

## Implementasi Data Mining Untuk Prediksi Stok Penjualan Keramik dengan Metode K-Means

Ferdian Arya Dinata, Alwis Nazir<sup>\*</sup>, Muhammad Fikry, Iis Afrianty

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1</sup>11950115065@students.uin-suska.ac.id, <sup>2,\*</sup>alwis.nazir@uin-suska.ac.id, <sup>3</sup>muhammad.fikry@uin-suska.ac.id, <sup>4</sup>iis.afrianty@uin-suska.ac.id

Email Penulis Korespondensi: alwis.nazir@uin-suska.ac.id

Submitted: 20/05/2024; Accepted: 31/05/2024; Published: 31/05/2024

**Abstrak**—Keramik menjadi salah satu produk yang diminati pelanggan dari tahun ke tahun sehingga banyaknya perusahaan tertarik untuk melakukan penjualan kermik. Namun penjualan keramik harus memenuhi dan mengimbangi kebutuhan pelanggan yang berubah-ubah serta ditemukannya permasalahan mengenai produk keramik dan pelanggan seperti kurangnya stok produk keramik yang mengakibatkan pelanggan tidak jadi melakukan pemesanan dan penjualan produk tidak memenuhi target. Pengelompokan keramik dibutuhkan untuk mengantisipasi resiko yang akan diterima perusahaan dengan pemanfaatan proses data mining menggunakan data dimasa lampau. Pada penelitian ini menggunakan metode K-Means yang terdapat pada data mining. Tujuan penelitian ini adalah dapat mengelompokkan dan menentukan penjualan merek yang potensial untuk penambahan stok dimasa depan serta melakukan pengujian data menggunakan DBI (*Davies Bouldin Index*) yang dilakukan dengan menguji nilai jarak antar *cluster* melalui serangkaian eksperimen. Penelitian ini menggunakan data 1 tahun terakhir dari Januari 2022 hingga Desember 2022 dengan total 156 data menggunakan 9 atribut yaitu merek, kode barang (FT,WT) dan ukuran (40x40, 25x25, 50x50, 25x40, 60x60, 20x40). Hasil penelitian menggunakan metode K-Means penjualan yang sangat laris adalah klaster 2, penjualan yang laris adalah klaster 1 dan yang tidak laris adalah klaster 0. Merek yang sangat laris adalah HRM, yang laris adalah VALENSIA dan yang tidak laris adalah MCC. Hasil uji menggunakan metode DBI dengan validitas 1.013 menunjukkan bahwa *cluster* terbaik diperoleh pada k=3 dengan menggunakan metode *elbow*. Diharapkan penelitian ini berkontribusi untuk perusahaan terkait sebagai pendukung pengambilan keputusan.

**Kata Kunci:** Clustering; Data Mining; K-Means; Keramik; Penjualan

**Abstract**—Ceramics has become one goods that consumers show interest in every year, so many companies are interested in selling ceramics. However, ceramic sales must meet and balance changing customer needs as well as problems found regarding ceramic products and customers, such as a lack of stock of ceramic products which results in customers not placing orders and product sales not meeting targets. So it is necessary to group ceramics to anticipate the risks that the company will accept by utilizing the data mining process using past data. This research uses the K-Means method found in data mining. The objective of this research is to group determine sales of brands that have potential for additional stock in the future and to test the data using the DBI (*Davies Bouldin Index*) which is carried out by testing the distance values between clusters through a series of experiments. This research uses data for the last 1 year from January 2022 to December 2022 with a total of 156 data using 9 attributes, namely brand, item code (FT, WT) and size (40x40, 25x25, 50x50, 25x40, 60x60, 20x40). The results of the research using the K-Means method, the best-selling brand is cluster 2, the best-selling brand is cluster 1 and the best-selling brand is cluster 0. The best-selling brand is HRM, the best-selling brand is VALENSIA and the best-selling brand is MCC. Test results using the DBI method with a validity of 01.013 show that the best cluster is obtained at k=3 using the elbow method. It is hoped that this research will contribute to related companies as support for decision making.

**Keywords:** Clustering; Data Mining; K-Means; Ceramics; Sale

### 1. PENDAHULUAN

Penjualan merupakan suatu upaya untuk memenuhi dan mengimbangi kebutuhan pelanggan yang berubah-ubah [1]. Salah satu peluang usaha yang ada yaitu dibidang penjualan keramik [2]. Hal ini dikarenakan banyaknya pelanggan yang ingin melakukan pemesanan produk keramik pada toko tersebut [3]. Keramik adalah salah satu produk yang dibuat sebagai bahan dasarnya yaitu tanah liat dan berfungsi sebagai pelapis lantai, dinding serta menambah keindahan. Kebutuhan keramik pada saat sekarang meningkatkan tinggi berdasarkan permintaan pelanggan.

Penjualan keramik sering kali mengalami beberapa permasalahan terhadap produk dan pelanggan [3]. Salah satu perusahaan yang bergerak dibidang penjualan keramik yaitu CV. XYZ yang selalu memperhatikan penjualan terkhusus keramik. Persaingan perusahaan merupakan suatu hal yang umum dan perusahaan dituntut untuk selalu memperbaharui serta mengembangkan peluang bisnis yang ada [4]. Berdasarkan hasil wawancara dengan salah satu karyawan CV. XYZ, beliau menjabarkan bahwa ditemukannya permasalahan mengenai produk keramik dan pelanggan seperti kurangnya stok produk keramik yang mengakibatkan pelanggan tidak jadi melakukan pemesanan dan penjualan produk tidak memenuhi target. Urgensi penelitian ini mengangkat permasalahan yaitu perusahaan mengalami kerugian seperti kerugian penjualan, banyaknya pengeluaran biaya pada produk keramik yang kurang diminati serta target penjualan yang tidak tercapai. Oleh karena itu perusahaan harus tepat dalam pengambilan keputusan terhadap strategi persediaan keramik yang diminati oleh pelanggan.

Beberapa perusahaan menggunakan metode *clustering* dalam data mining untuk memprediksi stok penjualan keramik [5]. Data mining merupakan suatu kegiatan pengambilan untuk memperbaiki keputusan di

masa depan dengan pengumpulan dan pemakaian data historis dalam menemukan keteraturan pola [6]. Dalam literatur data mining beberapa teknik yang sering digunakan salah satunya yaitu *clustering*. Metode *clustering* adalah metode untuk memecahkan masalah dalam suatu pengelompokan data menggunakan analisa data [7]. Salah satu metode *clustering* yaitu metode K-Means. Suatu metode yang melakukan pengelompokan data dalam jumlah yang cukup besar dengan perhitungan waktu relative cepat dan efisien merupakan pengertian dari K-Means [8]. Metode ini banyak digunakan dalam pengelompokan pada bidang bisnis, kesehatan, pendidikan, dan bidang bisnis lainnya [9].

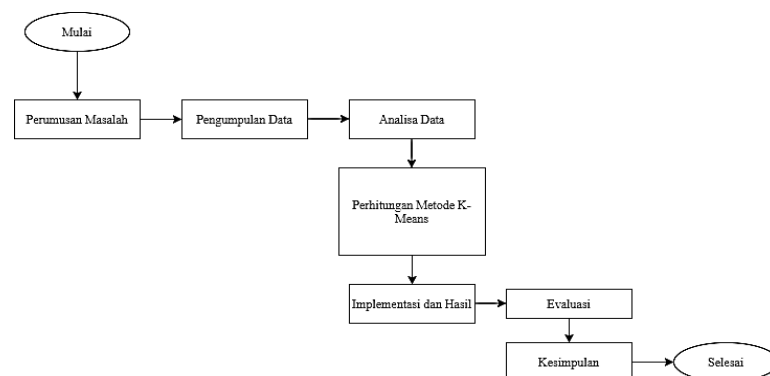
Berdasarkan *research gap* penelitian sebelumnya mengenai klasterisasi penjualan pakaian untuk meningkatkan strategi penjualan barang menggunakan K-Means yang bertujuan untuk membantu dalam meningkatkan strategi penjualan yang efektif dengan hasil penelitian memberikan rekomendasi berupa nama produk, jenis produk, banyaknya produk yang terjual, bulan produk yang terjual paling banyak dan warna berdasarkan data yang digunakan melalui proses klasterisasi [1]. Penelitian selanjutnya menggunakan K-Means untuk pemilihan keramik dan pelanggan potensial pada CV. Jaya Tunggal Keramik dan dapat disimpulkan bahwa hasil pengujian 34 data keramik bulan November hingga Desember 2017 hanya 6 keramik yang mendekati perkiraan stok sedangkan 23 keramik tidak sesuai prediksi dan 5 keramik tidak dilakukannya stok barang karena kurang laku dan pelanggan potensial adalah Laguna karena sering melakukan transaksi [3].

Penelitian selanjutnya yaitu pengklusteran penjualan bahan bangunan menggunakan algoritma K-Means dan hasil penelitian ini menggunakan 30 data sebanyak 5 iterasi serta didapatkan bahan bangunan yang laris dibeli konsumen yaitu: Batu Ultra Tipis, Besi 4 Pas, Cat Bravo 5kg, Gembok 40 Risk, Keni 3 inci, Keramik Alberta Brown 40/40 [6]. Selanjutnya penelitian mengenai penerapan data mining untuk prediksi penjualan mobil menggunakan metode K-Means *clustering* dengan menggunakan data 5 tahun dari 900 data yaitu *cluster* 0 laris, *cluster* 1 kurang laris, dan *cluster* 2 sangat laris. Membantu pihak perusahaan dengan penjualan mobil yang masuk *cluster* 1 dapat melakukan strategi untuk meningkatkan penjualan [10]. Selanjutnya penelitian dengan judul pendekatan K-Means *clustering* metode elbow pada motivasi pengunjung festival halal dengan mendapatkan kesimpulan bahwa berdasarkan metode elbow yang digunakan penelitian motivasional dibagi menjadi tiga *cluster* yang berbeda seperti edukasional, rekreasional dan transaksional [11].

Berdasarkan penelitian sebelumnya data yang diperoleh sesuai dan efektif jika menggunakan algoritma K-Means untuk memprediksi stok penjualan [10][12]. Pengelompokan produk keramik merupakan suatu teknik yang dapat digunakan untuk pengambilan keputusan oleh perusahaan. Permasalahan yang telah dijabarkan, maka penelitian ini akan mengelompokkan produk keramik berdasarkan pembelian data pelanggan dengan 9 atribut yaitu merek, kode barang (FT,WT) dan ukuran (40x40, 25x25, 50x50, 25x40, 60x60, 20x40). Penelitian ini mengimplementasikan metode K-Means untuk memperkirakan stok yang harus disediakan oleh perusahaan berdasarkan permintaan pelanggan [13] [14]. Dengan memanfaatkan metode *elbow* untuk menentukan jumlah *cluster* yang optimal dari suatu algoritma proses melihat siku pada sebuah grafik titik dimana penurunan nilai kriteria evaluasi menjadi lebih lambat. Tujuan penelitian ini adalah dapat mengelompokkan dan menentukan penjualan merek yang potensial untuk penambahan stok dimasa depan. Data yang akan digunakan pada penelitian ini ialah data 1 tahun terakhir yaitu Januari 2022 hingga Desember 2022 dengan jumlah 156 data. Menentukan dan mengevaluasi hasil penelitian ini menggunakan DBI (*Davies Bouldin Index*) [15]. DBI merupakan suatu matrik *cluster* yang akan dianggap memiliki skema *clustering* yang optimal adalah yang memiliki DBI minimal [16].

## 2. METODOLOGI PENELITIAN

Penelitian ini akan mengikuti beberapa tahap, rangkaian proses dalam penelitian ini dapat dilihat sebagai mana pada gambar 1 berikut:



**Gambar 1.** Tahapan Penelitian

Pada gambar 1 diatas menjelaskan tentang beberapa tahapan yang akan dilakukan pada rangkaian proses penelitian ini.

## 2.1 Perumusan Masalah

Tahapan awal yang krusial dalam metodologi penelitian adalah menentukan permasalahan. Penelitian ini bertujuan untuk mengenali dan mengidentifikasi masalah yang ada, serta mencari solusi yang diharapkan. Menentukan masalah dengan akurat akan membantu peneliti merancang rencana penelitian yang sesuai dan mempermudah proses pengumpulan data serta analisis hasil penelitian.

## 2.2 Pengumpulan Data

Pengumpulan data adalah proses untuk memperoleh informasi dari permasalahan yang terkait. Data penelitian ini dikumpulkan dalam format excel yang diperoleh dari CV. XYZ dengan total 156 data.

## 2.3 Analisis Data

Tahapan ini merupakan tahapan dimana data yang dibutuhkan dalam penelitian dikumpulkan untuk dapat diolah selanjutnya. Data diperoleh dari CV. XYZ dari Januari 2022 hingga Desember 2022 dengan total 156 data. Data yang digunakan yaitu merek, kode barang (FT, WT) dan ukuran (40x40, 25x25, 50x50, 25x40, 60x60, 20x40). Selanjutnya data diolah yang akan menjadi sumber informasi yang dibutuhkan pada penelitian ini. Data ini akan diolah lebih lanjut dan dijadikan informasi untuk pihak selanjutnya.

### 2.3.1 Data Selection

Setelah proses pengumpulan data, langkah berikutnya yang diambil oleh peneliti adalah melakukan seleksi data. Seleksi data merupakan proses pemilihan atribut tertentu dari dataset yang lebih besar. Dalam penelitian ini peneliti memilih atribut merek, kode barang (FT, WT) dan ukuran (40x40, 25x25, 50x50, 25x40, 60x60, 20x40).

### 2.3.2 Data Preprocessing

Dalam tahap ini, dilakukan *transformasi* data sesuai kebutuhan penulis. *Transformasi* tersebut dalam dilihat pada tabel 1 berikut ini.

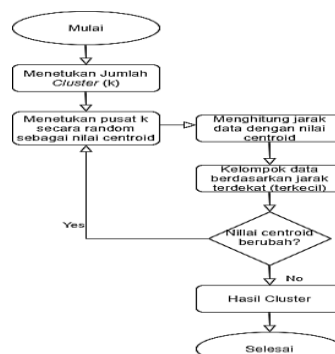
**Tabel 1.** Data Transformasi

Kode barang	Merek
1	ALBATROS
2	ANTARES
3	EXCELENTE
4	FORTUNER
5	GLD
6	HRM
7	MAJESTIC
8	MCC
9	MYLEANO
10	PICASSO
11	RAYMOND
12	TOTAL
13	VALENSIA

Tabel 1 diatas merupakan transformasi data yang telah dilakukan oleh penulis sesuai dengan kebutuhan berdasarkan data sebelumnya.

## 2.4 Perhitungan Metode K-Means

Penelitian ini menggunakan metode K-Means. Setelah melakukan pengumpulan data dan analisa data yang ada. Adapun tahapan pada proses metode K-Means dibawah ini gambar 2.



**Gambar 2.** Proses K-Means

Berikut rincian mengenai proses perhitungan metode K-Means yang dapat dilihat pada gambar 2:

- Langkah awal dengan menentukan jumlah *cluster* (*k*) terlebih dahulu dengan metode *elbow* dengan melihat grafik siku.
- Menentukan titik nilai *cluster* (*centroid*) secara random sebanyak *cluster* (*k*).
- Menghitung jarak setiap data terhadap titik *cluster* data dengan rumus *Euclidean distance* dibawah ini.

$$d(x,y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (1)$$

Dimana  $d(x,y)$  merupakan jarak data ke pusat *cluster*, simbol  $x$  adalah data asli, simbol  $y$  adalah data *centroid* dan  $i$  merupakan jumlah atribut atau parameter data.

- Mengelompokkan data yang telah dihitung berdasarkan jarak terdekat (terkecil) antara data yang tersebut dengan pusat *cluster* atau data *centroid* dan menjadikan sebuah *cluster* baru.
- Hitung ulang pusat *cluster* (*centroid*) baru berdasarkan data yang mengikuti *cluster* masing-masing. Nilai *centroid* baru diperoleh dengan menghitung rata-rata dari data setiap *cluster*. Rumus sebagai berikut.

$$CI = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x} \quad (2)$$

Penjelasan rumus diatas yaitu  $CI$  merupakan nilai *centroid* baru,  $x_1$  merupakan nilai data *cluster* ke-1,  $x_n$  merupakan nilai data *cluster* ke-n dan  $\sum x$  adalah jumlah data

- Didapatkan titik *centroid* baru, maka lakukan iterasi 2 atau ulangi langkah c sampai e hingga tidak ada data yang berpindah *cluster*.

## 2.5 Implementasi dan Hasil

Tahapan yang dilakukan yaitu implementasi pengolahan data dengan *Python* sehingga menghasilkan model dari data tersebut.

## 2.6 Evaluasi

Evaluasi ketepatan hasil pengelompokan data berdasarkan data yang telah diolah dengan metode K-Means dengan mengukur menggunakan DBI. Metode DBI merupakan metode untuk mengevaluasi kualitas *cluster* dengan menghitung nilai baik suatu *cluster* data dengan cara menghitung rasio antara jarak *cluster* dengan *cluster* lainnya.

## 2.7 Kesimpulan

Hasil yang telah didapatkan berdasarkan dari perhitungan metode K-Means tahapan akhir ini selesai.

# 3. HASIL DAN PEMBAHASAN

Pada penelitian ini, sumber data yang digunakan berasal dari CV. XYZ. Atribut yang digunakan pada penelitian ini merek, kode barang (FT, WT) dan ukuran (40x40, 25x25, 50x50, 25x40, 60x60, 20x40). Data dapat dilihat pada tabel 2 berikut.

Tabel 2. Dataset

No	Merek	Kode Barang		Ukuran					
		FT	WT	40X40	25X25	50X5	25X4	60X6	20X4
1	9	106	0	103	0	0	0	0	0
2	13	29	0	29	0	0	0	0	0
3	9	120	103	31	111	55	26	0	0
4	7	0	0	0	0	0	0	0	0
5	3	48	17	27	11	0	27	0	0
6	4	47	25	0	28	0	44	0	0
7	1	17	0	4	0	13	0	0	0
8	6	56	76	9	2	21	96	4	0
9	11	20	5	17	2	0	6	0	0
10	5	27	0	0	0	27	0	0	0
--	--	--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--	--	--
154	10	3	0	3	0	0	0	0	0
155	13	0	2	0	1	0	1	0	0
156	2	40	0	1	0	39	0	0	0

Tabel 2 merupakan sumber data yang digunakan pada penelitian ini.

### 3.1 Pengolahan Data

Data di atas akan diproses dengan menggunakan metode K-Means. Berikut langkah-langkah dalam pengolahan data menggunakan metode K-Means.

- Tentukan jumlah kluster ( $k$ ) berdasarkan metode *elbow*. Pada kasus ini kluster optimal yang diambil adalah 3 kluster.
- Pilih nilai awal untuk titik kluster (*centroid*) sejumlah kluster ( $k$ ). Nilai *centroid* dapat dipilih secara acak. Berikut merupakan titik awal kluster yang terdapat dalam tabel 3.

**Tabel 3.** Nilai kluster awal (Iterasi pertama)

Centroid	Merek	Kode Barang		Ukuran					
		FT	WT	40X40	25X25	50X50	25X40	60X60	20X40
0	3	48	17	27	11	0	27	0	0
1	11	20	5	17	2	0	6	0	0
2	13	19	24	19	9	0	12	0	3

Tabel 3 diatas merupakan titik awal kluster yang mana nilai *centroid* dipilih secara acak.

- Menghitung jarak terdekat data terhadap setiap nilai *centroid*.

- Perhitungan data pertama terhadap setiap nilai *cluster* 0

$$d(1,0) = \sqrt{(9-3)^2 + (106-48)^2 + (0-17)^2 + (103-27)^2 + (0-11)^2 + (0-0)^2 + \sqrt{(0-27)^2 + (0-0)^2(0-0)^2} = 101,5627885$$

- Perhitungan data ke pertama terhadap pusat *cluster* 1

$$d(1,1) = \sqrt{(9-11)^2 + (106-20)^2 + (0-5)^2 + (103-17)^2 + (0-2)^2 + (0-0)^2 + \sqrt{(0-6)^2 + (0-0)^2(0-0)^2} = 121,9057013$$

- Perhitungan data ke pertama terhadap pusat *cluster* 2

$$d(1,1) = \sqrt{(9-13)^2 + (106-19)^2 + (0-24)^2 + (103-19)^2 + (0-9)^2 + (0-0)^2 + \sqrt{(0-12)^2 + (0-0)^2(0-3)^2} = 124,3020515$$

Dengan menggunakan rumus *Euclidean Distance* pada data pertama, ditemukan bahwa data tersebut memiliki jarak terdekat dengan *cluster* 0 berdasarkan nilai terkecil di antara perhitungan jarak antara *cluster* 1. Oleh karena itu, data pertama termasuk dalam kelompok *cluster* 0. Proses ini kemudian diulang hingga data 156 sehingga setiap data dapat menemukan hasil jarak yang terdekat dengan *cluster*.

- Jika sudah menghitung jarak data ke centroid, langkah berikutnya adalah mengelompokkan data berdasarkan nilai terdekat (yang paling kecil). Berikut adalah hasil pengelompokannya pada tabel 4.

**Tabel 4.** Hasil Pengelompokan Data

No	C0	C1	C2	Jarak Terdekat	Cluster
1	101,5627885	121,9057013	124,3020515	101,5627885	C0
2	40,0499	17,1464	31,7804	17,1464	C1
3	160,1811	187,3766	173,8850	160,1811	C0
4	64,7147	27,7488	39,5979	27,7488	C1
5	0	40,1746	35,9444	0	C0
6	37,0540	59,8915	51,1957	37,0540	C0
--	--	--	--	--	--
--	--	--	--	--	--
155	45,4092	31,3049	31,2409	31,2409	C2
156	58,3108	48,1975	56,7186	48,1975	C1

Hasil pengelompokan data dapat dilihat pada tabel 4 berdasarkan nilai terdekat (yang paling kecil) setelah menghitung jarak data ke centroid.

- Melakukan pembaharuan nilai *centroid* baru pada iterasi selanjutnya dilakukan menghitung nilai rata-rata dari setiap atribut data yang terdapat dalam *cluster*. Berikut ini adalah hasil perhitungan nilai *centroid* yang baru.

- Rata-rata data pada *cluster* 0

$$CI(\text{Merek}) = \frac{432}{60} = 7,2$$

$$CI(50x50) = \frac{801}{60} = 13,35$$

$$CI(\text{FT}) = \frac{4040}{60} = 67,333$$

$$CI(25x40) = \frac{1285}{60} = 21,416$$

$$CI(WT) = \frac{1221}{60} = 20,35$$

$$CI(60x60) = \frac{24}{60} = 0,4$$

$$CI(40x40) = \frac{2200}{60} = 36,666$$

$$CI(20x40) = \frac{0}{60} = 0$$

$$CI(25x25) = \frac{960}{60} = 16$$

## 2. Rata-rata pada *cluster* 1

$$CI(\text{Merek}) = \frac{471}{74} = 6,364$$

$$CI(50x50) = \frac{952}{74} = 12,86$$

$$CI(FT) = \frac{1262}{74} = 17,054$$

$$CI(25x40) = \frac{139}{74} = 1,878$$

$$CI(WT) = \frac{148}{74} = 2$$

$$CI(60x60) = \frac{9}{74} = 0,121$$

$$CI(40x40) = \frac{195}{74} = 2,645$$

$$CI(20x40) = \frac{34}{74} = 0,459$$

$$CI(25x25) = \frac{73}{74} = 0,986$$

## 3. Rata-rata pada *cluster* 2

$$CI(\text{Merek}) = \frac{202}{22} = 9,181$$

$$CI(50x50) = \frac{53}{22} = 2,409$$

$$CI(FT) = \frac{351}{22} = 15,954$$

$$CI(25x40) = \frac{640}{22} = 29,090$$

$$CI(WT) = \frac{768}{22} = 34,909$$

$$CI(60x60) = \frac{19}{22} = 0,863$$

$$CI(40x40) = \frac{128}{22} = 5,818$$

$$CI(20x40) = \frac{79}{22} = 3,590$$

$$CI(25x25) = \frac{202}{22} = 9,181$$

**Tabel 5.** Nilai *cluster* awal (Iterasi kedua)

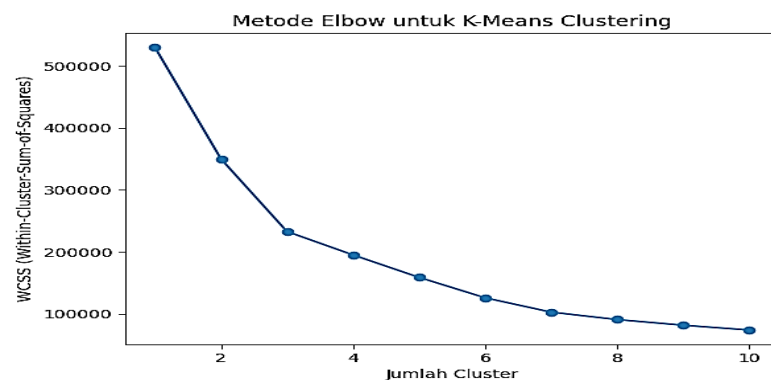
Centroid	Merek	Kode Barang				Ukuran			
		FT	WT	40X40	25X25	50X50	25X40	60X60	20X40
0	7,2	67,333	20,35	36,666	16	13,35	21,416	0,4	0
1	6,364	17,054	2	2,635	0,986	12,86	1,878	0,121	0,459
2	9,181	15,954	34,909	5,818	9,181	2,409	29,090	0,863	3,590

Tabel 5 merupakan nilai *cluster* awal pada iterasi kedua.

- f. Dilakukan perhitungan jarak kembali antara data dan nilai *centorid* baru. Lakukan perulangan pada proses ini sehingga tidak ada data yang beralih ke *cluster* lain atau iterasi terakhir memiliki data yang sama dengan iterasi sebelumnya

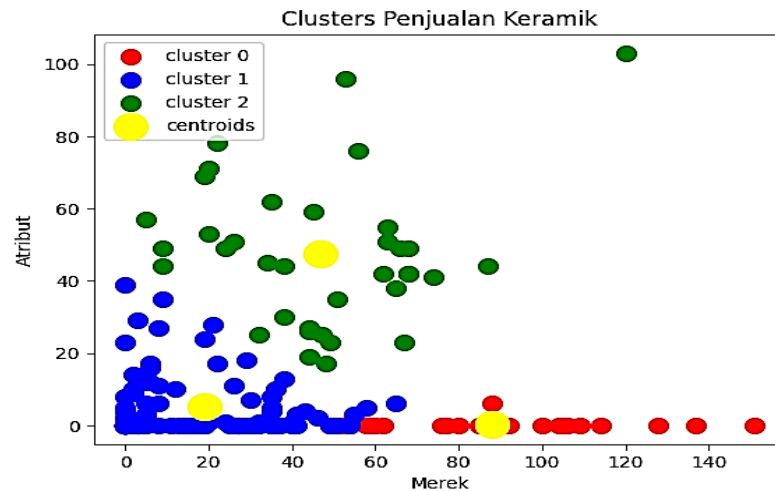
## 3.2 Implementasi dan Hasil

Jika data telah diproses, langkah berikutnya adalah mengimplementasi data ke dalam *python* untuk menghasilkan model hasil. Proses *python* menggunakan *Google Colab* sebagai media. Berikut adalah hasil model dari implementasi data *python* menggunakan *Google Colab* yang terlihat pada gambar 3 dan 4 berikut.



**Gambar 3.** Proses metode *elbow*

Hasil dari metode *elbow* yang diperoleh dari *python* menggunakan *Google Colab* menunjukkan *cluster* atau *k* yang terbaik yaitu  $k = 3$  dikarenakan melihat siku pada grafik titik dimana penurunan nilai kriteria evaluasi menjadi lebih lambat. Menerapkan metode *elbow* untuk mencari jumlah *cluster* yang optimal dari penelitian ini dengan mencatat nilai SSE (*Sum Of Square Error*) yang berubah ketika nilai *k* ditambahkan satu per satu. Keputusan jumlah *cluster* berdasarkan aturan *elbow* yang mana melihat siku pada grafik yang dihasilkan dan titik *elbow*.



Gambar 4. Hasil Pengelompokkan

Hasil pengelompokkan pada gambar 4 diatas menunjukkan pembagian 3 klaster sangat laris, laris dan tidak laris. Pada kluster 2 di kategorikan sangat laris, kluster 1 laris dan kluster 0 tidak laris. Berdasarkan penjualan merek-merek keramik tersebut yang menunjukkan bahwa keramik dengan merek HRM adalah yang sangat laris kluster 2, keramik dengan merek VALENSIA adalah laris kluster 1 dan merek MCC adalah tidak laris kluster 0 berdasarkan data dari Januari 2022 sampai Desember 2022.

### 3.3 Pengujian Data

Pengujian data dilakukan untuk menentukan informasi terkait jumlah klaster yang optimal. Dalam penelitian ini, metode DBI digunakan untuk pengujian data, yang dilakukan dengan menguji nilai jarak antar klaster melalui serangkaian eksperimen mulai dari 2 sampai 6 klaster diperoleh dari *python* menggunakan *Google Colab*. Nilai DBI optimal adalah  $<0$ . Hasil validitas dari penelitian ini adalah 1.013 yang dapat dilihat pada tabel 6 berikut ini dengan beberapa kali percobaan didapatkan hasil optimal penelitian ini.

Tabel 6. Evaluasi DBI

Hasil Davies Bouldin Index
1.013

Tabel 6 hasil validitas penelitian menggunakan evaluasi DBI.

## 4. KESIMPULAN

Setelah melakukan penelitian ditemukan kesimpulan bahwa metode *clustering* dengan metode K-Means dapat digunakan untuk mengelompokkan data penjualan keramik di Riau. Terdapat 9 atribut merek, kode barang (FT, WT) dan ukuran (40x40, 25x25, 50x50, 25x40, 60x60, 20x40). Hasil uji diperoleh dari *python* menggunakan *Google Colab* pada 156 data penjualan keramik di CV.XYZ pada Januari hingga Desember tahun 2022 menghasilkan 3 *cluster* dengan menggunakan metode *elbow* yaitu penjualan yang sangat laris adalah klaster 2, penjualan yang laris adalah klaster 1 dan yang tidak laris adalah klaster 0. Merek yang sangat laris adalah HRM, yang laris adalah VALENSIA dan yang tidak laris adalah MCC. Hasil uji menggunakan metode DBI dengan beberapa kali percobaan didapatkan nilai validitas optimal penelitian ini 1.013 menunjukkan hasil pengujian penelitian menggunakan metode K-Means. Penelitian selanjutnya diharapkan dapat menambahkan atribut dan *cluster* yang berbeda dan lebih bervariasi.

## REFERENCES

- [1] C. Purnama, W. Witanti, and P. N. Sabrina, "KLAUSTERISASI PENJUALAN PAKAIAN UNTUK MENINGKATKAN STRATEGI PENJUALAN BARANG MENGGUNAKAN K-MEANS," *JOINT (Journal of Information Technology)*, vol. 04, no. 01, pp. 35-38, Mar. 2022.

- [2] D. Handoko, H. Satria Tambunan, and J. T. Hardinata, "Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, vol. 6, pp. 111–119, 2021, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [3] M. Richie and dan Halim Agung, "Implementation of K-Means Algorithm For Ceramic Selection and Potential Customer at CV. Jaya Tunggal Keramik," *Jurnal Algoritma, Logika dan Komputasi*, vol. III, no. 2, 2020, doi: 10.30813/j-alu.v2i2.2157.
- [4] W. W. Kristianto and C. Rudianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki)," *JUKANTI (Jurnal Pendidikan Teknologi Informasi)*, vol. 5, no. 2, Nov. 2022.
- [5] J. Jupriyanto and S. Nurlela, "KERANGKA PENGAMBILAN KEPUTUSAN UNTUK PEMASARAN PRESISI MENGGUNAKAN METODE RFM, ALGORITMA K-MEANS DAN DECISION TREE," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 227–234, Sep. 2019, doi: 10.33480/pilar.v15i2.618.
- [6] F. Hadi and Y. Diana, "PENGKLUSTERAN PENJUALAN BAHAN BANGUNAN MENGGUNAKAN ALGORITMA K-MEANS," *JOISIE Journal Of Information System And Informatics Engineering*, vol. 4, no. 1, pp. 22–28, 2020.
- [7] Izzah Lailil and Jananto Arief, "Penerapan Algoritma K-Means Clustering untuk Perencanaan kebutuhan Obat di Klinik Citra," *Jurnal Ilmiah Komputer*, vol.18, no. 1, Feb. 2022.
- [8] Rahmah S Aulia, "KLASTERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI TOKO JUANDA TANI KECAMATAN HUTABAYU RAJA)," *Djtechno : Journal of Information Technology Research*, vol.1, no. 1, 2020.
- [9] Muni Abdul, "Analisis Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Penjualan Sepeda Motor Studi Kasus PT. Alfa Scorpii," *JUTI-UNISI (Jurnal Teknik Industri UNISI)*, vol.4, no. 1, Jun. 2020.
- [10] S. Butsianto and N. T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," 2020.
- [11] R. S. Wicaksana, D. Heksaputra, T. A. Syah, and F. F. Nur'aini, "Pendekatan K-Means Clustering Metode Elbow Pada Analisis Motivasi Pengunjung Festival Halal JHF#2," *Jurnal Ilmiah Ekonomi Islam*, vol. 9, no. 3, p. 4162, Nov. 2023, doi: 10.29040/jiei.v9i3.10591.
- [12] Negara I Setiawan, Purwono and Ashari I Ahmad, "Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means," vol. 7, no. 1, 2018.
- [13] R. Mayang Sari, V. Tasril, and Y. M. Apridonol, "Prediksi Jumlah APBD Kota Payakumbuh dengan metode K-Means," *JURNAL IPTEKS TERAPAN*, 2020, doi: 10.22216/jit.2020.v14i1.5323.
- [14] Indriyani Fintri and Irfiani Eni, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA: Jurnal Informatika*, vol. 7, no. 2, nov. 2019.
- [15] F. Nurdiansyah and I. Akbar, "Implementasi Algoritma K-Means untuk Menentukan Persediaan Barang pada Poultry Shop Article," *Jurnal Teknologi dan Manajemen Informatika*, vol. 7, no. 2, pp. 86–94, 2021, [Online]. Available: <http://http://jurnal.unmer.ac.id/index.php/jtmi>
- [16] A. Wahid, A. Nazir, S. Kurnia Gusti, F. Syafria "Pengelompokan Keberhasilan Produksi Peternak Ayam Broiler di Riau Berdasarkan Index Performance Menggunakan K-Means Clustering," *Techno.COM*, vol. 22, no. 1, Feb. 2023.
- [17] S. W. Harjono *et al.*, "Klasterisasi Tingkat Penjualan pada Startup Panak.id dengan Algoritma K-Means," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 17, no. 1, 2023.
- [18] W. Tities Pambudi and A. Witanti, "PENERAPAN ALGORITMA K-MEANS UNTUK MENGANALISIS DATA PENJUALAN PADA TOKO AYU COLLECTION BERASIS WEB APPLICATION OF K-MEANS ALGORITHM TO ANALYZE SALES DATA AT WEB-BASED AYU COLLECTION STORE," *Jurnal Sistem Informasi Dan Bisnis Cerdas*, vol. 15, no. 1. Februari 2022.
- [19] W. Warisa and N. Nurahman, "Perbandingan Performa Cluster Model Algoritma K-Means Dalam Mengelompokkan Penerima Bantuan Program Keluarga Harapan," *J. Sistem Info. Bisnis*, vol. 13, no. 1, pp. 20–28, Jun. 2023, doi: 10.21456/vol13iss1pp20-28.
- [20] H. Fitriyah, E. M. Safitri, N. Muna, M. Khasanah, D. A. Aprilia, and D. Nurdiansyah, "IMPLEMENTASI ALGORITMA CLUSTERING DENGAN MODIFIKASI METODE ELBOW UNTUK MENDUKUNG STRATEGI PEMERATAAN BANTUAN SOSIAL DI KABUPATEN BOJONEGORO," *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 4, no. 3, pp. 1598–1607, Dec. 2023, doi: 10.46306/lb.v4i3.453.