)^2 + (600 - 750)^2} = 158,113$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(150 - 250)^2 + (450 - 750)^2} = 316,227$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(300 - 250)^2 + (900 - 750)^2} = 158,113$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(90 - 250)^2 + (270 - 750)^2} = 505,964$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(300 - 250)^2 + (900 - 750)^2} = 158,113$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(90 - 250)^2 + (270 - 750)^2} = 505,964$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(90 - 250)^2 + (270 - 750)^2} = 505,964$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(250 - 250)^2 + (750 - 750)^2} = 0$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(150 - 250)^2 + (450 - 750)^2} = 316,227$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(80 - 250)^2 + (240 - 750)^2} = 537,587$$

Hitung *Euclidian Distance* dari semua data ke setiap titik pusat ketiga:

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(200 - 150)^2 + (600 - 450)^2} = 158,113$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(150 - 150)^2 + (450 - 450)^2} = 0$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(300 - 150)^2 + (900 - 450)^2} = 474,341$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(90 - 150)^2 + (270 - 450)^2} = 189,736$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(300 - 150)^2 + (900 - 450)^2} = 474,341$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(90 - 150)^2 + (270 - 450)^2} = 189,736$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(90 - 150)^2 + (270 - 450)^2} = 189,736$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(250 - 150)^2 + (750 - 450)^2} = 316,227$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(150 - 150)^2 + (450 - 450)^2} = 0$$

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2} = \sqrt{(80 - 150)^2 + (240 - 450)^2} = 221,359$$

Tabel 2. *Euclidian Distance*

Nama Barang	C1	C2	C3
M1	316,227	158,113	158,113
M2	474,341	316,227	0
M3	0	158,113	474,341
M4	664,078	505,964	189,736

Nama Barang	C1	C2	C3
M5	0	158,113	474,341
M6	664,078	505,964	189,736
M7	664,078	505,964	189,736
M8	158,113	0	316,227
M9	474,341	316,227	0
M10	695,701	537,587	221,359

Sumber: Hasil Penelitian (2022)

Anggota C1 = {M3,M5}

Anggota C2 = {M1,M8}

Anggota C3 = {M2,M4,M6,M7,M9,M10}

Iterasi ke-2

a. Hitung titik pusat baru

Tentukan posisi *centroid* baru (C_k) dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_1 \quad (2)$$

Dimana n_k adalah jumlah dokumen dalam *cluster* k dan d_1 adalah dokumen dalam *cluster* k. Sehingga didapatkan titik pusat atau *centroid* yang baru yaitu:

C1 = (300,900)

C2 = (225,675)

C3 = (108.3, 325)

b. Perhitungan jarak pusat cluster

Hitung *Euclidian Distance* dari semua data ke titik pusat yang baru (C1,C2,C3) seperti yang telah dilakukan pada tahap 1 setelah hasil perhitungan telah didapatkan, sama dengan posisi iterasi pertama, maka proses dihentikan, namun jika tidak maka proses dilanjutkan ke iterasi ke3.

Anggota C1 = {M3,M5}

Anggota C2 = {M1,M8}

Anggota C3 = {M2,M4,M6,M7,M9,M10}

Karena hasil posisi cluster tidak berubah maka perhitungan dihentikan.

Kesimpulan: Nama barang pepsodent dan minyak bimoli merupakan anggota C1 dan termasuk barang dengan tingkat penjualan yang tinggi juga memberikan keuntungan yang tinggi, sedangkan indomie dan sarden maya termasuk anggota C2 dimana dikategorikan sebagai barang dengan tingkat keuntungan serta penjualan yang sedang, dan coca-cola, panadol, teh pucuk, freshcare, royco, dan pantene merupakan anggota C3 dan dikategorikan sebagai barang dengan tingkat penjualan dan keuntungan yang rendah.

3.2. Evaluasi Clustering

Evaluasi *clustering* dilakukan dengan tujuan untuk mengetahui seberapa baik kualitas dari hasil *clustering*. Pada penelitian ini, evaluasi hasil *clustering* yang digunakan adalah Davies-Bouldin Index. Untuk mendapatkan nilai Davies-Bouldin Index, terlebih dahulu dihitung nilai Sum of Square Within-cluster (SSW), Sum of Square Between-cluster (SSB) dan Ratio. Untuk memulai perhitungan manual Davies-Bouldin Index, dibutuhkan data yang sudah *diclustering*. Ilustrasi data sample yang telah selesai *diclustering* adalah seperti pada Tabel 3.

Tabel 3. Data Sample yang *Diclustering*

Cluster	Stok	Penjualan
1	300	900
2	225	675
3	108,3	325

Sumber: Hasil Penelitian (2022)

Langkah 1. Hitung SSW

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$$

Cluster 1

$$d(x_3, c_1) = \sqrt{(300 - 300)^2 + (900 - 900)^2} = 0$$

$$d(x_5, c_1) = \sqrt{(300 - 300)^2 + (900 - 900)^2} = 0$$

Sehingga SSW cluster 1 didapatkan:

$$SSW_1 = \frac{1}{2}(0 + 0) = 0$$

Cluster 2

$$d(x_1, c_1) = \sqrt{(200 - 300)^2 + (600 - 900)^2} = 316,227$$

$$d(x_8, c_1) = \sqrt{(250 - 300)^2 + (750 - 900)^2} = 158,113$$

Sehingga SSW cluster 2 didapatkan:

$$SSW_1 = \frac{1}{2}(316,227 + 158,113) = 237,17$$

Cluster 3

$$d(x_2, c_1) = \sqrt{(150 - 300)^2 + (450 - 900)^2} = 474,341$$

$$d(x_4, c_1) = \sqrt{(90 - 300)^2 + (270 - 900)^2} = 664,078$$

$$d(x_6, c_1) = \sqrt{(90 - 300)^2 + (270 - 900)^2} = 664,078$$

$$d(x_7, c_1) = \sqrt{(90 - 300)^2 + (270 - 900)^2} = 664,078$$

$$d(x_9, c_1) = \sqrt{(150 - 300)^2 + (450 - 900)^2} = 474,341$$

$$d(x_{10}, c_1) = \sqrt{(80 - 300)^2 + (240 - 900)^2} = 695,701$$

Sehingga SSW cluster 3 didapatkan:

$$SSW_1 = \frac{1}{6}(474,341 + 664,078 + 664,078 + 664,078 + 474,341 + 695,701) = 606,102$$

Langkah 2. Hitung SSB

$$SSB_{1,2} = d(c_i, c_j) = \sqrt{(x_{c1} - x_{c2})^2 + (y_{c1} - y_{c2})^2}$$

$$\sqrt{(300 - 225)^2 + (900 - 675)^2} = 237,170$$

$$SSB_{1,3} = d(c_i, c_j) = \sqrt{(x_{c1} - x_{c3})^2 + (y_{c1} - y_{c3})^2}$$

$$\sqrt{(300 - 108,3)^2 + (900 - 325)^2} = 606,11$$

$$SSB_{2,3} = d(c_i, c_j) = \sqrt{(x_{c2} - x_{c3})^2 + (y_{c2} - y_{c3})^2}$$

$$\sqrt{(225 - 108,3)^2 + (675 - 325)^2} = 368,94$$

Table 4. Hasil Matric SSB1

SBB	1	2	3
1	0	237,170	606,11
2	237,170	0	368,94
3	606,11	368,94	0

Sumber: Hasil Penelitian (2022)

Langkah 3. Hitung Rasio

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

$$R_{1,2} = \frac{0 + 237,17}{237,17} = 1$$

$$R_{1,3} = \frac{0 + 606,102}{606,11} = 0,999$$

$$R_{2,3} = \frac{237,17 + 606,102}{368,94} = 2,285$$

Table 5. Hasil Perhitungan Rasio 2

R	1	2	3	R max
1	0	1	0,999	1
2	1	0	2,285	2,285
3	0,999	2,285	0	2,285

Sumber: Hasil Penelitian (2022)

Langkah 4. Hitung DBI

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

$$DBI = \frac{1}{3} (1 + 2,285 + 2,285) = 1,856$$

Nilai DBI mendekati nol *cluster* sudah cukup baik.

4. Kesimpulan

Berdasarkan masalah bagaimana menggunakan algoritma *K-Means* dapat membantu *clustering* produk untuk penentuan prediksi persediaan barang? Hasil pengujian metode *clustering K-Means* menggunakan *Davies-Bouldin Index* (DBI) adalah sebesar 1,856 dimana nilai DBI mendekati nol *cluster* sudah cukup baik.

Daftar Pustaka

- Bates, A., & Kalita, J. (2016). Counting Clusters in Twitter Posts. *ACM International Conference Proceeding Series, 04-05-Marc*. <https://doi.org/10.1145/2905055.2905295>
- Fahmi, M. (2013). Penentuan Prioritas Rehabilitasi DAS Menggunakan Algoritma *K-Means Clustering*. *JAVA Journal of Electrical and Electronics Engineering*, 11(2), 14–20.
- Indrawan, B. R. (2018). *Penerapan Algoritma K-Means Untuk Menentukan Strategi Promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta*. Universitas Islam Negeri Sunan Kalijaga.
- Murti, M. A. W. K. (2017). *Penerapan Metode K-Means Clustering Untuk Mengelompokkan Potensi Produksi Buah-buahan di Provinsi Daerah Istimewa Yogyakarta*. Universitas Sanata Dharma Yogyakarta.
- Nasari, F., & Sianturi, C. J. M. (2016). Penerapan Algoritma *K-Means Clustering* Untuk Pengelompokan Penyebaran Diare di Kabupaten Langkat. *CogITo Smart Journal*, 2(2), 108–119. <https://doi.org/http://dx.doi.org/10.31154/cogito.v2i2.19.108-119>
- Nawrin, S., Rahatur, M., & Akhter, S. (2017). Exploring *K-Means* with Internal Validity Indexes for Data *Clustering* in Traffic Management System. (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, 8(3), 264–272. <https://doi.org/10.14569/ijacsa.2017.080337>
- Prianti, K. R., & Wijaya, H. (2014). Aplikasi Teks Mining Untuk Automasi Penentuan Tren Topik Skripsi Dengan Metode *K-Means Clustering*. *Jurnal Cybermatika*, 2(1).
- Qolbi, A. A. (2016). *Penerapan Metode Clustering K-Means Terhadap Dosen Berdasarkan Publikasi Jurnal Nasional dan Internasional*. Universitas Negeri Semarang.
- Safriaty, U. (2017). *Pembelajaran Mesin (Machine Learning) Clustering & Algoritma K-Means*. Universitas Hasannudin.
- Wani, M. ., & Riyaz, R. (2017). A Novel Point Density Based Validity Index For *Clustering* Gene Expression Datasets. *International Journal of Data Mining and Bioinformatics* 17, 17(1), 66–84.