

Penerapan *K-Means Clustering* Untuk Menentukan Peminatan Barang Pada Jayyid Shop

Fikri Dermawan^{1*}, Gunawan Pria Utama²

^{1,2}Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹*fikridermawan121@gmail.com, ²gunawan.priautama@budiluhur.ac.id
(* : corresponding author)

Abstrak-Dalam dunia bisnis yang kompetitif perlu meningkatkan strategi penjualan barang dan meningkatkan manajemen penyimpanan yang akurat. Jayyid Shop selama ini masih kesulitan untuk menentukan barang yang sangat diminati, diminati, dan kurang diminati oleh pembeli. Akibat dari belum dapatnya menentukan peminatan barang mengakibatkan seringnya kehabisan barang yang sangat diminati dan menumpuknya barang yang kurang diminati. Serta kurang akuratnya strategi penjualan selama ini. Jayyid Shop sendiri terletak di daerah depok, toko ini merupakan sebuah toko *retail* yang menjual berbagai jenis barang seperti kartu perdana, *sparepart handphone*, dan aksesoris *handphone*. Terdapat 221 tipe barang, data penjualan januari dengan total 7903 data, data penjualan februari dengan total 6696 data, dan data penjualan maret dengan total 7259 data. Tujuan penelitian ini agar membantu manajemen Jayyid Shop dalam menentukan barang yang sangat diminati, diminati, dan kurang diminati berdasarkan data laporan penjualan barang selama 3 bulan. Menggunakan *data mining* dengan algoritma *k-means clustering*. Dengan adanya *clustering* data dapat membantu Jayyid Shop untuk membuat pengelompokan data penjualannya agar dapat menentukan permintaan barang. Dalam menentukan peminatan barang menggunakan *field* nama barang, stok, dan laporan penjualan barang selama 3 bulan, kemudian data akan diproses dengan algoritma *k-means clustering*. Hasil akhir dengan 221 tipe barang didapatkan hasil akhir dengan 3 *cluster* yaitu, *cluster 1 (C1)* dengan 68 barang kurang diminati, *cluster 2 (C2)* dengan 69 barang diminati, dan *cluster 3 (C3)* dengan 84 barang sangat diminati. Evaluasi *cluster* menggunakan perhitungan *Davies Bouldin Index (DBI)* menghasilkan 1,94. Jika nilai *Davies Bouldin Index (DBI)* mendekati 0, hasil *cluster* yang diperoleh akan semakin bagus dan menunjukkan bahwa hasil *cluster* yang diperoleh relatif sangat baik.

Kata Kunci: *data mining, k-means, clustering, peminatan, barang*

Application Of K-Means Clustering To Determine The Interest Of Goods At Jayyid Shop

Abstract-In a competitive business world, it is necessary to improve the sales strategy of goods and improve accurate storage management. Jayyid Shop is still having trouble determining which items are in high demand, in demand, and less attractive to buyers. As a result of not being able to determine the specialization of goods, it results in frequent running out of goods that are in great demand and the accumulation of goods that are less desirable. And the lack of accurate sales strategy so far. Jayyid Shop itself is located in the Depok area, this store is a retail store that sells various types of goods such as starter packs, cellphone spare parts, and cellphone accessories. There are 221 types of goods, January sales data with a total of 7903 data, February sales data with a total of 6696 data, and March sales data with a total of 7259 data. The purpose of this study is to assist Jayyid Shop management in determining the items that are in high demand, in demand, and less in demand based on data from the sales report for 3 months. Using data mining with the *k-means clustering* algorithm. With data clustering, it can help Jayyid Shop to group its sales data in order to determine the specialization of goods. In determining the interest of goods using the *field* name of goods, stock, and reports of sales of goods for 3 months, then the data will be processed with the *k-means clustering* algorithm. The final result with 221 types of goods is obtained with 3 clusters, namely, *cluster 1 (C1)* with 68 items that are less desirable, *cluster 2 (C2)* with 69 items of interest, and *cluster 3 (C3)* with 84 items that are in great demand. Cluster evaluation using the *Davies Bouldin Index (DBI)* calculation yields 1.94. If the *Davies Bouldin Index (DBI)* value is close to 0, the cluster results obtained will be better and indicate that the cluster results obtained are relatively very good.

Keywords: *data mining, k-means, clustering, interest, goods*

1. PENDAHULUAN

Perkembangan teknologi semakin pesat di segala bidang khususnya teknologi informasi. Banyak sekali data yang bisa dihasilkan oleh teknologi informasi seperti pada ilmu pengetahuan, industri, ekonomi dan banyak aspek kehidupan lainnya. Dalam dunia bisnis yang kompetitif saat ini, kita perlu terus mengembangkan bisnis kita untuk bertahan dalam persaingan dunia bisnis. Beberapa hal yang perlu dilakukan pemilik bisnis untuk mencapai kemajuan dalam bisnis, yaitu dengan meningkatkan strategi penjualan barang dan meningkatkan manajemen penyimpanan yang akurat. Melalui analisis data penjualan dengan begitu dapat mempermudah pemilik dalam melakukan keputusan dalam manajemen penyimpanan dan strategi penjualan yang akurat, sehingga dapat meningkatkan penjualan.

Jayyid Shop merupakan sebuah toko *retail* yang terletak di daerah depok. Toko ini menjual berbagai kartu perdana, *sparepart handphone*, dan aksesoris *handphone*. Terdapat 221 tipe barang, data penjualan januari dengan total 7903 data, data penjualan february dengan total 6696 data, dan data penjualan maret dengan total 7259 data. Jayyid Shop memiliki banyaknya tipe barang sehingga pihak Jayyid Shop masih kesulitan dalam melakukan strategi penjualan dan mengelola manajemen penyimpanan, serta manajemen Jayyid Shop masih sulit menentukan barang mana saja yang sangat diminati, diminati, dan kurang diminati oleh pembeli. Akibatnya selama ini strategi penjualan masih kurang akurat serta terkadang terjadi penumpukan barang yang kurang diminati dan kehabisan barang yang sangat diminati. Serta dengan banyaknya transaksi penjualan setiap bulannya hingga ribuan data transaksi. Dimana toko ini masih menggunakan keterampilan manusia untuk memungkinkan melakukan analisis data sehingga mengakibatkan kurang akuratnya mengambil keputusan dalam mengelola manajemen penyimpanan barang dan strategi penjualan, sehingga dalam menentukan keputusan dalam manajemen dan menganalisis barang yang sangat diminati, diminati, dan kurang diminati oleh pembeli selama ini masih belum akurat.

Maka dari itu Jayyid Shop memerlukan sebuah sistem atau aplikasi yang bertujuan untuk melakukan perhitungan dalam menganalisis data penjualan agar mempermudah dan lebih akurat dalam menentukan barang mana yang sangat diminati, diminati, dan kurang diminati oleh pembeli. Dengan adanya hasil laporan barang yang sangat diminati, diminati, dan kurang diminati ini dapat digunakan sebagai bantuan keputusan untuk solusi dalam mengambil keputusan dalam manajemen penyimpanan agar meningkatkan penjualan. Dengan menerapkan *data mining* dengan metode *algoritma k-means clustering* pada program aplikasi yang dibuat untuk dapat menganalisis dari data transaksi penjualan selama 3 bulan, sehingga bisa didapatkan hasil mana barang yang sangat diminati, diminati, dan kurang diminati oleh pembeli.

Data mining adalah proses mengekstraksi data dari informasi yang sangat penting. *Data mining* juga merupakan proses analisis pola dalam data. Pola diambil dari berbagai jenis *database*, termasuk *database* relasional, data *warehouse*, data transaksional, dan data berorientasi objek. *Data mining* memungkinkan pemilik bisnis untuk membuat keputusan dengan cepat dan akurat. [1].

Algoritma *k-means* adalah salah satu teknik *clustering* yang paling populer. Algoritma tersebut menunjukkan metode *k-means clustering*. Ide dasarnya adalah dengan pengelompokan awal, pindahkan setiap titik ke pusat baru terdekat, hitung rata-rata titik untuk memperbarui pusat pengelompokan menggunakan kriteria konvergensi (untuk beberapa iterasi, memenuhi selisih nilai) [2].

Pada penelitian implementasi algoritma *k-means* untuk menentukan persediaan barang pada poultry shop [3], Poultry Shop dalam menjaga persediaan stok barang menggunakan algoritma *k-means clustering* dipilih untuk mengatasi masalah ini karena dapat mengelompokkan barang yang terjual dan masih tersedia menjadi beberapa klaster. Dari 3 klaster yang terbentuk, klaster 1 mencakup 2 barang, klaster 2 mencakup 9 barang, dan sisanya 25 barang termasuk dalam klaster 3. Dari hasil tersebut, manajemen Poultry Shop dapat memanfaatkannya untuk meningkatkan *inventory* dan mengambil keputusan untuk strategi penjualan.

Pada penelitian analisa penjualan mobil dengan menggunakan algoritma *k-means* di PT. Mulya Putra Kencana [4], penelitian ini bertujuan untuk mengelompokkan data penjualan ke dalam suatu *cluster* menggunakan metode algoritma *K-Means Clustering*. Data penjualan akan dikelompokkan berdasarkan kesamaan data sehingga data dengan karakteristik yang sama akan dikelompokkan menjadi satu *cluster*. Alhasil, peneliti dapat lebih mudah mengetahui seberapa banyak penjualan yang diminati konsumen. Dengan hasil akhir yaitu *cluster* 0 dengan 90 (paling laku), *cluster* 1 dengan 88 (laku) dan *cluster* 2 dengan 89 (kurang laku).

Pada penelitian implementasi algoritma *k-means* dalam mengkategorikan barang terlaris dan kurang laris pada toko alfamart cikarang [5], penelitian ini mengidentifikasi kelompok produk dari proses *pre-processing* data, 67 *record* data yang diperoleh akan diproses dengan algoritma *k-means clustering*. Dengan hasil akhir, *cluster* pertama adalah kelompok produk dengan persediaan lambat atau rendah dan termasuk dalam daftar produk bergerak lambat. Produk dalam kategori ini adalah 54 dari 67 produk dan *cluster* kedua merupakan kelompok stok cepat atau tinggi dan masuk kategori *fast moving products*, produk dalam kategori ini adalah 13 dari 67 produk dalam proses *cluster*.

Pada penelitian penerapan *data mining* untuk menentukan penjualan *sparepart* toyota dengan metode *k-means clustering* [6], pada permasalahan penelitian ini yaitu kurangnya peninjauan barang yang dibutuhkan konsumen, dan penyimpanan data kurang efisien. Untuk mengatasi permasalahan tersebut, analisis yang digunakan adalah *Clustering* yang menggunakan algoritma *K-Means*. Jadi, melalui proses *clustering* data, CV Terang Jaya dapat mengetahui item paling laris, laris, dan tidak laris. Sehingga barang di gudang tidak menumpuk. Dari penelitian ini, hasil yang dapat diketahui yaitu 15 barang paling laris, 45 barang laris, dan 13 barang tidak laris. Dengan hasil yang dapat pemilik CV Terang Jaya dapat membuat strategi penjualan dengan lebih tepat.

Pada penelitian pengelompokkan data pembelian tinta dengan menggunakan metode *k-means* [7], dari hasil proses *clustering* dapat dikelompokkan barang yang paling banyak keluar pada *cluster* 2 (C2) yang terdiri dari 3 barang, untuk barang yang keluaranya sedang pada *cluster* 1 (C1) yang terdiri 15 barang, dan untuk barang yang sedikit keluaranya pada *cluster* 3 (C3) yang terdiri 24 barang. Sehingga dengan hasil yang diperoleh dengan menggunakan data pembelian tinta dengan tiga atribut barang masuk, barang yang dibeli dan persediaan barang,

memudahkan dan membantu bagian pembelian untuk mengkategorikan barang harus dibeli banyak, sedang, dan sedikit.

Pada penelitian *clustering* data kecelakaan lalu lintas di kecamatan Cileungsi menggunakan metode *k-means* [8], pada penelitian ini bertujuan untuk mengetahui faktor penyebab, waktu dan usia korban kecelakaan yang paling umum. Bagi pemerintah untuk mengambil keputusan berdasarkan hasil penelitian. Dalam penelitian ini dengan diberikan 3 *cluster*, untuk mengklasifikasikan sebagai lokasi yang tidak rawan, rawan, dan sangat rawan. Klaster 1 menghasilkan nilai evaluasi 0.35, klaster 2 menghasilkan nilai evaluasi 0.22, dan kelompok 3 menghasilkan nilai evaluasi 0.38.

Dari penelitian-penelitian sebelumnya penggunaan metode algoritma *k-means clustering* sedikit berbeda dari atribut-atribut yang digunakan, namun pada dasarnya proses algoritma *k-means clustering* masih sama yaitu berfungsi untuk mengelompokkan data menjadi ke beberapa *cluster*. Hanya berbeda dari data yang digunakan, atribut yang digunakan, dan jumlah datanya. Dari hal itu mendapatkan hasil yang berbeda-beda sesuai dari data yang didapat dari tempat penelitian dan keperluan penyelesaian dari tempat penelitian.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Dalam melakukan penelitian, penulis membutuhkan data yang berkaitan dengan topik yang dibahas. Metode pengumpulan data dilakukan dengan tujuan untuk memperoleh data, yaitu:

a. Observasi

Observasi dilakukan untuk meninjau langsung kegiatan dari proses transaksi pada Jayyid Shop. Dari yang diamati berdasarkan observasi, proses transaksi pada Jayyid Shop yaitu pelanggan langsung datang ke toko dan melakukan transaksi secara langsung dengan Jayyid Shop. Pada pendataan Jayyid Shop masih menggunakan buku dan menggunakan Excel.

b. Wawancara

Wawancara dilakukan dengan kepala toko Jayyid Shop yang terkait dengan data yang diperlukan untuk memperoleh data dan informasi tentang proses transaksi saat ini dan tentang sistem penyimpanan data setelah atau sebelum transaksi selesai. Informasi yang didapat akan digunakan sebagai acuan dalam pembuatan sistem aplikasi.

c. Dokumentasi

Dalam dokumentasi, penulis meminta kepada kepala toko Jayyid Shop terkait data yang diperlukan dan tersedia pada Jayyid Shop yang akan digunakan sebagai masukan dalam proses pembuatan sistem aplikasi.

d. Tinjauan Pustaka

Tinjauan pustaka dilakukan untuk mengumpulkan informasi yang tersedia dalam bentuk jurnal yang akan digunakan penulis sebagai acuan dalam melakukan penelitian ini. Dengan jurnal yang terkait *data mining*, dan *k-means clustering*.

2.2 Knowledge Discovery Database (KDD)

Knowledge discovery in database (KDD) adalah aktivitas yang melibatkan pengumpulan dan penggunaan data *historis* untuk menemukan pola atau hubungan yang sering terjadi dalam manajemen data besar [9].

2.2.1 Data Selection

Proses *data selection* dilakukan dengan tujuan untuk memilih atribut yang relevan dalam data sehingga dapat dilakukan analisis data. Dari beberapa atribut yang telah didapatkan dari Jayyid Shop, dipilih 5 atribut seperti berikut:

a. Nama Barang

b. Stok Barang

c. Data Transaksi Bulan Januari 2022

d. Data Transaksi Bulan Februari 2022

e. Data Transaksi Bulan Maret 2022

2.2.2 Pre-Processing / Cleaning

Pada tahap *cleaning* ini, data dibersihkan dari duplikat, kesalahan pada data, dan perubahan data. Pada tahap ini penulis menghapus beberapa data yang memiliki penjualan kosong pada bulan januari, februari, dan maret. Dari 573 baris data yang didapatkan, beberapa data dihapus sehingga menggunakan 221 baris data.

2.2.3 Transformation

Transformasi data yaitu mentransformasikan data terpilih sehingga data tersebut cocok untuk *data mining*. Proses transformasi di KDD adalah proses yang kreatif dan sangat bergantung pada jenis atau pola informasi yang diambil dari *database*. Pada tahap ini penulis tidak melakukan perubahan dikarenakan data yang didapat dari Jayyid Shop sudah sesuai untuk proses *data mining*.

2.2.4 Data Mining

Data yang telah dibersihkan dan dimodifikasi, kemudian diolah untuk dicari informasinya dalam data menggunakan algoritma *k-means clustering*. Data disimpan pada *database* dan diproses secara otomatis dengan aplikasi menggunakan teknik dan perhitungan algoritma *k-means clustering*.

2.2.5 Interpretation / Evaluation

Pada tahap ini setelah data yang telah diolah dengan *data mining* tersebut kemudian diperiksa dalam prosedur evaluasi untuk menentukan informasi yang dihasilkan yang akan digunakan sebagai acuan untuk perencanaan yang optimal pada Jayyid Shop. Pada tahap ini menggunakan metode *Davies Bouldin Index (DBI)*, yang dimana metode ini bertujuan untuk menguji suatu kualitas dari klaster yang dihasilkan.

2.3 Tahapan K-Means Clustering

Referensi [10] menunjukkan pada tahapan dalam melakukan *clustering* atau pengelompokan dengan metode *k-means* adalah sebagai berikut:

- Tentukan jumlah klaster yang ingin dibentuk dan ditetapkan pusat klaster (k).
- Menggunakan jarak *Euclidean* dengan persamaan 1 kemudian hitung setiap data ke pusat klaster.

$$d_{ij} = \sqrt{\sum_{k=1}^m (X_{ij} - C_{ij})^2} \quad (1)$$

Dimana:

- D_{ij} : Jarak objek antar nilai data dan nilai pusat klaster
- m : Jumlah dimensi data
- X_{ij} : Nilai data dari dimensi ke- k
- C_{jk} : Nilai pusat klaster dari dimensi ke- k .

- Kelompokkan data ke dalam klaster dengan jarak yang paling pendek.
- Hitung pusat klaster yang baru dengan persamaan 2.

$$C = \frac{\sum m}{n} \quad (2)$$

Dimana:

- C : Merupakan centroid data
- m : Anggota data yang termasuk ke dalam centroid tertentu
- n : Jumlah data yang menjadi anggota centroid tertentu

- Ulangi langkah dua sampai dengan empat sehingga sudah tidak ada lagi data yang berpindah ke klaster yang lain.

2.4 Pengujian

Dalam penelitian ini digunakan uji *Davies Bouldin Index (DBI)* dimana pada saat evaluasi model digunakan *Davies Bouldin Index (DBI)* untuk memaksimalkan jarak antar *cluster* dan meminimalkan jarak pada *cluster*. [11].

Dalam melakukan evaluasi *Davies Bouldin Index (DBI)* pertama menggunakan *sum of square within cluster (SSW)* yang merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah *cluster* ke- i yang dapat dilihat pada persamaan 3 berikut:

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

Dimana:

m_i adalah jumlah data dalam kluster- i
 c_i adalah centroid kluster ke- i
 $d(x_j, c_i)$ adalah jarak dari data ke- i ke titik kluster i

Setelah itu mencari nilai *sum of square between cluster* (SSB) yang merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan 4 berikut:

$$SSB_{i,j} = d(c_i, c_j) \quad (4)$$

Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran rasio R_{ij} untuk mengetahui nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j menggunakan persamaan 5 berikut:

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

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *Davies Bouldin Index* (DBI) dari persamaan 6 berikut:

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

Dari persamaan tersebut k adalah jumlah *cluster* yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan *K-means* yang digunakan.

3. HASIL DAN PEMBAHASAN

3.1 Penerapan *K-Means Clustering*

Pada sub bab ini akan menjelaskan proses *clustering* dari algoritma *k-means*. Dimana dari atribut yang didapat dari data Jayyid Shop dilakukan perhitungan setiap atribut dengan pusat *cluster*, sehingga dapat setelah didapatkan hasilnya dikelompokkan ke dalam setiap *cluster*. Proses perhitungan kali ini hanya mencontohkan dengan menggunakan 10 data barang yang dapat dilihat pada Tabel 1 di bawah ini:

Tabel 1. Data *Clustering*

No	Nama Barang	Stok	Penjualan		
			Januari	Februari	Maret
1	Vcr Simpati 11GB	280	95	89	83
2	Vcr Three Aon 2GB	67	23	13	26
3	Vcr Smartfren 10GB	100	21	27	30
4	Sp Simpati 17GB	270	97	99	71
5	Sp Isat Biasa	150	36	47	46
6	Sp Smartfren 6GB Non Stop	80	21	22	37
7	Sp Three 4,5 GB	90	28	31	28
8	Sp Simpati 8GB	73	31	15	27
9	Vcr Aigo 3GB	59	21	9	27
10	Vcr Isat 1GB (7 Hari)	75	25	26	23

Pertama kita menentukan nilai centroid awal secara acak. Nilai centroid dapat dilihat pada Tabel 2 di bawah ini:

Tabel 2. Centroid Awal

Centroid	Pusat cluster	Pusat cluster	Pusat cluster	Pusat cluster
C1	4	3	2	1
C2	6	5	4	3
C3	9	8	7	6

Iterasi 1

Vcr Simpati 11GB

$$\text{Jarak ke C1} = \sqrt{(280 - 4)^2 + (95 - 3)^2 + (89 - 2)^2 + (83 - 1)^2} = 314.54$$

$$\text{Jarak ke C2} = \sqrt{(280 - 6)^2 + (95 - 5)^2 + (89 - 4)^2 + (93 - 3)^2} = 311.13$$

$$\text{Jarak ke C3} = \sqrt{(280 - 9)^2 + (95 - 8)^2 + (89 - 7)^2 + (93 - 6)^2} = 306.04$$

Hasil : C3

Vcr Three Aon 2GB

$$\text{Jarak ke C1} = \sqrt{(67 - 4)^2 + (23 - 3)^2 + (13 - 2)^2 + (26 - 1)^2} = 71.52$$

$$\text{Jarak ke C2} = \sqrt{(67 - 6)^2 + (23 - 5)^2 + (13 - 4)^2 + (26 - 3)^2} = 68.23$$

$$\text{Jarak ke C3} = \sqrt{(67 - 9)^2 + (23 - 8)^2 + (13 - 7)^2 + (26 - 6)^2} = 63.44$$

Hasil : C3

Vcr Smartfren 10GB

$$\text{Jarak ke C1} = \sqrt{(100 - 4)^2 + (21 - 3)^2 + (27 - 2)^2 + (30 - 1)^2} = 104.91$$

$$\text{Jarak ke C2} = \sqrt{(100 - 6)^2 + (21 - 5)^2 + (27 - 4)^2 + (30 - 3)^2} = 101.73$$

$$\text{Jarak ke C3} = \sqrt{(100 - 9)^2 + (21 - 8)^2 + (27 - 7)^2 + (30 - 6)^2} = 97.09$$

Hasil : C3

Sp Simpati 17GB

$$\text{Jarak ke C1} = \sqrt{(270 - 4)^2 + (97 - 3)^2 + (99 - 2)^2 + (71 - 1)^2} = 306.43$$

$$\text{Jarak ke C2} = \sqrt{(270 - 6)^2 + (97 - 5)^2 + (99 - 4)^2 + (71 - 3)^2} = 303.00$$

$$\text{Jarak ke C3} = \sqrt{(270 - 9)^2 + (97 - 8)^2 + (99 - 7)^2 + (71 - 6)^2} = 297.88$$

Hasil : C3

Sp Isat Biasa

$$\text{Jarak ke C1} = \sqrt{(150 - 4)^2 + (36 - 3)^2 + (47 - 2)^2 + (46 - 1)^2} = 162.65$$

$$\text{Jarak ke C2} = \sqrt{(150 - 6)^2 + (36 - 5)^2 + (47 - 4)^2 + (46 - 3)^2} = 159.36$$

$$\text{Jarak ke C3} = \sqrt{(150 - 9)^2 + (36 - 8)^2 + (47 - 7)^2 + (46 - 6)^2} = 154.48$$

Hasil : C3

Sp Smartfren 6GB Non Stop

$$\text{Jarak ke C1} = \sqrt{(80 - 4)^2 + (21 - 3)^2 + (22 - 2)^2 + (37 - 1)^2} = 88.29$$

$$\text{Jarak ke C2} = \sqrt{(80 - 6)^2 + (21 - 5)^2 + (22 - 4)^2 + (37 - 3)^2} = 84.92$$

$$\text{Jarak ke C3} = \sqrt{(80 - 9)^2 + (21 - 8)^2 + (22 - 7)^2 + (37 - 6)^2} = 79.97$$

Hasil : C3

Sp Three 4,5 GB

$$\text{Jarak ke C1} = \sqrt{(90 - 4)^2 + (28 - 3)^2 + (31 - 2)^2 + (28 - 1)^2} = 97.93$$

$$\text{Jarak ke C2} = \sqrt{(90 - 6)^2 + (28 - 5)^2 + (31 - 4)^2 + (28 - 3)^2} = 94.55$$

$$\text{Jarak ke C3} = \sqrt{(90 - 9)^2 + (28 - 8)^2 + (31 - 7)^2 + (28 - 6)^2} = 89.56$$

Hasil : C3

Sp Simpati 8GB

$$\text{Jarak ke C1} = \sqrt{(73 - 4)^2 + (31 - 3)^2 + (15 - 2)^2 + (27 - 1)^2} = 79.94$$

$$\text{Jarak ke C2} = \sqrt{(73 - 6)^2 + (31 - 5)^2 + (15 - 4)^2 + (27 - 3)^2} = 76.56$$

$$\text{Jarak ke C3} = \sqrt{(73 - 9)^2 + (31 - 8)^2 + (15 - 7)^2 + (27 - 6)^2} = 71.62$$

Hasil : C3

Vcr Aigo 3GB

$$\text{Jarak ke C1} = \sqrt{(59 - 4)^2 + (21 - 3)^2 + (9 - 2)^2 + (27 - 1)^2} = 63.83$$

$$\text{Jarak ke C2} = \sqrt{(59 - 6)^2 + (21 - 5)^2 + (9 - 4)^2 + (27 - 3)^2} = 60.55$$

$$\text{Jarak ke C3} = \sqrt{(59 - 9)^2 + (21 - 8)^2 + (9 - 7)^2 + (27 - 6)^2} = 55.80$$

Hasil : C3

Vcr Isat 1GB (7 Hari)

$$\text{Jarak ke C1} = \sqrt{(75 - 4)^2 + (25 - 3)^2 + (26 - 2)^2 + (23 - 1)^2} = 81.15$$

$$\text{Jarak ke C2} = \sqrt{(75 - 6)^2 + (25 - 5)^2 + (26 - 4)^2 + (23 - 3)^2} = 77.75$$

$$\text{Jarak ke C3} = \sqrt{(280 - 9)^2 + (95 - 8)^2 + (89 - 7)^2 + (93 - 6)^2} = 72.77$$

Hasil : C3

Centroid iterasi ke-2

C1 = 0,0,0,0

$$C2 = 0,0,0,0$$

$$C3 = \frac{280+67+100+270+150+80+90+75+73+59}{10} = 124.4$$

$$= \frac{95+23+21+97+36+21+28+25+31+25}{10} = 2.1$$

$$= \frac{89+13+27+99+47+22+31+26+15+9}{10} = 0.9$$

$$= \frac{83+26+30+71+46+37+28+23+27+27}{10} = 2.7$$

Untuk iterasi ke-2 dapat dilihat hasilnya pada Tabel 3 di bawah ini:

Tabel 3. Hasil *Clustering* Iterasi Ke-2

No	Nama Barang	C1	C2	C3	Hasil
1	Vcr Simpati 11GB	319.74	319.74	216.91	C3
2	Vcr Three Aon 2GB	76.57	76.57	66.49	C3
3	Vcr Smartfren 10GB	109.86	109.86	48.78	C3
4	Sp Simpati 17GB	311.69	311.69	210.94	C3
5	Sp Isat Biasa	167.69	167.69	76.19	C3
6	Sp Smartfren 6GB Non Stop	93.24	93.24	62.85	C3
7	Sp Three 4,5 GB	103.10	103.10	58.31	C3
8	Vcr Isat 1GB (7 Hari)	86.34	86.34	63.30	C3
9	SP SIMPATI 8GB	85.11	85.11	65.32	C3
10	VCR AIGO 3GB	68.79	68.79	72.74	C1

Jika hasil *clustering* masih berubah maka menentukan nilai centroid untuk iterasi selanjutnya. Nilai centroid dapat dilihat pada Tabel 4 di bawah ini:

Tabel 4. Nilai Centroid Iterasi Ke-3

Centroid	Pusat cluster	Pusat cluster	Pusat cluster	Pusat cluster
C1	59	21	9	27
C2	0	0	0	0
C3	131.67	3.44	1.67	3

Untuk iterasi ke-3 dapat dilihat hasilnya pada Tabel 5 di bawah ini:

Tabel 5. Hasil *Clustering* Iterasi Ke-3

No	Nama Barang	C1	C2	C3	Hasil
1	Vcr Simpati 11GB	252.69	319.74	210.74	C3
2	Vcr Three Aon 2GB	9.22	76.57	72.26	C1
3	Vcr Smartfren 10GB	44.88	109.86	51.79	C1
4	Sp Simpati 17GB	245.63	311.69	204.91	C3
5	Sp Isat Biasa	101.54	167.69	72.80	C3
6	Sp Smartfren 6GB Non Stop	26.65	93.24	67.43	C1
7	Sp Three 4,5 GB	38.67	103.10	61.84	C1
8	Vcr Isat 1GB (7 Hari)	24.02	86.34	68.32	C1
9	Sp Simpati 8GB	18.22	85.11	70.39	C1
10	Vcr Aigo 3GB	0.00	68.79	78.86	C1

Jika hasil *clustering* masih berubah maka menentukan nilai centroid untuk iterasi selanjutnya. Nilai centroid dapat dilihat pada Tabel 6 di bawah ini:

Tabel 6. Nilai Centroid Iterasi Ke-4

Centroid	Pusat cluster	Pusat cluster	Pusat cluster	Pusat cluster
C1	77.71	24.29	20.43	28.29
C2	0	0	0	0
C3	233.33	43.67	45.33	43.00

Untuk iterasi ke-4 dapat dilihat hasilnya pada Tabel 7 di bawah ini:

Tabel 7. Hasil *Clustering* Iterasi Ke-4

No	Nama Barang	C1	C2	C3	Hasil
1	Vcr Simpati 11GB	231.55	319.74	91.21	C3

2	Vcr Three Aon 2GB	13.30	76.57	171.55	C1
3	Vcr Smartfren 10GB	23.53	109.86	137.10	C1
4	Sp Simpati 17GB	224.19	311.69	88.62	C3
5	Sp Isat Biasa	79.89	167.69	83.76	C1
6	Sp Smartfren 6GB Non Stop	9.72	93.24	156.86	C1
7	Sp Three 4,5 GB	16.63	103.10	145.67	C1
8	Vcr Isat 1GB (7 Hari)	8.18	86.34	161.84	C1
9	Sp Simpati 8GB	9.92	85.11	164.45	C1
10	Vcr Aigo 3GB	22.21	68.79	180.23	C1

Jika hasil *clustering* masih berubah maka menentukan nilai centroid untuk iterasi selanjutnya. Nilai centroid dapat dilihat pada Tabel 8 di bawah ini:

Tabel 8. Nilai Centroid Iterasi Ke-5

Centroid	Pusat cluster	Pusat cluster	Pusat cluster	Pusat cluster
C1	86.75	25.75	23.75	30.5
C2	0	0	0	0
C3	275	96	94	77

Untuk iterasi ke-5 dapat dilihat hasilnya pada Tabel 9 di bawah ini:

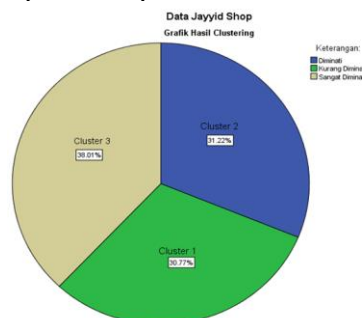
Tabel 9. Hasil *Clustering* Iterasi Ke-5

No	Nama Barang	C1	C2	C3	Hasil
1	Vcr Simpati 11GB	221.71	319.74	9.33	C3
2	Vcr Three Aon 2GB	23.10	76.57	240.32	C1
3	Vcr Smartfren 10GB	14.45	109.86	207.24	C1
4	Sp Simpati 17GB	214.38	311.69	9.33	C3
5	Sp Isat Biasa	69.90	167.69	149.65	C1
6	Sp Smartfren 6GB Non Stop	10.65	93.24	224.58	C1
7	Sp Three 4,5 GB	8.63	103.10	212.65	C1
8	Vcr Isat 1GB (7 Hari)	14.14	86.34	229.31	C1
9	Sp Simpati 8GB	17.48	85.11	231.88	C1
10	Vcr Aigo 3GB	31.98	68.79	249.01	C1

Pada iterasi 4 dan iterasi 5 hasil perhitungan *clustering* data menunjukkan hasil setiap data tidak berpindah *cluster*, maka iterasi selanjutnya tidak perlu dilakukan. Maka hasil akhir dikelompokkan sehingga dari perhitungan yaitu bahwa *cluster* C3 yaitu dengan 2 barang sangat diminati, *cluster* C2 yaitu dengan 0 barang diminati, dan *cluster* C1 yaitu dengan 8 barang kurang diminati.

3.2 Hasil K-Means Clustering

Dalam penerapan metode *data mining* algoritma *k-means clustering* untuk menentukan produk mana saja yang sangat diminati, diminati, dan kurang diminati oleh pembeli. Menggunakan data dari Jayyid Shop dengan *field* nama barang, stok, penjualan januari, februari, dan maret. Menggunakan 221 baris data atau tipe barang, dengan dilakukan *clustering*, dimana pertama ditentukan nilai awal centroid secara *random*, setelah itu perhitungan *clusterisasi* dari iterasi 1 hingga iterasi 8, pada iterasi 8 berhenti dikarenakan pada iterasi 8 dan iterasi 7 hasilnya sama, sehingga pada iterasi 8 diberhentikan dan hasil akhir yang didapatkan dengan 3 *cluster* yaitu, *cluster* 1 (C1) dengan 68 barang kurang diminati, *cluster* 2 (C2) dengan 69 barang diminati, dan *cluster* 3 (C3) dengan 84 barang sangat diminati. Grafik hasil *clustering* dapat dilihat pada Gambar 1 di bawah ini:



Gambar 1. Grafik Hasil *Clustering*

Dengan didapatkan hasil mana barang yang sangat diminati, diminati, dan kurang diminati maka manajemen Jayyid Shop dapat melakukan strategi penjualan, serta mengelola manajemen penyimpanannya. Sehingga kedepannya manajemen Jayyid Shop tidak lagi kehabisan barang yang sangat diminati, dan menumpuknya barang yang kurang diminati. Serta dengan hasil itu manajemen Jayyid Shop dapat melakukan strategi penjualan berupa potongan harga untuk barang yang kurang diminati.

3.3 Hasil Pengujian Data

Pada tahap pertama untuk menentukan SSW_i , dengan persamaan 3 sehingga didapatkan hasil, yaitu:

$$SSW_1 = 4,925147059$$

$$SSW_2 = 67,39405797$$

$$SSW_3 = 104,509881$$

Pada tahap kedua untuk menentukan SSB_{ij} , dengan persamaan 4 didapatkan hasil pada Tabel 10, yaitu:

$$SSB_{1,2} = 107,0330692$$

$$SSB_{1,3} = 173,8618871$$

$$SSB_{2,3} = 66,83628131$$

Tabel 10. Hasil Matrik SSB

SSB	1	2	3
1		107,0330692	173,8618871
2	107,0330692		66,83628131
3	173,8618871	66,83628131	

Pada tahap ketiga untuk menentukan R_{ij} , dengan persamaan 5 didapatkan hasil pada Tabel 11, yaitu:

$$R_{12} = \frac{4,925147059 + 67,39405797}{107,0330692} = 0,675671599$$

$$R_{13} = \frac{4,925147059 + 104,509881}{173,8618871} = 0,629436559$$

$$R_{23} = \frac{67,39405797 + 104,509881}{66,83628131} = 2,572015312$$

Tabel 11. Hasil Matrik R max

R	1	2	3	R max
1		0,675671599	0,629436559	0,675671599
2	0,675671599		2,572015312	2,572015312
3	0,629436559	2,572015312		2,572015312

Setelah itu kita melakukan perhitungan *Davies Bouldin Index* (DBI), dengan persamaan 6 sehingga didapatkan hasil, yaitu:

$$DBI = \frac{1}{3} (0,675671599 + 2,572015312 + 2,572015312) = 1,939900741$$

Evaluasi *cluster* menggunakan perhitungan *Davies Bouldin Index* (DBI) menghasilkan 1,94. Jika nilai *Davies Bouldin Index* (DBI) mendekati 0, hasil *cluster* yang diperoleh akan semakin bagus dan menunjukkan bahwa hasil *cluster* yang diperoleh relatif sangat baik.

4. KESIMPULAN

Dalam melalui proses pengerjaan dan pengujian yang dilakukan pada penelitian ini, maka didapatkan beberapa kesimpulan yaitu, dari 221 tipe barang didapatkan hasil akhir dengan 3 cluster yaitu, cluster 1 (C1) dengan 68 barang kurang diminati, cluster 2 (C2) dengan 69 barang diminati, dan cluster 3 (C3) dengan 84 barang sangat diminati. Evaluasi cluster menggunakan perhitungan *Davies Bouldin Index* (DBI) menghasilkan 1,94. Jika nilai *Davies Bouldin Index* (DBI) mendekati 0, hasil cluster yang diperoleh akan semakin bagus dan menunjukkan bahwa hasil cluster yang diperoleh relatif sangat baik. Dengan adanya pengelompokan data ini, manajemen Jayyid Shop dapat mengetahui barang sangat diminati, diminati, kurang diminati oleh pembeli dengan mudah. Saran untuk penelitian lebih lanjut bisa ditambahkan parameter atau field lain untuk proses pengelompokan penelitian. Serta pada penelitian selanjutnya dapat dilanjutkan dengan menambahkan metode yang lain untuk klasifikasi data.

DAFTAR PUSTAKA

[1] Normah, S. Nurajizah, dan A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan

- Pada Toko Fashion Hijab Banten,” *J. Tek. Komput. AMIK BSI*, vol. 7, hal. 158–163, 2021, doi: 10.31294/jtk.v7i2.10553.
- [2] M. S. Nawawi, F. Sembiring, dan A. Erfina, “Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris,” *Semin. Nas. Teknol. Inf. dan Komun.*, hal. 789-797, 2021, [Daring]. Tersedia pada: <http://prosiding.unipma.ac.id/index.php/SENATIK/article/view/1837>.
- [3] F. Nurdyansyah dan I. Akbar, “Implementasi Algoritma K-Means untuk Menentukan Persediaan Barang pada Poultry Shop,” *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 2, hal. 86–94, Des 2021, doi: 10.26905/jtmi.v7i2.6377.
- [4] Erliyana, O. Nurdiawan, N. R. A. I. Purnamasari, dan I. Ali, “Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana,” *J. DATA Sci. Inform.*, vol. 1, hal. 32–35, 2021, [Daring]. Tersedia pada: <http://publikasi.bigdatascience.id/index.php/jdsi/article/view/12>.
- [5] I. Nawangsih, R. Puspita, dan Suherman, “Implementasi Algoritma K-Means Dalam Mengkategorikan Produk Terlaris Dan Kurang Laris Pada Toko Alfamart Cikarang,” *Pelita Teknol.*, vol. 16, no. 1, hal. 79–87, 2021, [Daring]. Tersedia pada: <https://jurnal.pelitabangsa.ac.id/index.php/pelitatekno/article/view/674>.
- [6] S. P. Tamba, F. T. Kesuma, dan Feryanto, “Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering,” *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, hal. 67–72, 2019, doi: 10.34012/jusikom.v2i2.376.
- [7] Susliansyah, H. Sumarno, H. Priyono, dan N. Hikmah, “Pengelompokan Data Pembelian Tinta Dengan Menggunakan Metode K-Means,” *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 3, no. 2, hal. 381–392, 2019, doi: 10.30645/j-sakti.v3i2.156.
- [8] T. Kurniawan dan M. Jajuli, “Clustering Data Kecelakaan Lalu Lintas di Kecamatan Cileungsi Menggunakan Metode K-Means,” *Gener. J.*, vol. 6, no. 1, hal. 1–12, 2022, doi: 10.29407/gj.v6i1.16103.
- [9] R. Rustam, S. Rahmatullah, S. Supriyanto, dan S. Wahyuni, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Triplek Pada Pt Puncak Menara Hijau Mas,” *J. Inf. dan Komput.*, vol. 8, no. 2, hal. 74–84, 2020, doi: 10.35959/jik.v8i2.186.
- [10] L. Maulida, “Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, hal. 167–174, 2018, doi: 10.14421/jiska.2018.23-06.
- [11] I. R. Mahartika dan A. Wibowo, “Data Mining Klasterisasi dengan Algoritme K-Means untuk Pengelompokan Provinsi Berdasarkan Konsumsi Bahan Bakar Minyak Nasional,” *Pros. Semin. Nas. SISFOTEK (Sistem Inf. dan Teknol.)*, vol. 3, no. 1, hal. 87–91, 2019, [Daring]. Tersedia pada: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/108>.