

Pemanfaatan Data Mining untuk Mengelompokkan Data Penjualan Menggunakan Algoritma K-Medoids

Nurul Cahaya Kamila¹, Fitria², M. Hafid³

^{1,2,3}Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati, Tarakan, Kalimantan Utara
Email: ¹nurulck05@gmail.com, ²fitria@ppkia.ac.id, ³hafid@ppkia.ac.id

Abstrak

Penelitian ini bertujuan untuk membantu toko pernak-pernik dalam mencapai target penjualan bulanan melalui penerapan teknik data mining dengan metode klusterisasi K-Medoids. Penelitian dilakukan melalui enam tahapan utama, yaitu: (1) pengumpulan data, (2) pembersihan data (data cleaning), (3) implementasi data mining, (4) evaluasi hasil klusterisasi menggunakan nilai Davies-Bouldin Index (DBI), (5) penentuan jumlah kluster optimal (k-terbaik), dan (6) visualisasi hasil klusterisasi. Data yang digunakan terdiri dari tiga atribut yang dipilih dari enam atribut yang tersedia. Proses klusterisasi dengan metode K-Medoids menghasilkan kluster yang bervariasi karena pemilihan titik pusat (centroid) dilakukan secara acak. Berdasarkan evaluasi DBI, jumlah kluster optimal ditemukan pada k=3, yang memberikan hasil klusterisasi terbaik untuk mendukung strategi pemasaran toko.

Kata Kunci: Clustering, Data Mining, K-Medoids, Davies Bouldin Index (DBI).

Implementation of Data Mining for Sales Data Clustering Using K-Medoids Algorithm

Abstract

This study aims to assist a trinket shop in achieving its monthly sales targets by applying data mining techniques using the K-Medoids clustering method. The research was conducted in six main stages: (1) data collection, (2) data cleaning, (3) data mining implementation, (4) evaluation of clustering results using the Davies-Bouldin Index (DBI), (5) determination of the optimal number of clusters (best k), and (6) visualization of clustering results. The data used consists of three selected attributes out of six available attributes. The clustering process with the K-Medoids method produced varying clusters due to the random selection of centroids. Based on the DBI evaluation, the optimal number of clusters was found to be k=3, providing the best clustering results to support the shop's marketing strategies.

Keywords: Clustering, Data Mining, K-Medoids, Davies Bouldin Index (DBI).

I. PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi terus mendorong masyarakat untuk berinovasi di berbagai bidang guna mencapai efisiensi dalam waktu dan hasil yang optimal. Salah satu bidang yang terpengaruh secara signifikan adalah perdagangan, di mana aktivitas membeli produk dari produsen atau pihak tertentu untuk dijual kembali kepada konsumen menjadi bagian penting dari strategi untuk memperoleh keuntungan secara maksimal.

Ketatnya persaingan dalam dunia bisnis memaksa pelaku usaha untuk mengembangkan ide-ide baru dan strategi inovatif agar bisnis mereka tetap kompetitif. Salah satu sektor yang berkembang pesat adalah penjualan pernak-pernik rumah tangga, didukung oleh tren konten digital yang semakin populer, terutama dari kalangan ibu rumah tangga yang sering memamerkan dekorasi rumah atau dapur mereka melalui

media sosial. Salah satu toko pernak-pernik yang berlokasi di Tarakan menjual berbagai peralatan rumah tangga dan dekorasi, seperti pisau, mangkuk, toples, hiasan dinding, hingga perabotan unik lainnya. Dengan banyaknya jenis produk yang dijual dan meningkatnya permintaan konsumen, toko ini memerlukan pendekatan yang lebih terorganisir untuk mengelompokkan produk-produk tersebut agar dapat dikelola secara lebih efisien.

Data mining, sebagai sebuah disiplin ilmu, bertujuan untuk menemukan pola-pola tersembunyi, hubungan yang relevan, dan tren tertentu dalam kumpulan data yang besar. Proses ini mengintegrasikan berbagai teknik seperti statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi informasi yang bermanfaat dari database [1]. Salah satu metode data mining yang sering digunakan adalah clustering, yang bertujuan untuk mengelompokkan objek atau

data ke dalam beberapa kluster berdasarkan kesamaan atribut atau karakteristik tertentu [2].

Dalam kasus toko pernak-pernik ini, data penjualan dapat dimanfaatkan untuk menggali informasi penting seperti pola pembelian konsumen, perencanaan pengelolaan stok barang, dan peningkatan efisiensi operasional. Analisis data penjualan dengan menggunakan teknik data mining memungkinkan pengambilan keputusan yang lebih cerdas dan berbasis data. Penelitian ini menggunakan metode K-Medoids, sebuah teknik clustering partisional yang meminimalkan jarak antara data di dalam kluster dengan pusat kluster yang dipilih. Berbeda dengan metode lain seperti K-Means, K-Medoids lebih tahan terhadap keberadaan data pencilan (outliers) karena menggunakan titik-titik data aktual sebagai pusat kluster [3].

Metode K-Medoids dipilih karena kemampuannya dalam menghasilkan pengelompokan data yang lebih stabil dan akurat, terutama ketika berhadapan dengan data yang memiliki distribusi tidak merata [4]. Dengan menerapkan metode ini, diharapkan toko pernak-pernik dapat memperoleh wawasan baru terkait pengelompokan produk, sehingga dapat mendukung pengambilan keputusan strategis, seperti penentuan promosi, alokasi stok, dan perencanaan distribusi barang. Penelitian ini bertujuan untuk memberikan solusi yang praktis dan efektif dalam mengelola data penjualan dengan memanfaatkan teknologi data mining.

II. METODOLOGI PENELITIAN

A. Algoritma K-Medoids

K-Medoids, yang juga dikenal sebagai algoritma Partition Around Medoids (PAM), pertama kali diperkenalkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada tahun 1987. Algoritma ini termasuk dalam kategori partitioning clustering, yaitu teknik untuk mengelompokkan sejumlah objek ke dalam beberapa kluster berdasarkan kemiripan tertentu. Pada metode ini, medoid digunakan sebagai representasi dari setiap kluster. Medoid adalah objek nyata dalam kumpulan data yang paling mewakili kluster tersebut [5]. Berbeda dengan algoritma K-Means, yang menentukan pusat kluster berdasarkan nilai rata-rata (mean) dari data dalam kluster, K-Medoids menggunakan titik data yang sebenarnya sebagai pusat kluster. Pendekatan ini memberikan keunggulan signifikan dalam menghadapi data yang mengandung noise atau outlier—yaitu data yang secara signifikan berbeda dari distribusi mayoritas. Sementara K-Means cenderung sensitif terhadap keberadaan noise dan outlier karena perubahan pada nilai rata-rata dapat memengaruhi pusat kluster, K-Medoids lebih robust, menjadikannya pilihan yang lebih stabil untuk data dengan karakteristik tersebut.

Langkah-langkah dalam algoritma K-Medoids mencakup proses iteratif untuk memilih medoid, menghitung jarak antar objek, dan mengoptimalkan pengelompokan dengan mengganti medoid jika diperlukan. Secara umum, proses ini dimulai dengan pemilihan awal medoid secara acak dari data. Selanjutnya, setiap objek dihubungkan ke medoid yang paling dekat berdasarkan jarak tertentu, misalnya jarak Euclidean atau Manhattan. Setelah itu, algoritma mengevaluasi apakah ada perubahan medoid yang dapat mengurangi biaya total dalam kluster. Jika perubahan tersebut meningkatkan kualitas

kluster, medoid baru akan dipilih, dan proses diulang hingga tidak ada lagi perubahan yang signifikan atau kriteria penghentian tertentu terpenuhi [6].

Keunggulan utama dari K-Medoids terletak pada kemampuannya untuk menghasilkan kluster yang lebih representatif dan tahan terhadap anomali dalam data. Algoritma ini cocok digunakan dalam berbagai aplikasi, termasuk pengelompokan data penjualan, analisis pola pelanggan, hingga pengelompokan dalam data medis, di mana akurasi dan stabilitas hasil sangat penting. Dengan mengikuti langkah-langkah yang terstruktur, K-Medoids menawarkan solusi yang andal untuk analisis data kompleks yang memerlukan pendekatan berbasis kluster [7].

Beberapa tahapan yang dilakukan oleh penulis dalam penelitian ini :

- 1) Pengumpulan data dan data *Clean*
- 2) Implementasi Algoritma K-Medoids
 - a. Inisialisasi pusat cluster sebanyak
 - b. Menghitung jarak (Manhattan Distance).

$$d(x, y) = \sum_{i=1}^n |X_i - Y_i| \quad (1)$$

Keterangan :

$$d(x, y) = \text{Jarak}$$

$$X = \text{Koordinat 1}$$

$$Y = \text{Koordinat 2}$$
 - c. Memilih acak data untuk setiap cluster, kemudian dijadikan medoids baru.
 - d. Menghitung Total Distance.
- 3) Evaluasi jumlah cluster terbaik (Davies Bouldin Index)

Langkah penyelesaian evaluasi sebagai berikut :

- a. *Sum Of Square Within- Cluster (SSW)*
Mengetahui kohesi dalam sebuah cluster ke-i, persamaan sebagai berikut :

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - P_i)^2} \quad (2)$$

Keterangan:

(p, q) = Dua titik di ruang n Euclidean

q_i, P_i = Vector Euclidean

n = ruang - n

Menghitung nilai SSW sebagai berikut :

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, C_i) \quad (3)$$

Keterangan :

m_i = jumlah data dalam cluster ke-i

C_i = centroid cluster ke-i

d = jarak setiap data

- b. *Sum of square between cluster (SSB)*

Mengetahui separasi atau jarak antar cluster, persamaan sebagai berikut :

$$SSB_{ij} = d(X_i, X_j) \quad (4)$$

Keterangan:

$d(X_i, X_j)$ = Jarak antara data ke i dengan data ke j di cluster lain

c. *Ratio* (Rasio)

Mengetahui nilai perbandingan antara cluster ke-i dan cluster ke-j, persamaan sebagai berikut :

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (5)$$

Keterangan:

SSW_i = Sum Of Square Within- Cluster pada centroid i

$SSB_{i,j}$ = Sum of Square Between Cluster data ke i dengan j pada cluster yang berbeda

d. *Davies Bouldin Index* (DBI)

Mengevaluasi atau mempertimbangkan hasil algoritma clustering menggunakan persamaan berikut :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (6)$$

Keterangan:

k = jumlah cluster

Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik cluster yang diperoleh.

- 4) Penentuan hasil K-terbaik
- 5) Hasil clustering

III. HASIL DAN PEMBAHASAN

Toko Pernak Pernik ingin meningkatkan pengetahuan pelanggan dan efisiensi operasionalnya, Jumlah klaster yang digunakan sebanyak tiga atau $K=3$. Toko pernak pernik ingin menggunakan teknik clustering untuk mengelompokkan produk agar toko dapat memberikan rekomendasi produk, manajemen stok, strategi harga dan promosi produk. Penerapan perhitungan metode Algoritma k-medoids dilakukan pemilihan centroid secara random (acak). Berdasarkan tahapan penelitian penulis melakukan :

1) Pengumpulan Data

Sampel data pada penelitian ini diambil sebanyak 50 item dari data penjualan. Namun dalam penjelasan berikut ini hanya menampilkan 20 item data penjualan seperti tampak pada Tabel 1, data yang digunakan dari bulan Januari sampai dengan Agustus 2023.

Tabel 1. Data Penjualan

| No | Merk | Nama Barang | K1 | K2 | K3 |
|----|---------|-----------------|----|----|--------|
| 1 | Ikea | Storage box | 15 | 3 | 120000 |
| 2 | Informa | Stroge box 6,5L | 10 | 2 | 85000 |

Tabel 1. Data Penjualan (lanjutan)

| No | Merk | Nama Barang | K1 | K2 | K3 |
|----|----------|-------------------------|----|----|--------|
| 3 | Informa | Pedal dust bin rect | 7 | 3 | 210000 |
| 4 | Ikea | Rigga | 10 | 5 | 360000 |
| 5 | Ikea | Gladom | 15 | 3 | 425000 |
| 6 | Ikea | Harnavan Troli | 10 | 2 | 350000 |
| 7 | Informa | Boston Shoe Rack 4 Tier | 15 | 5 | 360000 |
| 8 | Krishome | Slimwaggon | 6 | 2 | 260000 |
| 9 | Ikea | Bekvam | 5 | 2 | 385000 |
| 10 | Ace | Wir Leaer Shelving Kit | 10 | 5 | 380000 |
| 11 | Ikea | Lersta | 12 | 6 | 390000 |
| 12 | Stora | Wall Hook | 20 | 2 | 270000 |
| 13 | Krisbo | bucket shelf 2 | 5 | 2 | 245000 |
| 14 | Krisbo | bucket shelf 4, tier | 5 | 3 | 450000 |
| 15 | Informa | Drawer Shelf | 5 | 1 | 450000 |
| 16 | Ikea | Mulig | 10 | 8 | 300000 |
| 17 | Ikea | Jall | 10 | 6 | 215000 |
| 18 | Ikea | Fejan | 10 | 1 | 250000 |
| 19 | Ikea | Marius | 20 | 15 | 160000 |
| 20 | Ikea | Dacey, Chair | 10 | 5 | 425000 |

2) Implementasi Algoritma

Berdasarkan data penjualan bulan Januari 2023 pada tabel 3.1 di atas dilakukan pengelompokkan dengan menerapkan metode k-medoids dimana dengan teknik pengukuran jarak menggunakan Manhattan Distance. Adapun tahapan penerapan k-medoids clustering pada penelitian ini sebagai berikut :

a. Menentukan nilai K (Jumlah Cluster) yang diinginkan, Pada penelitian ini nilai K sudah ditentukan yaitu sebanyak 3 ($K=3$).

b. Pilih secara acak medoid awal sebanyak K dari n data. Medoids awal pada penelitian ini diambil secara acak dari data penjualan sebanyak 3, diasumsikan sebagai berikut :

Medoid 1: (15,3,120000)

Medoid 2: (10,5,38000)

Medoid 3: (8,3,85000)

Data medoid awal dapat dilihat pada Tabel 2 berikut:

Tabel 2. Medoid Awal

| K1 | K2 | K3 |
|----|----|--------|
| 15 | 3 | 120000 |
| 10 | 5 | 38000 |
| 8 | 3 | 85000 |

c. Hitung jarak masing-masing obyek ke medoid (m) sementara, kemudian tandai jarak terdekat obyek ke medoid, lalu hitung total menggunakan pengukuran jarak Manhattan Distance berdasarkan persamaan (1) sebagai berikut :

$$D1: |15-15| + |3-3| + |120000-120000| = 0$$

$$D2: |15-10| + |3-5| + |120000-38000| = 260007$$

$$D3: |15-8| + |3-3| + |120000-85000| = 350007$$

Lakukan perhitungan jarak untuk semua obyek menggunakan cara yang sama. hasil keseluruhan perhitungan jarak dapat terlihat pada tabel 3 :

Tabel 3. Hasil Perhitungan Jarak Iterasi 1

| D1 | D2 | D3 | MD |
|--------|--------|--------|--------|
| 0 | 260007 | 35007 | 0 |
| 35006 | 295003 | 3 | 3 |
| 90008 | 170005 | 125001 | 90008 |
| 240007 | 20000 | 275004 | 20000 |
| 305000 | 45007 | 340007 | 45007 |
| 230006 | 30003 | 265003 | 30003 |
| 240002 | 20005 | 275009 | 20005 |
| 140010 | 120007 | 175003 | 120007 |
| 265011 | 5008 | 300004 | 5008 |
| 260007 | 0 | 295004 | 0 |
| 270006 | 10003 | 305007 | 10003 |
| 150006 | 110013 | 185013 | 110013 |
| 125011 | 135008 | 160004 | 125011 |
| 330010 | 70007 | 365003 | 70007 |
| 330012 | 70009 | 365005 | 70009 |
| 180010 | 80003 | 215007 | 80003 |
| 95008 | 165001 | 130005 | 95008 |
| 130007 | 130004 | 165004 | 130004 |
| 40017 | 220020 | 75024 | 40017 |
| 305007 | 45000 | 340004 | 45000 |

Selanjutnya menghitung jarak terdekat dengan menjumlahkan seluruh nilai minimum sebanyak 50 data sehingga diperoleh jarak terdekat atau total distance pada iterasi 1 adalah 2934320.

d. Lakukan Iterasi Medoid

Tahapan ini menentukan kembali medoid baru yang diambil acak, selanjutnya menghitung jarak menggunakan perhitungan manhattan distance sama seperti perhitungan pada tabel 3, namun dalam iterasi ini akan ditentukan centroid yang sama seperti tampak pada tabel 4 berikut :

Tabel 4. Hasil Perhitungan Jarak Iterasi 2

| D1 | D2 | D3 | MD | Cluster |
|--------|--------|--------|--------|---------|
| 350006 | 90008 | 35007 | 35006 | D1 |
| 0 | 125004 | 3 | 0 | D1 |
| 125004 | 0 | 125001 | 0 | D2 |
| 275003 | 150005 | 275004 | 150005 | D2 |
| 340006 | 215008 | 340007 | 215008 | D2 |
| 265000 | 140004 | 265003 | 140004 | D2 |
| 275008 | 150010 | 275009 | 150010 | D2 |
| 175004 | 50002 | 175003 | 50002 | D2 |
| 300005 | 175003 | 300004 | 175003 | D2 |
| 295003 | 170005 | 295004 | 170005 | D2 |
| 305006 | 180008 | 305007 | 180008 | D2 |
| 185010 | 60014 | 185013 | 60014 | D2 |
| 160005 | 35003 | 160004 | 35003 | D2 |
| 365006 | 240002 | 365003 | 240002 | D2 |
| 365006 | 240004 | 365005 | 240004 | D2 |
| 215006 | 90008 | 215007 | 90008 | D2 |
| 13000 | 5006 | 130005 | 5006 | D2 |
| 165001 | 40005 | 165004 | 40005 | D2 |
| 165023 | 40025 | 165024 | 40025 | D2 |
| 340003 | 215005 | 340004 | 215005 | D2 |
| 350006 | 90008 | 35007 | 35006 | D1 |
| 0 | 125004 | 3 | 0 | D1 |
| 125004 | 0 | 125001 | 0 | D2 |
| 275003 | 150005 | 275004 | 150005 | D2 |
| 340006 | 215008 | 340007 | 215008 | D2 |
| 265000 | 140004 | 265003 | 140004 | D2 |
| 275008 | 150010 | 275009 | 150010 | D2 |
| 175004 | 50002 | 175003 | 50002 | D2 |

Dalam tabel 4 terlihat cluster dari masing-masing medoid. Selanjutnya dilakukan perhitungan nilai minimum sebanyak 50 data sehingga diperoleh jarak terdekat pada iterasi 2 adalah 4489234.

e. Hitung Total Simpangan (S)

Jika a adalah jumlah jarak terdekat antara obyek ke medoid awal, dan b adalah jumlah jarak terdekat antara obyek ke medoid baru, maka total simpangan adalah $S=b-a$. Jika $S=0$, maka tukar obyek dengan data untuk membentuk sekumpulan K baru sebagai medoid.

$$S = 4489234 - 2934320$$

$$S = 1554914$$

Karena nilai $S > 0$ maka proses kluster dihentikan atau nilai mendekati 0.

3) Evaluasi jumlah cluster terbaik (Davies Bouldin Index)
Pengukuran dilakukan untuk mengetahui apakah jumlah cluster sudah optimal untuk digunakan. Adapun langkah-langkah perhitungan DBI adalah sebagai berikut:

a. Sum of Square Within - Cluster (SSW)

Menghitung nilai SSW berdasarkan data hasil iterasi terakhir. Selanjutnya menentukan titik pusat cluster baru

dengan menghitung rata-rata untuk dijadikan pusat cluster pada perhitungan SSW. Nilai rata-rata jarak tiap cluster sebagai berikut:

Tabel 5. Titik Pusat Cluster DBI

| Cluster | K1 | K2 | K3 |
|---------|----|----|--------|
| C1 | 10 | 4 | 195833 |
| C2 | 8 | 3 | 383435 |
| C3 | 11 | 3 | 82000 |

Selanjutnya titik pusat tersebut digunakan pada perhitungan nilai kohesi sebuah cluster, cara memperoleh nilainya adalah dengan menghitung jarak Euclidean menggunakan persamaan (2) sebagai berikut :

$$d = \sqrt{(15 - 10)^2 + (3 - 4)^2 + (120000 - 195833)^2} = 75833$$

Perhitungan jarak untuk data pertama menghasilkan nilai 75833, selanjutnya dilakukan perhitungan yang sama pada data ke 2 hingga data terakhir cluster 1, 2 hingga cluster 3. Sebagai contoh hasil dari perhitungan cluster 1 terlihat pada tabel 6.

Tabel 6. Hasil Perhitungan SSW Cluster 1

| Nama Barang | K1 | K2 | K3 | Distance |
|-----------------------------------|----|----|--------|--------------|
| Storage box | 15 | 3 | 120000 | 75833 |
| Pedal dust bin rect | 7 | 3 | 210000 | 14167 |
| Blanket Marshmallow bed cover | 10 | 0 | 240000 | 44167 |
| Blanket Marshmallow bed cover | 10 | 0 | 240000 | 44167 |
| Eftertanka | 10 | 2 | 120000 | 75833 |
| Artificial Pottend Plant With Pot | 8 | 5 | 120000 | 75833 |
| SSW Cluster 1 | | | | 43889 |

Setelah mengetahui nilai jarak sebuah cluster, selanjutnya dilakukan perhitungan nilai SSW menggunakan persamaan (3) sebagai berikut :

$$= 1 \frac{1}{3} * (75833+14167+49167+19167+35833+4167+44167+44167+44167+44167+75833+75833) = 175556$$

Berikut tabel hasil perhitungan jarak dan perhitungan SSW cluster 1 dan cluster 2.

Tabel 7. Hasil Perhitungan SSW Cluster 2

| Nama Barang | K1 | K2 | K3 | Distance |
|----------------|----|----|--------|----------|
| Rigga | 10 | 5 | 360000 | 23435 |
| Gladom | 15 | 3 | 425000 | 41565 |
| Harnavan Troli | 10 | 2 | 350000 | 33435 |

Tabel 7. Hasil Perhitungan SSW Cluster 2 (lanjutan)

| Nama Barang | K1 | K2 | K3 | Distance |
|-------------------------|----|----|--------|--------------|
| Boston Shoe Rack 4 Tier | 15 | 5 | 360000 | 23435 |
| Slimwagoon | 6 | 2 | 260000 | 123435 |
| Bekvam | 5 | 2 | 385000 | 1565 |
| Wir Leaer Shelving Kit | 10 | 5 | 380000 | 3435 |
| Lersta | 12 | 6 | 390000 | 6565 |
| Wall Hook | 20 | 2 | 270000 | 113435 |
| bucket shelf 4 tier | 5 | 3 | 450000 | 66565 |
| Drawer Shelf | 5 | 1 | 450000 | 66565 |
| Mulig | 10 | 8 | 300000 | 83435 |
| Fejan | 10 | 1 | 250000 | 133435 |
| Dacey chair | 10 | 5 | 425000 | 41565 |
| Curtain Set | 5 | 0 | 320000 | 63435 |
| Curtain Set | 5 | 0 | 350000 | 33435 |
| Bed Cover set 180K200 | 6 | 3 | 779000 | 395565 |
| Bed Cover 160K200 | 6 | 3 | 540000 | 156565 |
| Bed Cover 160K200 | 6 | 2 | 520000 | 136565 |
| Bed Cover 120K200 | 6 | 2 | 400000 | 16565 |
| Table Runner set | 5 | 1 | 285000 | 98435 |
| Table Runner set | 5 | 0 | 285000 | 98435 |
| Table Runner set | 5 | 0 | 285000 | 98435 |
| SSW Cluster 2 | | | | 80839 |

Tabel 8. Hasil Perhitungan SSW Cluster 3

| Nama Barang | K1 | K2 | K3 | Distance |
|----------------------|----|----|-------|--------------|
| Stroge box 6,5L | 10 | 2 | 85000 | 3000 |
| Cushion 45x45cm | 20 | 2 | 75000 | 7000 |
| Cushion 30x50cm | 20 | 2 | 60000 | 22000 |
| Bath Towel | 5 | 0 | 70000 | 12000 |
| Storage box 29K23 cm | 10 | 3 | 70000 | 12000 |
| Foldable box | 10 | 0 | 10000 | 18000 |
| Foldable box | 10 | 0 | 10000 | 18000 |
| Jall | 10 | 7 | 10000 | 18000 |
| Albrunna | 10 | 4 | 75000 | 7000 |
| Tromma | 10 | 6 | 10000 | 18000 |
| Bjarsen | 10 | 3 | 90000 | 8000 |
| Bastis | 10 | 4 | 50000 | 32000 |
| Nereby | 5 | 2 | 85000 | 3000 |
| Ostbit | 10 | 6 | 85000 | 3000 |
| Varriera | 8 | 3 | 85000 | 3000 |
| SSW Cluster 3 | | | | 12267 |

b. Menghitung Nilai SSB untuk setiap cluster menggunakan nilai centroid setiap cluster dan nilai dari SSW 1, 2 dan 3. Hasil perhitungan SSB menggunakan persamaan (4) berupa Matriks sebagai berikut:

Tabel 9. Hasil Matrik SSB

| SSB | 1 | 2 | 3 |
|-----|--------|--------|--------|
| 1 | 0 | 187602 | 113833 |
| 2 | 187602 | 0 | 301435 |
| 3 | 113833 | 301435 | 0 |

Sebagai contoh perhitungan hasil pada SSB (1,2) diperoleh dari perhitungan jarak pada titik pusat C1 dan C2 sebagai berikut :

$$SSB_{1,2} = \sqrt{(10 - 8)^2 + (4 - 3)^2 + (195833 - 383435)^2}$$

$$= 187602$$

Pada titik pusat (C1,C3) dan (C2,C3) dilakukan dengan cara yang sama pada perhitungan (C1,C2).

c. *Ratio* (Rasio)

Perhitungan ratio menggunakan persamaan (5), hasil perhitungan terdapat pada tabel 10 berikut:

Tabel 10. Hitung R

| R | 1 | 2 | 3 | Rmax |
|---|-------|-------|-------|-------|
| 1 | 0 | 0,665 | 0,493 | 0,665 |
| 2 | 0,665 | 0 | 0,309 | 0,665 |
| 3 | 0,493 | 0 | 0 | 0,493 |

Sebagai contoh perhitungan hasil hitung rasio (1,2) sebagai berikut :

$$R = \frac{(175556 + 619768)}{187602} = 4$$

Nilai Tmax (1,1), (1,2) dan (1,3) yaitu 4 maka diambil nilai tertinggi pada baris 1

d. *Davies Bouldin Index* (DBI)

Berdasarkan persamaan (6), maka diperoleh :

$$DBI = \frac{1}{3} * (4 + 23 + 14) = 14$$

Semakin kecil nilai DBI yang diperoleh (non-negatif >= 0), maka semakin baik cluster. Berdasarkan perhitungan DBI maka jumlah cluster K=3 dianggap baik.

IV. KESIMPULAN

Pengujian evaluasi DBI (*Davies Bouldin Index*) dengan nilai K=3 yang telah dilakukan pada uji coba di sub bab sebelumnya dari bulan Januari sampai dengan Agustus 2023 sebagai berikut:

Tabel 10. Hasil Evaluasi

| Bulan | Uji 1 | Uji 2 | Terbaik |
|----------|-------|-------|---------|
| Januari | 31 | 14 | 14 |
| Februari | 27 | 10 | 10 |
| Maret | 31 | 6 | 6 |
| April | 15 | 25 | 15 |
| Mei | 11 | 13 | 11 |
| Juni | 17 | 161 | 17 |
| Juli | 7 | 6 | 6 |
| Agustus | 8 | 6 | 6 |

Disimpulkan bahwa diperoleh nilai DBI terbaik berada di bulan Juli dan Agustus. Berdasarkan hasil uji coba yang dilakukan bahwa pengelompokkan produk sangat laris, laris, dan kurang laris berhasil dilakukan dengan perhitungan jarak yang telah dilakukan. Hasil clustering yang dilakukan setiap kali uji coba menghasilkan anggota cluster yang berbeda dikarenakan titik pusat (*centroid*) pada metode k-medoids ditentukan secara acak (*random*). Penerapan teknik data mining klastering sehingga dapat membantu Toko Pernak Pernik dalam menentukan strategi pemasaran sehingga target yang diinginkan bisa tercapai.

REFERENSI

- [1] Asmiatun, Siti, dkk. (2020). "Penerapan Metode K-Medoids Untuk Pengelompokkan Kondisi Jalan Di Kota Semarang". *Jurnal Teknik Informatika dan Sistem Informasi*, Vol. 6 No. 2.
- [2] F. Erika, S. R. W, and Suhada, "Mengkluster jumlah kabupaten/kota yang melaksanakan kawasan tanpa rokok (ktr) di 50% sekolah menurut provinsi dengan k-medoids," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, pp. 637–644, 2019.
- [3] Kusriani, Lutfi, Emha Taufiq. (2009). "Algoritma Data Mining". Yogyakarta. ANDI Yogyakarta.
- [4] P. N. Safitri, R. Aristawidya, and S. B. Faradilla, "KLAUSTERISASI FAKTOR-FAKTOR KEMISKINAN DI PROVINSI JAWA BARAT MENGGUNAKAN K-MEDOIDS CLUSTERING", *JaMES*, vol. 4, no. 2, pp. 75–80, Oct. 2021.
- [5] G. B. Kaligis and S. Yulianto, "ANALISA PERBANDINGAN ALGORITMA K-MEANS, K-MEDOIDS, DAN X-MEANS UNTUK PENGELOMPOKKAN KINERJA PEGAWAI", *itexplore*, vol. 1, no. 3, pp. 179–193, Oct. 2022.
- [6] S. Darma, "Klasterisasi Teknik Promosi dalam Meningkatkan Mutu Kampus Menggunakan Algoritma K-Medoids", *INFEB*, vol. 3, no. 3, pp. 89-94, Mar. 2021.
- [7] M. Nuranti, M. N. Aini, and U. Enri, "Komparasi Distance Measure Pada K-Medoids Clustering untuk Pengelompokkan Penyakit Ispa", *Edumatic*, vol. 5, no. 1, pp. 99–107, Jun. 2021.