

Penentuan Penjualan Barang Berdasarkan Pengelompokan Produk dengan K-Means Clustering Metode CRISP-DM Pada CV.Sembako Dina

¹Narayana Sakti Aji, ²Fauzan Natsir, ³Siti Istianah

^{1,2,3} Teknik Informatika, Universitas Indraprasta PGRI, Jakarta

¹saktiajinarayana@gmail.com, ²fauzan.natsir@gmail.com, ³dewekebae@gmail.com

Abstract - CV Sembako Dina is a staple distributor company that sells products in the form of food and non-food products. Sales of goods require analysis of sales results to increase sales. CV Sembako Dina in determining the sale of goods still does not use a special method as a basis for mapping goods based on the grouping of product selling power. Unbalanced product mapping results in sales that are arranged unevenly, causing a buildup of stock of goods that exceeds its capacity. In addition, uneven sales can also result in no goods being sold, thus reducing company profits. In this research with the CRISP-DM business method using the K-Means algorithm model as one of the algorithms used to cluster products to obtain a product clustering model. K-Means is widely used because of the ease and simplicity of the algorithm. The product group formed is validated by evaluating the Davies-Bouldin Index to obtain the optimal group model. The results obtained the optimal number of groups is 3 categories. Sales of goods and stock management are then mapped based on these 3 categories according to the company's policy in increasing sales.

Keywords - Sale, K-Means, Clustering, Davies-Bouldin Index, CRISP-DM

Abstrak - CV Sembako Dina adalah perusahaan distributor bahan pokok yang menjual produk berupa produk makanan dan non makanan. Penjualan barang memerlukan analisis hasil penjualan sebagai peningkatan penjualan. CV Sembako Dina dalam menentukan penjualan barang masih belum menggunakan metode khusus sebagai landasan untuk memetakan barang berdasarkan pengelompokan daya jual produk. Pemetaan produk yang tidak seimbang mengakibatkan penjualan yang disusun tidak merata sehingga menyebabkan adanya penumpukan stok barang sehingga melebihi kapasitasnya. Selain itu ketidakrataaan penjualan bisa juga mengakibatkan tidak adanya barang yang terjual sehingga menurunkan profit perusahaan. Pada penelitian ini dengan metode bisnis CRISP-DM menggunakan model algoritma K-Means sebagai salah satu algoritma yang digunakan untuk mengklaster produk untuk mendapatkan model pengelompokan produk. K-Means banyak dipakai karena kemudahan dan kesederhanaan algoritmanya. Kelompok produk yang dibentuk divalidasi dengan evaluasi *Davies-Bouldin Index* untuk mendapatkan model kelompok yang optimal. Hasil penelitian didapatkan jumlah kelompok yang optimal yaitu 3 kategori. Penjualan barang dan manajemen stok selanjutnya dipetakan berdasarkan 3 kategori tersebut sesuai kebijakan perusahaan dalam meningkatkan penjualan.

Kata Kunci - Penjualan, K-Means, Clustering, Davies-Bouldin Index, CRISP-DM

I. Pendahuluan

Pada industri distribusi bahan pokok sehari-hari dengan berbagai jenis produk, tantangan kompetisi yang dinamis mendorong inovasi, kelangsungan usaha, dan pertumbuhan. Tiga kebutuhan krusial untuk mencapai tujuan ini melibatkan penambahan produk, peningkatan pemasaran, dan pengurangan biaya operasional. Analisis data penjualan menjadi kunci dalam upaya ini.

Setiap produk memiliki daya jual yang berbeda dan tergantung pada konsumen dan wilayah. Keberhasilan diukur dari seberapa banyak produk terjual. Namun, perlu keseimbangan antara produk laris dan tidak laris, menghindari akumulasi stok dan hilangnya produk. Distributor seperti CV Sembako Dina menghadapi tugas ini dan menggunakan pengelompokan produk berdasarkan penjualan sebelumnya.

Namun, ada kesulitan dalam pengelompokan ini. Produk terlaris memiliki rentang penjualan yang besar, dan ini mempengaruhi analisis. Ini menyebabkan produk rendah bergabung dengan produk populer, mengurangi peluang terjual. Masalah ini mempengaruhi efisiensi, kestabilan penjualan, dan keuntungan. Oleh karena itu, diperlukan metode yang lebih canggih. Penelitian ini memfokuskan pada penggunaan data mining untuk mengekstraksi data demi mendapatkan informasi penting yang bersifat implisit dan sebelumnya tidak diketahui dari suatu data [1], dalam data mining memiliki beberapa teknik, yaitu Estimasi, Prediksi, Klasifikasi, Klastering, dan Asosiasi [2].

CV Sembako Dina bergerak dalam bisnis distribusi bahan pokok sehari-hari dan menghadapi persaingan yang ketat. Pengelompokan produk berdasarkan angka penjualan tahun sebelumnya memberikan gambaran kasar. Analisis ini tidak memadai untuk memahami kebutuhan konsumen, meningkatkan penjualan, atau mengoptimalkan stok [3].

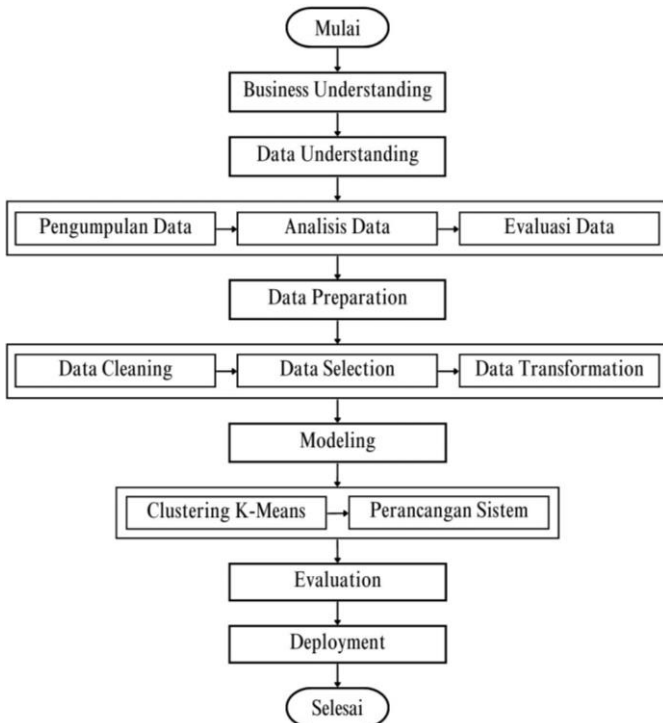
K-Means merupakan salah satu algoritma dalam fungsi *clustering* atau pengelompokan, dengan algoritma K-Means mengacu pada pengelompokan data, observasi atau kasus yang berdasarkan kemiripan suatu objek yang diteliti [4]. Algoritma K-Means cukup mudah dalam implementasi, relatif cepat, mudah disesuaikan dan banyak digunakan [5], dalam penggunaan algoritma K-Means untuk pengelompokan dimulai dengan pembentukan partisi/pusat dari *cluster* di awal, kemudian secara iteratif partisi/pusat dari *cluster* ini diperbaiki sehingga tidak terjadi perubahan yang signifikan pada

partisi/pusat dari *cluster* [6]. Algoritma *K-means* menggunakan pendekatan “*top-down*” yang semula dari awal sudah ditetapkan jumlah *cluster* kemudian dilakukan pengamatan [7]. Dalam mengetahui jumlah dari kelompok *cluster* terbaik digunakan sebuah metode validasi *cluster* untuk memaksimalkan antara jarak *inter cluster* serta meminimalkan jarak *intra cluster*[5].

Pengklasteran dengan menggunakan *Clustering K-Means* didapatkan untuk produk yang laku dengan yang tidak laku. Semakin banyak data penjualan barang yang diinput, maka semakin banyak *clustering centroid* yang terbentuk di proses metode *K-Means* [8]. Tujuan utama dari pengelompokan *clustering* data ini adalah menghasilkan kelompok *cluster* terbaik dan optimal untuk meningkatkan penjualan, manajemen stok barang berlebih, serta mengatasi ketidakseimbangan penjualan produk.

II. Metode Penelitian

Metode yang digunakan pada penelitian kali ini adalah penelitian eksperimen dengan menggunakan *Cross Industry Standard Process Model for Data Mining (CRISP-DM)*. Metodologi data mining *CRISP-DM* sebagai pemecahmasalah yang umum untuk bisnis dan penelitian [9], digunakan karena memiliki tahapan yang kompleks sesuai dengan bisnis yang terdiri dari enam fase [10], yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*.



Gambar 1. Tahapan Penelitian CRISP-DM

1. Business Understanding

Tahap ini adalah tahapan memahami kebutuhan serta menentukan tujuan dari sudut pandang bisnis, menentukan fenomena dan persyaratan dengan jelas secara keseluruhan,

kemudian mengartikan pengetahuan ke dalam bentuk definisikan masalah tersebut serta menentukan pembatasan dalam perumusan masalah yang diaplikasikan dalam data mining dan selanjutnya mempersiapkan rencana serta strategi awal untuk mencapai tujuan data mining.

2. Data Understanding

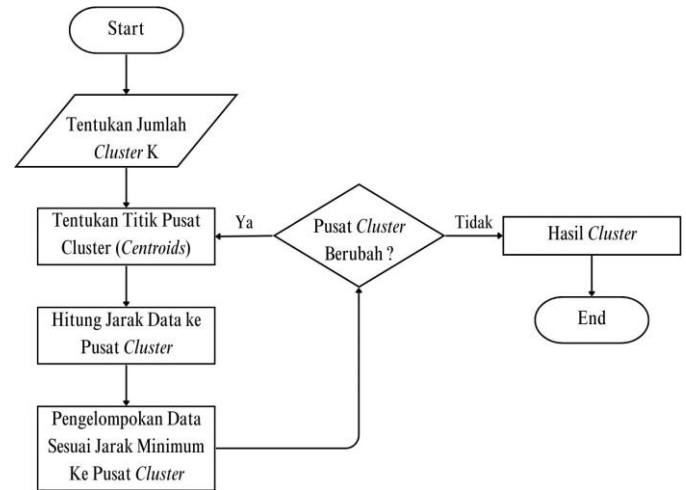
Pemahaman data sebagai awal pengetahuan data sebelum pengolahan data, tahapan ini diawali dengan pengumpulan data, analisis data, dan evaluasi data.

3. Data Preparation

Tahapan ini mencakup pembersihan data (*Data Cleaning*), melakukan pemilihan data (*Data Selection*), record dan atribut-atribut, dan juga melakukan transformasi terhadap data (*Data Transformation*) untuk dijadikan masukan dalam tahap pemodelan[11].

4. Modeling

Tahapan modeling ini secara langsung dengan menggunakan model *clustering* dengan algoritma yang dipilih *K-Means*. Cara algoritma ini mengelompokkan dapat dikatakan serupa dari sebuah data dengan cara mengukur jarak nilai dari satu data dengan data lain.



Gambar 2. Algoritma K-Means Clustering

- Menentukan jumlah *cluster* *k* yang diinginkan.
- Inisialisasi *K* pusat *cluster* (*centroid*) secara random.
- Tempatkan setiap data atau objek ke *cluster* terdekat. Jarak yang dipakai pada algoritma *K-Means* adalah *Euclidean distance* (d).

$$d_{\text{Euclidean distance}}(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

- $x = x_1, x_2, \dots, x_n$, dan $y = y_1, y_2, \dots, y_n$ merupakan banyaknya *n* atribut (kolom) antara 2 record.
- Menghitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata (mean) dari semua data atau objek dalam *cluster* tertentu atau alokasikan masing-masing data ke *centroid* terdekat.

e. Apabila masih terdapat perpindahan data dari satu *cluster* ke *cluster* lainnya, atau apabila perubahan pada nilai centroid masih di atas nilai *threshold* yang ditentukan, kembali lagi ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi (stabil) atau tidak ada penurunan yang signifikan dari nilai SSE (*Sum of Squared Errors*). Jika pusat *cluster* sudah tidak berubah lagi, maka proses pengklasteran selesai.

5. Evaluation

Model evaluasi pada penelitian ini dilakukan dengan melihat tingkat performa dari pola yang dihasilkan oleh algoritma yang diterapkan dengan evaluasi *clustering*. Evaluasi penelitian ini menggunakan algoritma *Davies-Bouldin Index* (DBI).

Algoritma yang digunakan ini Jika nilai *Davies-Bouldin Index* semakin kecil maka semakin baik kelompok *cluster* yang terbentuk, sehingga jumlah *cluster* yang terbentuk lebih efektif terhadap penelitian.

$$DB = \frac{1}{K} \sum_{i=1}^K Ri, qt \quad (2)$$

- $Ri, qt = \max_{j, j \neq i} \left\{ \frac{Si, q + Sj, q}{Dij, t} \right\}$, dengan K adalah banyaknya *cluster*.

Skema *cluster* terbaik ini digunakan untuk menentukan barang yang masuk kedalam kelompok *cluster* terbaik, setelah dilakukan evaluasi *clustering* dengan pengukuran *performance* dari *cluster* yang terbaik dengan menggunakan algoritma *Davies-Bouldin Index* (DBI).

Hasil dari sebuah model *cluster* terbaik antara barang yang kurang laku dengan barang yang paling laku tidak bercampur kedalam data kelompok yang kurang optimal untuk dapat meningkatkan penjualan dan manajemen stok barang.

6. Deployment

Penerapan hasil dari tahapan pengumpulan data hingga tahapan sebuah *modeling* data yang dilakukan dan telah sesuai dengan tujuan bisnis, kemudian hasil keseluruhan digunakan kedalam suatu bisnis tersebut untuk pembuatan laporan hasil *clustering* dan grafik untuk mengetahui peningkatan penjualan dan manajemen stok barang yang dilakukan.

III. Hasil dan Pembahasan

1. Business Understanding

Penerapan data mining pada penelitian ini berhubungan langsung dengan data penjualan barang untuk menggali pengetahuan tentang suatu kelompok *cluster* terbaik terhadap penjualan barang yang mempengaruhi peningkatan penjualan barang, serta manajemen stok barang dengan melihat parameter-parameter apa saja yang mempengaruhinya terhadap tidak meratanya stok dan penjualan barang tersebut.

2. Data Understanding

Pemahaman data sebagai awal pengetahuan data sebelum pengolahan data, tahapan ini diawali dengan pengumpulan data, analisis data, dan evaluasi data.

a. Pengumpulan Data

Tabel 1.
Data Penjualan Barang Periode 1 (Januari - Juli) 2022

| No | Nama Barang | Satuan | Transaksi | Rata-rata Penjualan |
|----|-------------------|--------|-----------|---------------------|
| 1 | Beras | Kg | 10434 | 869,495 |
| 2 | Gula Pasir | Kg | 4497 | 374,745 |
| 3 | Minyak Goreng | Ltr | 8204 | 683,665 |
| 4 | Mentega | Kg | 3286 | 273,83 |
| 5 | Telur | Kg | 7879 | 656,08 |
| 6 | Garam | Kg | 2925 | 243,75 |
| 7 | Gula Merah | Kg | 1447 | 120,58 |
| 8 | Sabun Mandi | Pcs | 1807 | 150,58 |
| 9 | Shampo | Pcs | 2019 | 168,245 |
| 10 | Sikat Gigi | Pcs | 267 | 22,245 |
| 11 | Pasta Gigi | Pcs | 473 | 39,41 |
| 12 | Deterjen | Pcs | 1825 | 152,08 |
| 13 | Sabun Cuci Piring | Pcs | 1386 | 128,995 |
| 14 | Pembersih Lantai | Pcs | 924 | 76,995 |
| 15 | Air Mineral | Dus | 168 | 13,995 |
| 16 | Sirup | Pcs | 97 | 8,08 |
| 17 | Teh | Pcs | 1822 | 151,83 |
| 18 | Kopi | Pcs | 2156 | 179,665 |
| 19 | Mie Instan | Pcs | 8007 | 667,25 |
| 20 | Santan | Pcs | 341 | 28,415 |
| 21 | Biskuit | Pcs | 481 | 40,08 |
| 22 | Roti | Pcs | 354 | 29,495 |
| 23 | Tepung Terigu | Kg | 6424 | 535,33 |
| 24 | Popok | Pcs | 268 | 22,33 |
| 25 | Susu | Pcs | 1726 | 143,83 |
| 26 | Gas Elpiji | Pcs | 697 | 58,08 |
| 27 | Alat Tulis | Pcs | 77 | 6,415 |
| 28 | Kecap | Pcs | 333 | 27,75 |
| 29 | Saus | Pcs | 221 | 18,415 |
| 30 | Obatan Umum | Pcs | 100 | 8,33 |
| 31 | Bumbu Dapur | Pcs | 2146 | 178,83 |
| 32 | Sandal Jepit | Pcs | 34 | 2,83 |
| 33 | Makanan Ringan | Pcs | 1727 | 143,915 |
| 34 | Pembalut | Pcs | 132 | 10,995 |
| 35 | Minuman Sachet | Pcs | 1843 | 153,58 |
| 36 | Baterai | Pcs | 28 | 2,33 |
| 37 | Tisu | Pcs | 845 | 70,415 |
| 38 | Sarden Kaleng | Pcs | 72 | 6 |
| 39 | Bohlam Lampu | Pcs | 28 | 16,33 |
| 40 | Obat Serangga | Pcs | 57 | 4,75 |
| 41 | Sambal | Pcs | 162 | 13,495 |
| 42 | Kantong Plastik | Pcs | 967 | 80,58 |
| 43 | Rokok | Pcs | 4187 | 348,91 |

Tabel 1 menunjukkan sebuah data penjualan barang pada periode 1 yaitu, periode pada bulan Januari sampai dengan Juli 2022 dimana terdapat 43 barang yang terjual.

Tabel 2.
Data Penjualan Barang Periode 2 (Agustus - Desember) 2022

| No | Nama Barang | Satuan | Transaksi | Rata-rata Penjualan |
|----|-------------------|--------|-----------|---------------------|
| 1 | Beras | Kg | 4997 | 832,83 |
| 2 | Gula Pasir | Kg | 1928 | 321,33 |
| 3 | Minyak Goreng | Ltr | 4164 | 694 |
| 4 | Mentega | Kg | 1308 | 218 |
| 5 | Telur | Kg | 4120 | 686,66 |
| 6 | Garam | Kg | 1458 | 243 |
| 7 | Gula Merah | Kg | 649 | 108,16 |
| 8 | Sabun Mandi | Pcs | 842 | 140,33 |
| 9 | Shampo | Pcs | 1015 | 169,16 |
| 10 | Sikat Gigi | Pcs | 100 | 16,66 |
| 11 | Pasta Gigi | Pcs | 259 | 43,16 |
| 12 | Deterjen | Pcs | 839 | 139,83 |
| 13 | Sabun Cuci Piring | Pcs | 701 | 116,83 |
| 14 | Pembersih Lantai | Pcs | 448 | 74,66 |
| 15 | Air Mineral | Dus | 70 | 11,66 |
| 16 | Sirup | Pcs | 30 | 5 |
| 17 | Teh | Pcs | 870 | 145 |
| 18 | Kopi | Pcs | 1124 | 187,33 |
| 19 | Mie Instan | Pcs | 3978 | 663 |
| 20 | Santan | Pcs | 159 | 26,5 |
| 21 | Biskuit | Pcs | 250 | 41,66 |
| 22 | Roti | Pcs | 200 | 33,33 |
| 23 | Tepung Terigu | Kg | 2998 | 499,66 |
| 24 | Popok | Pcs | 123 | 20,5 |
| 25 | Susu | Pcs | 739 | 123,16 |
| 26 | Gas Elpiji | Pcs | 345 | 57,5 |
| 27 | Alat Tulis | Pcs | 20 | 3,33 |
| 28 | Kecap | Pcs | 198 | 33 |
| 29 | Saus | Pcs | 119 | 19,83 |
| 30 | Obatan Umum | Pcs | 57 | 9,5 |
| 31 | Bumbu Dapur | Pcs | 1123 | 187,16 |
| 32 | Sandal Jepit | Pcs | 11 | 1,83 |
| 33 | Makanan Ringan | Pcs | 780 | 130 |
| 34 | Pembalut | Pcs | 65 | 10,83 |
| 35 | Minuman Sachet | Pcs | 997 | 166,16 |
| 36 | Baterai | Pcs | 10 | 1,66 |
| 37 | Tisu | Pcs | 356 | 59,33 |
| 38 | Sarden Kaleng | Pcs | 30 | 5 |
| 39 | Bohlam Lampu | Pcs | 9 | 1,5 |
| 40 | Obat Serangga | Pcs | 30 | 5 |
| 41 | Sambal | Pcs | 65 | 10,83 |
| 42 | Kantong Plastik | Pcs | 510 | 85 |
| 43 | Rokok | Pcs | 2083 | 347,16 |

b. Analisis Data

Pengumpulan data penjualan sembako yang digunakan memiliki 6 atribut data. Pada tabel dibawah ini menunjukkan 6 ringkasan atribut data yang digunakan pada pengumpulan data.

Tabel 3. Atribut yang Digunakan

| No | Field | Tipe |
|----|---------------------|---------|
| 1 | Nama Barang | ID |
| 2 | Satuan | Atribut |
| 3 | Transaksi Periode 1 | Atribut |
| 4 | Transaksi Periode 2 | Atribut |
| 5 | Rata-Rata Penjualan | Atribut |

c. Evaluasi Data

Berdasarkan analisis terhadap data penjualan 1 tahun terakhir dengan membagi menjadi 2 periode dengan rentang per periode adalah 6 bulan, tujuan membagi menjadi data adalah untuk mengetahui tingkat jumlah *cluster* terbaik yang dapat dilakukan analisis dengan melakukan percobaan *clustering* terhadap *dataset* dimulai dengan penentuan nilai $K=2$ sampai dengan penentuan nilai K hingga ditemukannya *cluster* yang tidak memiliki centroid atau anggota.

3. Data Preparation

a. Data Cleaning

Pembersihan data dari dataset penjualan sembako dari 43 data, pembersihan yang dilakukan pembersihan *Missing Value* seperti data kosong, data tidak terukur, dan data yang tidak dilakukan pengukuran. Hal tersebut diatasi dengan menghitung nilai pengganti (*Imputation*) seperti mengganti dengan nilai rata-rata.

b. Data Selection

Melakukan analisis dan memilih setiap variabel atau atribut data yang digunakan untuk proses *transformation*, diantaranya nama barang, transaksi periode 1 dan transaksi periode 2 dilakukan penjumlahan total transaksi, dan total rata-rata penjualan periode 1 dan periode 2 selama 1 tahun [12]. Data tersebut terdapat 43 dataset dengan 4 atribut yang digunakan pada proses *transformation* data, kemudian dataset tersebut digabungkan dengan Excel dan disimpan ke dalam dataset baru menggunakan Microsoft Excel. Hasil pengelolaan data pada tahap ini terdapat pada tabel dibawah ini.

Tabel 4. Atribut Dataset

| No | Atribut | Tipe | Keterangan |
|----|---------------------|------------|----------------------------------------------------|
| 1 | Nama Barang | Nominal/Id | Nama barang yang dijual |
| 2 | Total Transaksi | Numerical | Jumlah total transaksi terjual dalam 1 tahun |
| 3 | Rata-Rata Penjualan | Numerical | Rata-rata penjualan dalam 1 tahun jumlah transaksi |

c. Data Transformation

Tabel 5. Dataset Transformation

| No | Nama Barang | Total Transaksi | Rata-rata Penjualan |
|----|-------------------|-----------------|---------------------|
| 1 | Beras | 10434 | 869,495 |
| 2 | Gula Pasir | 4497 | 374,745 |
| 3 | Minyak Goreng | 8204 | 683,665 |
| 4 | Mentega | 3286 | 273,83 |
| 5 | Telur | 7879 | 656,08 |
| 6 | Garam | 2925 | 243,75 |
| 7 | Gula Merah | 1447 | 120,58 |
| 8 | Sabun Mandi | 1807 | 150,58 |
| 9 | Shampo | 2019 | 168,245 |
| 10 | Sikat Gigi | 267 | 22,245 |
| 11 | Pasta Gigi | 473 | 39,41 |
| 12 | Deterjen | 1825 | 152,08 |
| 13 | Sabun Cuci Piring | 1386 | 128,995 |
| 14 | Pembersih Lantai | 924 | 76,995 |
| 15 | Air Mineral | 168 | 13,995 |
| 16 | Sirup | 97 | 8,08 |

| No | Nama Barang | Total Transaksi | Rata-rata Penjualan |
|----|-----------------|-----------------|---------------------|
| 17 | Teh | 1822 | 151,83 |
| 18 | Kopi | 2156 | 179,665 |
| 19 | Mie Instan | 8007 | 667,25 |
| 20 | Santan | 341 | 28,415 |
| 21 | Biskuit | 481 | 40,08 |
| 22 | Roti | 354 | 29,495 |
| 23 | Tepung Terigu | 6424 | 535,33 |
| 24 | Popok | 268 | 22,33 |
| 25 | Susu | 1726 | 143,83 |
| 26 | Gas Elpiji | 697 | 58,08 |
| 27 | Alat Tulis | 77 | 6,415 |
| 28 | Kecap | 333 | 27,75 |
| 29 | Saus | 221 | 18,415 |
| 30 | Obatan Umum | 100 | 8,33 |
| 31 | Bumbu Dapur | 2146 | 178,83 |
| 32 | Sandal Jepit | 34 | 2,83 |
| 33 | Makanan Ringan | 1727 | 143,915 |
| 34 | Pembalut | 132 | 10,995 |
| 35 | Minuman Sachet | 1843 | 153,58 |
| 36 | Baterai | 28 | 2,33 |
| 37 | Tisu | 845 | 70,415 |
| 38 | Sarden Kaleng | 72 | 6 |
| 39 | Bohlam Lampu | 28 | 16,33 |
| 40 | Obat Serangga | 57 | 4,75 |
| 41 | Sambal | 162 | 13,495 |
| 42 | Kantong Plastik | 967 | 80,58 |
| 43 | Rokok | 4187 | 348,91 |

4. Modeling

Menentukan Jumlah *Cluster* dan Nilai *Centroid* Awal Jumlah *cluster* dan nilai *centroid* awal harus ditentukan selanjutnya [13]. Penentuan pusat *cluster* untuk mengetahui tingkat *cluster* terbaik yang digunakan dengan menentukan mulai dari *cluster* 1 hingga percobaan *cluster* ke 5.

a. Penentuan Pusat Awal *Cluster* (*Centroid*)

Penentuan pusat awal *cluster* pada data penjualan total transaksi dan rata-rata penjualan untuk dilakukan perhitungan.

Centroid untuk cluster 1 menggunakan data beras dengan nilai, *Cluster* 1 = 10434 dan 869,495

Centroid untuk cluster 2 menggunakan data minyak goreng dengan nilai, *Cluster* 2 = 8204 dan 683,665

Centroid untuk cluster 3 menggunakan data tepung terigu dengan nilai, *Cluster* 3 = 6424 dan 535,33

Centroid untuk cluster 4 menggunakan data teh dengan nilai, *Cluster* 4 = 1822 dan 151,83

Centroid untuk cluster 5 menggunakan data bohlam lampu dengan nilai, *Cluster* 5 = 28 dan 16,33

b. Perhitungan Manual

Berikut perhitungan jarak setiap data ke *cluster* C1 – C5 dengan atribut total transaksi dan rata-rata transaksi, perhitungan ini berdasarkan jarak terdekat dari setiap produk dengan *centroid* :

Produk 1 = Jarak data 1 ke *cluster* 1 (10434 dan 869,495)

$$\sqrt{(10432 - 10434)^2 + (869,495 - 869,495)^2} = 0$$

Produk 1 = Jarak data 1 ke *cluster* 2 (8204 dan 683,665)

$$\sqrt{(10432 - 8204)^2 + (869,495 - 683,665)^2} = 2.237,7291$$

Produk 1 = Jarak data 1 ke *cluster* 3 (6424 dan 535,33)

$$\sqrt{(10432 - 6242)^2 + (869,495 - 535,33)^2} = 67,2619$$

Produk 1 = Jarak data 1 ke *cluster* 4 (1822 dan 151,83)

$$\sqrt{(10432 - 1822)^2 + (869,495 - 151,83)^2} = 96,5798$$

Produk 1 = Jarak data 1 ke *cluster* 5 (28 dan 16,33)

$$\sqrt{(10432 - 28)^2 + (869,495 - 16,33)^2} = 106,0997$$

Perhitungan dilakukan terhadap seluruh data hingga 43 data penjualan barang terhadap *cluster* C1-C5 dengan dengan melakukan perulangan perhitungan data dari satu *cluster* ke *cluster* lainnya.

c. Penentuan Nilai K Terbaik *Davies-Bouldin Indeks* (DBI)

Setelah diadakan eksperimen dengan melakukan pengolahan data menggunakan algoritma K-Means dengan jumlah kelompok mulai dari 2 atau K=2 sampai K=5 untuk mendapatkan hasil model *cluster* terbaik berdasarkan *Davies-Bouldin Indeks* (DBI), didapatkan beragam skema *cluster* sebagai berikut ini dengan perhitungan DBI [14].

$$DB = \frac{1}{K} \sum_{i=1}^K Ri, qt \quad (3)$$

Hasil *clustering* dengan K=2 ditunjukkan pada Tabel 4.2 didapatkan ringkasan *centroid* dan jumlah item. C2 adalah *cluster* yang memiliki nilai rata-rata *centroid* terbesar dengan jumlah anggota sebanyak 7 barang. Sisanya adalah anggota C1 dengan anggota sebanyak 36 barang.

Tabel 6. Nilai *Centroid* K=2

| <i>Cluster</i> | Atribut | | Jumlah Item |
|-----------------|-----------------|---------------------|-------------|
| | Total Transaksi | Rata-rata Penjualan | |
| <i>Centroid</i> | C1 | 923,361 | 36 Barang |
| | C2 | 7090,286 | 7 Barang |

Hasil *clustering* dengan K=3 ditunjukkan pada Tabel 4.3 didapatkan ringkasan *centroid* dan jumlah item. C3 adalah *cluster* yang memiliki nilai rata-rata *centroid* terbesar dengan jumlah anggota sebanyak 5 barang. Sisanya adalah anggota C1 sebanyak 23 barang dan C2 sebanyak 15 barang.

Tabel 7. Nilai *Centroid* K=3

| <i>Cluster</i> | Atribut | | Jumlah Item |
|-----------------|-----------------|---------------------|-------------|
| | Total Transaksi | Rata-rata Penjualan | |
| <i>Centroid</i> | C1 | 309,826 | 23 Barang |
| | C2 | 2319,933 | 15 Barang |
| | C3 | 8189,6 | 5 Barang |

Hasil *clustering* dengan K=4 ditunjukkan pada Tabel 4.4 dengan urutan *centroid* tertinggi yaitu C3 dengan anggota sebanyak 5 barang. Berikutnya adalah C4 sebanyak 4 barang, lalu disusul dengan C1 sebanyak 23 barang dan sisanya adalah C2 sebanyak 11 barang.

Tabel 8. Nilai *Centroid* K=4

| Cluster | | Atribut | | Jumlah Item |
|----------|----|-----------------|---------------------|-------------|
| | | Total Transaksi | Rata-rata Penjualan | |
| Centroid | C1 | 1089,455 | 152,012 | 23 Barang |
| | C2 | 309,826 | 26,424 | 11 Barang |
| | C3 | 8189,6 | 682,364 | 5 Barang |
| | C4 | 3723,75 | 310,309 | 4 Barang |

Hasil *clustering* dengan K=5 ditunjukkan pada Tabel 4.5 dengan urutan *centroid* tertinggi yaitu C5 dengan anggota sebanyak 1 barang. Berikutnya adalah C3 sebanyak 4 barang, C1 sebanyak 4 barang, C4 sebanyak 11 barang, C2 sebanyak 23 barang.

Tabel 9. Nilai *Centroid* K=5

| Cluster | | Atribut | | Jumlah Item |
|----------|----|-----------------|---------------------|-------------|
| | | Total Transaksi | Rata-rata Penjualan | |
| Centroid | C1 | 3723,75 | 310,309 | 4 Barang |
| | C2 | 309,826 | 26,424 | 23 Barang |
| | C3 | 7628,5 | 635,581 | 4 Barang |
| | C4 | 1809,455 | 152,012 | 11 Barang |
| | C5 | 10434 | 869,495 | 1 Barang |

5. Evaluation

Hasil evaluasi *Davies-Bouldin Indeks* (DBI) terhadap 5 model *cluster* yang diuji ditunjukkan pada Tabel 4.6.

Tabel 10. Nilai *Davies-Bouldin Index* setiap Model *Cluster*

| | Cluster | | | |
|-----|-----------|-----------|-----------|-----------|
| | 2 Cluster | 3 Cluster | 4 Cluster | 5 Cluster |
| DBI | -0,413 | -0,418 | -0,359 | -0,324 |

Berdasarkan nilai DBI yang ditampilkan pada Tabel 4.6 nilai DBI terendah ditunjukkan pada K=3 atau *cluster* paling ideal dengan nilai DBI -0,418 yaitu sejumlah 3 buah *cluster*. *Cluster* lainnya memiliki nilai DBI lebih besar dibandingkan dengan nilai DBI pada K=3 [15].

Adapun produk-produk yang masuk ke dalam 3 *cluster* tersebut ditampilkan pada Tabel dibawah ini.

Tabel 11. Hasil 3 *Cluster* Skema Terbaik

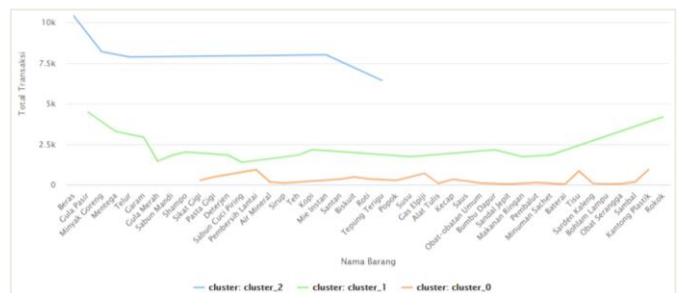
| No | Nama Barang | Cluster | Total Transaksi | Rata-rata Penjualan |
|----|-------------------|-----------|-----------------|---------------------|
| 1 | Beras | cluster_3 | 10434 | 869,495 |
| 2 | Gula Pasir | cluster_1 | 4497 | 374,745 |
| 3 | Minyak Goreng | cluster_3 | 8204 | 683,665 |
| 4 | Mentega | cluster_1 | 3286 | 273,83 |
| 5 | Telur | cluster_3 | 7879 | 656,08 |
| 6 | Garam | cluster_1 | 2925 | 243,75 |
| 7 | Gula Merah | cluster_1 | 1447 | 120,58 |
| 8 | Sabun Mandi | cluster_1 | 1807 | 150,58 |
| 9 | Shampo | cluster_1 | 2019 | 168,245 |
| 10 | Sikat Gigi | cluster_2 | 267 | 22,245 |
| 11 | Pasta Gigi | cluster_3 | 473 | 39,41 |
| 12 | Deterjen | cluster_1 | 1825 | 152,08 |
| 13 | Sabun Cuci Piring | cluster_1 | 1386 | 128,995 |
| 14 | Pembersih Lantai | cluster_2 | 924 | 76,995 |

| No | Nama Barang | Cluster | Total Transaksi | Rata-rata Penjualan |
|----|-----------------|-----------|-----------------|---------------------|
| 15 | Air Mineral | cluster_2 | 168 | 13,995 |
| 16 | Sirup | cluster_2 | 97 | 8,08 |
| 17 | Teh | cluster_1 | 1822 | 151,83 |
| 18 | Kopi | cluster_1 | 2156 | 179,665 |
| 19 | Mie Instan | cluster_3 | 8007 | 667,25 |
| 20 | Santan | cluster_2 | 341 | 28,415 |
| 21 | Biskuit | cluster_1 | 481 | 40,08 |
| 22 | Roti | cluster_2 | 354 | 29,495 |
| 23 | Tepung Terigu | cluster_3 | 6424 | 535,33 |
| 24 | Popok | cluster_2 | 268 | 22,33 |
| 25 | Susu | cluster_1 | 1726 | 143,83 |
| 26 | Gas Elpiji | cluster_2 | 697 | 58,08 |
| 27 | Alat Tulis | cluster_2 | 77 | 6,415 |
| 28 | Kecap | cluster_2 | 333 | 27,75 |
| 29 | Saus | cluster_2 | 221 | 18,415 |
| 30 | Obatan Umum | cluster_2 | 100 | 8,33 |
| 31 | Bumbu Dapur | cluster_1 | 2146 | 178,83 |
| 32 | Sandal Jepit | cluster_2 | 34 | 2,83 |
| 33 | Makanan Ringan | cluster_1 | 1727 | 143,915 |
| 34 | Pembalut | cluster_2 | 132 | 10,995 |
| 35 | Minuman Sachet | cluster_1 | 1843 | 153,58 |
| 36 | Baterai | cluster_2 | 28 | 2,33 |
| 37 | Tisu | cluster_2 | 845 | 70,415 |
| 38 | Sarden Kaleng | cluster_2 | 72 | 6 |
| 39 | Bohlam Lampu | cluster_2 | 28 | 16,33 |
| 40 | Obat Serangga | cluster_2 | 57 | 4,75 |
| 41 | Sambal | cluster_2 | 162 | 13,495 |
| 42 | Kantong Plastik | cluster_2 | 967 | 80,58 |
| 43 | Rokok | cluster_1 | 4187 | 348,91 |

Tabel 4.7 hasil *clustering* menunjukkan ringkasan total barang yang masuk kedalam tiga *cluster* terbaik sebagai berikut.

- Cluster 1* : Produk-produk yang ada pada *cluster* ini memiliki angka tingkat penjualan berada di tengah antara produk terlaris dan tidak laris.
- Cluster 2* : Rata-rata angka penjualan produk pada *cluster* ini lebih sedikit dibandingkan dengan *cluster 1* bahkan terpaut jauh dengan *cluster 3*.
- Cluster 3* : Angka penjualan produk pada *cluster* ini paling laris dibandingkan dengan 2 *cluster* lainnya yaitu, *cluster 1* dan *cluster 2*.

Berdasarkan hasil ulasan ketiga *cluster* diatas, maka grafik dari tiga *cluster* digambarkan menjadi 3 warna *cluster* terbaik.



Gambar 3. Grafik Hasil 3 *Cluster* Terbaik

Berdasarkan tinjauan ulasan hasil grafik ketiga *cluster* diatas, maka penamaan kategori *cluster* berdasarkan dengan tinggi

rendahnya daya jual produk yang terjual yaitu menjadi 3 kategori *cluster*.

Tabel 12. Kategori *Cluster*

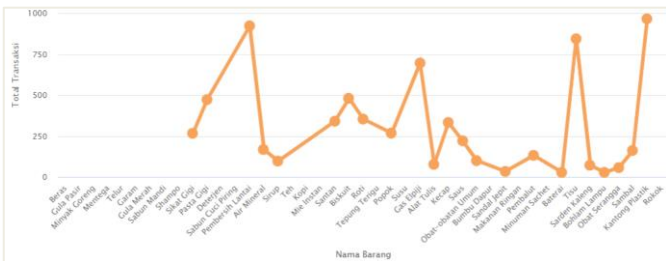
| Cluster | Kategori |
|-----------|--------------------|
| Cluster 1 | Produk Cukup Laku |
| Cluster 2 | Produk Kurang Laku |
| Cluster 3 | Produk Paling Laku |

Dari data Tabel 12 dengan kategori pada setiap *cluster* terbaik memiliki total barang yang terbagi menjadi kelompok 3 *cluster* terbaik diantaranya.

Tabel 13. Total Item 3 Cluster terbaik

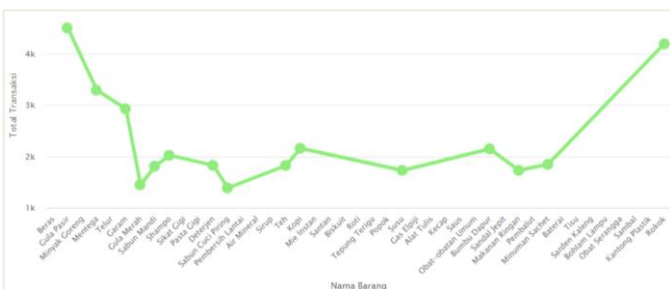
| | 3 Cluster Terbaik | | |
|------------|-------------------|-----------|-----------|
| | Cluster 1 | Cluster 2 | Cluster 3 |
| Total Item | 23 Barang | 15 Barang | 5 Barang |

Pada *cluster 1* yang merupakan kategori produk cukup laku digambarkan grafik nama barang dengan total transaksi penjualan selama periode 1 tahun terakhir.



Gambar 4. Grafik Kelompok *Cluster 1*

Dalam tiga *cluster* terbaik terdapat *cluster* dengan kategori produk kurang laku, yaitu *cluster 1* yang digambarkan pada grafik hasil *clustering* kategori produk kurang laku sebagai berikut.



Gambar 5. Grafik Kelompok *Cluster 2*

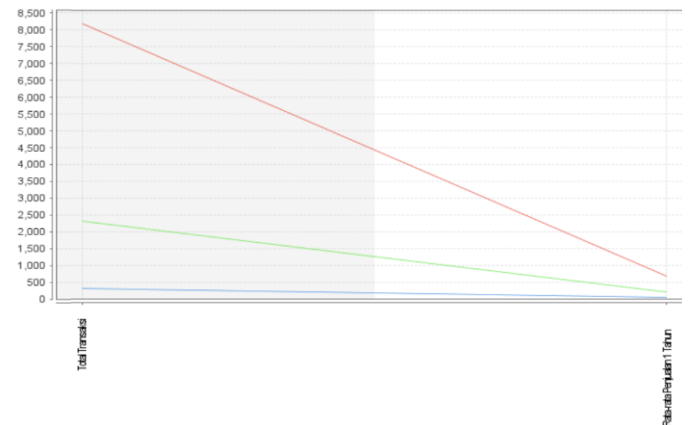
Pada hasil tiga *cluster* terbaik dari hasil perhitunga K-Means *clustering*, yaitu *cluster 3* yang menjadi *cluster* terbaik dengan tingkat penjualan barang paling laku.



Gambar 6. Grafik Kelompok *Cluster 3*

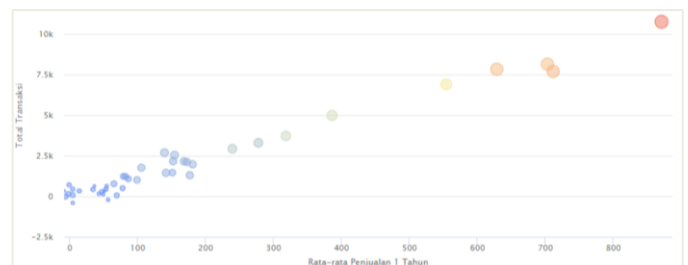
6. Deployment

Total penjualan pada *cluster 2* dengan total item barang terbanyak tetapi memiliki rata-rata penjualan yang rendah yang ditandai dengan warna biru, berbanding terbalik dengan *cluster 3* yang memiliki total item barang lebih sedikit tetapi memiliki tingkat rata-rata penjualan yang tinggi. Sedangkan *cluster 1* berada diantara *cluster 3* dan *cluster 2* yang berada dalam kategori cukup laku. Dari hasil grafik ini dapat mengetahui penjualan barang untuk meningkatkan penjualan dan manajemen stok barang dari hasil kelompok 3 *cluster* terbaik.



Gambar 7. Grafik Perbandingan 3 *Cluster*

Dari hasil grafik perbandingan 3 *cluster* terbaik didapatkan visualisasi dari data 3 *cluster* terbaik dengan penjualan terbanyak berwarna merah.



Gambar 8. Visualisasi Hasil Data Penjualan

IV. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan mendapatkan sebuah kesimpulan bahwa :

1. Hasil pengelompokan produk dengan K-Means terbukti

- efektif karena pemilihan jumlah kelompok divalidasi dengan metode evaluasi *Davies-Bouldin Index*.
2. Jumlah kelompok efisien yang didapatkan dengan *clustering* K-Means berdasarkan dataset penjualan barang tahun 2022 yaitu sebanyak 3 kelompok dengan nilai *Davies-Bouldin Index* terkecil sebesar -0,148 karena semakin kecil nilai DBI maka skema *cluster* semakin efisien.
 3. Hasil penerapan pengelompokan produk yang menggunakan *clustering* yang sudah dievaluasi dengan *Davies-Bouldin Index* berhasil menyelesaikan permasalahan penjualan barang yang tidak merata dan penumpukan stok.
 4. Meningkatnya hasil penjualan dan pengurangan stok barang yang tidak perlu pada periode dimana penjualan dengan *cluster* diterapkan, dibandingkan dengan hasil penjualan barang ditahun sebelumnya yang masih menerapkan penjualan tanpa *cluster*.

V. Daftar Pustaka

- [1] I. H. Witten, E. Frank, dan M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques*. 2011.
- [2] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc., Hoboken, New Jersey., 2005.
- [3] T. Gumelar *et al.*, “Perancangan Sistem Distributor Alat Tulis Kantor,” hal. 492–496, 2023.
- [4] Elly Muningsih dan S. Kiswati, “Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang,” *Jurnal Bianglala Informatika*, vol. 3, no. 1, hal. 10–17, 2015.
- [5] Mulyana, “Penentuan Penjadwalan Pelatihan Berdasarkan Pengelompokan Produk Dengan K-Means Clustering Pada Pt Brainmatics Cipta Informatika,” 2015.
- [6] Agusta Yudi, “K-Means – Penerapan, Permasalahan dan Metode Terkait,” *Jurnal Sistem dan Informatika STMIK STIKOM Bali*, vol. 3, hal. 47–60, 2007.
- [7] G. J. Myatt dan W. P. Johnson, *MAKING SENSE OF DATA: A Practical Guide to Exploratory Data Analysis and Data Mining*. 2014.
- [8] Y. D. Darmi dan A. Setiawan, “Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk,” *Jurnal Media Infotama*, vol. 12, no. 2, hal. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.
- [9] M. A. Hasanah, S. Soim, dan A. S. Handayani, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir,” *Journal of Applied Informatics and Computing*, vol. 5, no. 2, hal. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [10] P. Chapman *et al.*, “CRISP-DM 1.0 Step by Step Data Mining Guide,” *SPSS inc*, vol. 78, hal. 1–78, 2000.
- [11] F. Natsir, “Analisis Forensik Konten dan Timestamp pada Aplikasi Tiktok,” *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, no. 2, hal. 203–209, Des 2021, doi: 10.30998/STRING.V6I2.11454.
- [12] A. M. (Aisyah) Dawis *et al.*, *Rekayasa Perangkat Lunak Panduan Praktis Untuk Pengembangan Aplikasi Berkualitas*. Cv Widina Media Utama, 2023.
- [13] F. F. Nursaid, A. Hendra Brata, dan A. P. Kharisma, “Pengembangan Sistem Informasi Pengelolaan Persediaan Barang Dengan ReactJS Dan React Native Menggunakan Prototype (Studi Kasus : Toko Uda Fajri),” *J-Ptiik.Ub.Ac.Id*, vol. 4, no. 1, hal. 46–55, 2020, [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>.
- [14] F. Natsir, R. A. Sihombing, dan Triyadi, “Pelatihan Programming Junior Pembuatan Game Menggunakan Scratch untuk Sekolah Dasar (SD) Sebagai Upaya Kesiapan Menghadapi Industri Kreatif,” *Jurnal Pengabdian kepada Masyarakat UBJ*, vol. 6, no. 2, hal. 149–158, 2023.
- [15] M. Izzatillah, F. Natsir, dan S. Anisah, “Adaptive Affinity Propagation Untuk Pengelompokan Kehadiran Mahasiswa Pembelajaran,” *NJCA (Nusantara Journal of Computers and Its Applications)*, vol. 6, no. 2, hal. 31–35, 2021.