

KLASTERISASI PENJUALAN PAKAIAN UNTUK MENINGKATKAN STRATEGI PENJUALAN BARANG MENGUNAKAN K-MEANS

Chandra Purnama¹, Wina Witanti², Puspita Nurul Sabrina³

^{1,2,3}Program Studi Teknik Informatika, Universitas Jenderal Achmad Yani

e-mail: chandrapurnama17@if.unjani.ac.id

Abstrak— Semakin berkembangnya persaingan dalam dunia industri khususnya untuk memenuhi kebutuhan primer manusia yang selalu berkembang dan berubah-ubah maka dibutuhkan suatu cara dan juga solusi yang dapat menemukan suatu pola dalam meningkatkan penjualan dan pemasaran barang di perusahaan, salah satunya adalah dengan pemanfaatan data penjualan. Strategi penjualan yang tidak akurat, menyebabkan pembengkakan biaya yang tinggi dan tidak ekonomis, hal ini sangat berbahaya bagi para pelaku bisnis. Salah satu teknik yang dapat digunakan untuk membantu dalam meningkatkan strategi penjualan yang efektif yaitu dengan memanfaatkan data penjualan yang telah tersedia. Sebagai bentuk efisiensi penggunaannya maka dipandang perlu menambahkan suatu sistem yang digunakan untuk melakukan prediksi terhadap sesuatu objek. Klasterisasi yang merupakan sebuah proses untuk mengelompokkan data ke dalam beberapa klaster atau kelompok sehingga data dalam satu klaster memiliki tingkat kemiripan yang maksimum dan data antar klaster memiliki kemiripan yang minimum. Data Penjualan dikelompokkan berdasarkan kemiripan data tersebut sehingga data dengan karakteristik yang sama berada dalam satu klaster dan salah satu atribut yang digunakan adalah jenis pakaian dan penjualan.

Keywords: Algoritma, K-Means, Klaster, Klasterisasi, Strategi Penjualan

Abstract— The growing competition in the industrial world, especially to meet the primary needs of humans who are always evolving and changing, it requires a way and also a solution that can find a pattern in increasing sales and marketing of goods in the company, one of which is the use of sales data. Inaccurate sales strategy, causing high cost overruns and uneconomical, this is very dangerous for business people. One technique that can be used to assist in improving an effective sales strategy is by utilizing available sales data. As a form of efficiency in its use, it is deemed necessary to add a system that is used to predict an object. Clustering is a process to group data into several clusters or groups so that the data in one cluster has the maximum level of similarity and the data between clusters has the minimum similarity. Sales data are grouped based on the similarity of the data so that data with the same characteristics are in one cluster and one of the attributes used is the type of clothing and sales.

Keywords: Algoritma, K-Means, Cluster, Clustering, sales strategy

I. PENDAHULUAN

Salah satu upaya agar dapat menyeimbangi permintaan konsumen yaitu mengetahui perkembangan konsumen dan minat konsumen yang berubah-ubah [1]. Untuk menangani masalah tersebut maka dibutuhkan suatu strategi penjualan barang yang baik dalam pengolahannya, salah satu cara tersebut yaitu menggunakan data mining. Data mining merupakan suatu proses ekstraksi atau penggalian data dan informasi yang besar, yang belum diketahui sebelumnya, namun dapat dipahamidan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting [2].

Ada beberapa cara untuk menangani penjualan pakaian tersebut yaitu dengan cara memaksimalkan penjualan dengan merekomendasikan produk yang memiliki nilai jual lebih tinggi, banyaknya barang yang terjual atau dengan cara menawarkan produk-produk lain [3]. Hal ini bisa dilakukan dengan menggunakan *k-means* dengan cara menganalisa data dengan melakukan proses pemodelan tanpa supervisi (unsupervised) dimana data dalam satu kelompok mempunyai

karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain [4].

Dengan diadakannya penelitian agar dapat menghasilkan sebuah sistem dalam menentukan strategi penjualan barang dengan melakukan proses klasterisasi, sehingga dapat melakukan efisiensi secara cermat dalam hal penjualan barang.

II. METODE

Metode penelitian ini dilakukan beberapa tahapan seperti perolehan data, kemudian pra proses, serta proses data mining tersebut.

A. Perolehan Data

Dalam melakukan penelitian dibutuhkan data yang dapat digunakan sebagai pendukung penelitian, dimana data yang digunakan yaitu data penjualan milik Distro. Data Penelitian tersebut dapat dirumuskan suatu permasalahan dan menentukan objek yang digunakan

dalam penelitian. Data yang diperoleh yaitu data penjualan pakaian seperti pada Tabel 1.

Tabel 1 Data Penjualan Pakaian

No.	Kode Penjualan	Tanggal	Nama Produk	Qty
1	EPDC01	2021-01-01	Coach Jacket Basic EPDC I	1
2	EPDC02	2021-01-01	Hoodie EPDC Full Up DCS I	1
3	EPDC03	2021-01-02	Hoodie EPDC Full Up BDG	1
4	EPDC04	2021-01-02	T-Shirt EPDC Bad Shoes	2
5	EPDC05	2021-01-03	Jacket My Bullet CJ	1
6	EPDC06	2021-01-03	Backpack EPDC II	1
7	EPDC07	2021-01-03	T-Shirt EPDC Squad	1
8	EPDC08	2021-01-03	Hoodie EPDC Full Up United Heaven	1
9	EPDC09	2021-01-04	Hoodie EPDC Full Up Punish Power	1
10	EPDC010	2021-01-05	T-Shirt Hagibi Cj Rose I	2
...
1650	EPC01650	2021-12-31	Kemeja Adapt	1

B. Pra Proses

Pre-processing merupakan tahapan sebelum melakukan penggunaan data sebelum proses pengolahan yang dilakukan oleh sistem maka harus dilakukan terlebih dahulu metode *pre-processing* untuk melakukan perubahan data penjualan produk yang sudah tersedia agar selanjutnya dapat di proses menjadi data yang siap untuk di olah sebagai objek penelitian. Dalam proses *pre-processing* terdapat beberapa tahapan diantaranya yaitu *data selection*, dan *data cleaning*[5].

1. Data Selection

Data Selection merupakan tahapan yang dilakukan untuk menyeleksi data yang sudah dimiliki dan sudah mempunyai informasi yang lengkap pada setiap *attribute*. Data tersebut diseleksi agar dapat mengelompokkan *attribute* sesuai dengan yang dibutuhkan, dimana *attribute* yang dibutuhkan diantaranya Kode Penjualan, Nama Produk, Tanggal dan Qty

2. Data Cleaning

Pada tahap *Data Cleaning* dilakukan pembersihan data untuk melakukan analisa terhadap data yang tidak mempunyai kelengkapan data [6].

3. Data Transformasi

Data Transformation merupakan tahapan yang digunakan untuk merubah bentuk data asli menjadi data numerik untuk algoritma *K-Means Clustering* dan integrasi berdasarkan data yang sudah dipilih sehingga data yang dimiliki dapat diinisialisasi dan dapat digunakan,

C. Clustering dengan metode K-Means

Data Mining adalah proses yang melakukan pengolahan data histori kejadian-kejadian sebelumnya

dan digunakan sebagai dasar untuk membangun sebuah pengetahuan sebuah sistem prediksi dari sebuah kasus terdahulu[7]. serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [8]. Pada tahap ini merupakan proses perhitungan *K-Means Clustering*. Dalam metode ini data yang memiliki karakteristik yang sama dikelompokkan dalam satu kelompok dan data yang lain dikelompokkan dengan kelompok yang sesuai dengan karakteristik data tersebut. Tahap klasterisasi sebagai berikut:

1. Pilih jumlah banyaknya klaster, jumlah banyaknya k tidak boleh melebihi dari jumlah data.
2. Inisialisasi k pusat klister ini bisa dilakukan dengan diberi nilai awal dengan angka-angka acak.
3. Menempatkan semua data klister terdekkan. Kedekatan tersebut ditentukan berdasarkan jarak antar objek. Untuk menghitung jarak semua data dapat digunakan dengan rumus jarak *Euclidean Distance* dengan rumus sebagai berikut.

$$(i, j) = \sqrt{(x_i - y_j)^2 + (x_i - y_j)^2}$$

- D (i,j) = Jarak data ke i ke pusat klaster j
 - x_i = Data ke i pada atribut data ke k
 - y_j = Titik pusat ke j pada atribut ke k
4. Kelompokkan hasil perhitungan jarak tersebut berdasarkan jarak minimum. Kemudian hitung kembali pusat centroid dengan klaster yang baru.
 5. Melakukan kembali perhitungan jarak antar objek dengan menggunakan pusat centroid yang baru jika pusat klaster masih belum konvergen.
 6. Setelah mendapatkan nilai yang tidak berubah/konvergen, maka selanjutnya yaitu menghitung rata-rata nilai setiap titik data dengan menggunakan *Davies Bouldin Index (DBI)*
 - a. Cara pertama: Menentukan *Sum of square within cluster (SSW)* dengan data cara menjumlahkan data dari masing-masing klaster yang nantinya dibagi dengan jumlah data klaster tersebut.
 - b. Cara kedua: Menentukan *Sum of square between cluster (SSB)* dengan membutuhkan centroid terakhir pada iterasi terakhir.
 - c. Cara ketiga: Mencari jarak antar rasio, dimana hasil dari *Sum of square within cluster (SSW)* dibagi dengan *Sum of square between cluster (SSB)*.
 - d. Cara keempat: Mengitung *Davies Bouldin Index*.

III. HASIL DAN PEMBAHASAN

A. Pembahasan I

Pembahasan ini berupa hasil peneliti yang telah dilakukan menggunakan metode *K-Means Clustering* sebelumnya data telah dilakukan pra proses seperti *data selection* dan *data cleaning*, kemudian data diolah sehingga menghasilkan dataset yang dapat digunakan seperti pada Tabel 2.

Tabel 2 Data set

Nama Produk	Jan	Feb	Mar	Apr	...	Nov	Des
Coach Jacket Basic EPDC I	1	1	1	1	...	3	2
Hoodie EPDC Full Up DCS I	1	1	1	5	...	3	2
Hoodie EPDC Full Up BDG	2	1	1	1	...	1	1
T-Shirt EPDC Bad Shoes	2	7	8	22	...	2	6
Jacket My Bullet CJ	4	4	3	2	...	12	8
Backpack EPDC II	1	1	3	1	...	2	2
T-Shirt EPDC Squad	6	8	6	2	...	2	11
Hoodie EPDC Full Up United Heaven	1	1	1	4	...	4	2
Hoodie EPDC Full Up Punish Power	4	2	2	2	...	4	6
T-Shirt Hagibi Cj Rose I	3	2	1	1	...	1	2
...
Kemeja Adapt	10	8	6	22	...	6	4

Setelah menentukan dataset yang akan digunakan, kemudian menentukan centroid awal. Centroid awal yang digunakan dapat dilihat pada Tabel 3.

Tabel 3 Centroid Awal

	Jan	Feb	Mar	Apr	...	Nov	Des
C1	1	1	1	1	...	3	2
C2	1	1	1	5	...	3	2

Kemudian hitung Centroid Euliden Distance dengan menggunakan rumus :

$$D(i, j) = \sqrt{(x_i - y_j)^2 + (x_i - y_j)^2}$$

Contoh :

Perhitungan jarak pertama dan data pertama dengan pusat kluster pertama adalah:

Perhitungan data 1, centroid 1

$$\begin{aligned} &= \sqrt{(x_i - y_j)^2 + (x_i - y_j)^2} \\ &= \sqrt{(1 - 1)^2 + (1 - 1)^2 + (1 - 1)^2 + (1 - 1)^2 + \dots + (3 - 3)^2 + (2 - 2)^2} \\ &= 0 \end{aligned}$$

Perhitungan data 1, centroid 2

$$\begin{aligned} &= \sqrt{(x_i - y_j)^2 + (x_i - y_j)^2} \\ &= \sqrt{(1 - 1)^2 + (1 - 1)^2 + (1 - 1)^2 + (1 - 5)^2 + \dots + (3 - 4)^2 + (2 - 2)^2} \\ &= 8,24 \end{aligned}$$

Jika jarak data dengan Centroid Euliden Distance telah diketahui maka kelompokkan data berdasarkan Klasternya. Pengelompokkan dibedakan berdasarkan jarak terpendek seperti pada Tabel 4

Tabel 4 Hasil Pengelompokan data

No.	Nama Produk	C1	C2	Klaster
1	Coach Jacket Basic EPDC I	0	8,24	1
2	Hoodie EPDC Full Up DCS I	8,24	0	2
3	Hoodie EPDC Full Up BDG	8	8,59	1
4	T-Shirt EPDC Bad Shoes	24,93	20,97	2
5	Jacket My Bullet CJ	14,45	15,9	1
6	Backpack EPDC II82	18,76	12,16	2
7	T-Shirt EPDC Squad	19,28	20,445	1
8	Hoodie EPDC Full Up United Heaven	9,3	10,58	1
9	Hoodie EPDC Full Up Punish Power	7,68	11,09	1
10	T-Shirt Hagibi Cj Rose I	8,54	10,04	1
...
1650.	Kemeja Adapt	25,82	23,47	2

Untuk centroid baru didapatkan dari jumlah seluruh data sebuah atribut pada satu centroid dibagi dengan jumlah data dan berlaku untuk seterusnya untuk semua atribut centroid seperti pada Tabel 5.

Tabel 5 Centroid iterasi kedua

	Jan	Feb	Mar	Apr	...	Nov	Des
C1	7	2,71	2,41	1,85	...	3,85	4,57
C2	3,5	4,25	4,5	15	...	3,25	3,5

Setelah mendapatkan nilai rata-rata yang tidak berubah atau konvergen, maka selanjutnya menghitung rata-rata nilai setiap titik pada himpunan data atau bisa disebut juga dengan *Davies Bouldin Index (DBI)*

Cara pertama : Menentukan *Sum of square within cluster (SSW)* dimana data dari masing-masing kluster dijumlahkan, kemudian dibagi dengan jumlah data tersebut, dapat dilihat pada Tabel 6.

Tabel 6 Hasil Sum of square within cluster (SSW)

	Jan	Feb	Mar	Apr	...	Nov	Des
C1	3	2,71	2,28	1,85	...	3,85	4,57
C2	3,5	4,25	4,5	12,5	...	3,25	3,5

Cara kedua : Menentukan *Sum of square between cluster (SSB)*, cara ini membutuhkan centroid terakhir pada iterasi terakhir.

Contoh :

$$\begin{aligned} \text{SSB}_{1,2} &= \sqrt{(3 - 3,5)^2 + (2,71 - 4,25)^2 + \dots} \\ &= \sqrt{(0,25) + (0,9216) + \dots} \\ &= 15,3707449 \end{aligned}$$

Cara ketiga : Mencari rasio antar cluster dimana SSW dibagi dengan SSB.

Cara keempat : Mencari nilai *Davies Bouldin Index*.

$$\begin{aligned} \text{DBI} &= R1 + R2 / K \\ &= 0,438 \end{aligned}$$

B. Pembahasan II

Perhitungan yang telah dilakukan menggunakan metode *K-Means Clustering* selanjutnya dilakukan uji *performace* yang digunakan untuk menilai bagaimana baik atau tidaknya hasil klaster dilihat berdasarkan kedekatan dan kuantitas antar data dari hasil klaster. Berikut ini merupakan tabel hasil perhitungan pengujian dengan menggunakan *Davies Bouldien Index* dapat dilihat pada Tabel 7.

Tabel 7 Pengujian *K-Means Clustering*

No.	Perhitungan	Hasil
1.	<i>Average within centroid tance</i>	1573283272.446
2.	<i>Average within centroid tance cluster 1</i>	3637232728.513
3	<i>Average within centroid tance cluster 2</i>	8291345358.825
4.	<i>Davies Bouldien index</i>	0.438

Berdasarkan hasil perhitungan yang telah dilakukan seperti pada tabel diatas, dapat dilihat didapatkan nilai *Davies Bouldien index* sebesar 0,438. Dimana dengan nilai yang didapat tersebut dapat dikatakan kualitas klaster cukup baik.

Kemudian hasil akhir yang deperoleh setelah beberapa kali iterasi dan dilakukan nya perhitungan maka akan menghasilkan hasil seperti pada gambar Tabel 8 dan Tabel 9.

Tabel 8 Hasil Klasterisasi 1

No.	Nama Produk	Jenis Produk	QT Y	Bulan	Warna	Klaster
1	Kemeja Adapt	Kemeja	98	April	Dark	1
2	EPDC Bad Shoes	Sepatu	95	September	Grey	1

Tabel 9 Hasil Klasterisasi 2

No.	Nama Produk	Jenis Produk	QT Y	Bulan	Warna	Klaster
1	Jacket My Bullet CJ	Jaket	70	Oktober	Biru	2
2	T-Shirt Hagibi CJ Docktrin	T-Shirt	12	Oktober	Hitam	2

IV. KESIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan dapat disimpulkan bahwa penggunaan metode *K-Means* untuk meningkatkan strategi pejualan karena hasil dari penelitian ini yaitu memberikan suatu bentuk rekomendasi berupa nama produk, jenis produk, banyaknya produk yang terjual, bulan dimana produk tersebut terjual paling banyak dan warnanya berdasarkan data yang digunakan melalui proses klasterisasi

REFERENSI

- [1] Y. Hartati, S. Defit, and G. W. Nurcahyo, "Klasterisasi Bibit Terbaik Menggunakan Algoritma *K-Means* dalam Meningkatkan Penjualan," *J. Inform. Ekon. Bisnis*, vol. 3, pp. 6–11, 2021, doi: 10.37034/infeb.v3i1.56.
- [2] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode *K-Means* dalam pengelompokkan penjualan produk pada Swalayan Fadhlila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [3] . F., F. T. Kesuma, and S. P. Tamba, "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, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.
- [4] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan *K-Means Cluster* Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.
- [5] M. Mardalius, "Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma *K-Means*," *Jurteksi*, vol. 4, no. 2, pp. 123–132, 2018, doi: 10.33330/jurteksi.v4i2.36.
- [6] H. Leidiyana, "Penerapan Algoritma *K-Nearest Neighbor* Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, no. 1, pp. 65–76, 2013.
- [7] F. Gorunescu, *Intelligent System Reference Library, Volume 12*. 2005.
- [8] M. Bramer, *Principles of Data Mining*, no. January 2007. 2007.