

# Penerapan Data Mining Dalam Pengelompokan Bahan Sembako Laris Menggunakan K-Means Clustering (Studi Kasus: Toko Gunung Bumi)

Agie Sidik Permana<sup>1</sup>, Rangga Sanjaya<sup>2</sup>

<sup>1,2</sup> Program Studi Sistem Informasi, Universitas Adhirajasa Reswara Sanjaya

e-mail: agiesidik03@gmail.com<sup>1</sup>, rangga@ars.ac.id<sup>2</sup>

## Abstrak

Tujuan penelitian ini adalah untuk pengolahan dan pengkajian data secara cepat dan akurat dan membantu kelompok, maupun perusahaan untuk membuat keputusan yang kritis. Manfaat penelitian ini adalah untuk meningkatkan ke efisien kerja dalam pengelompokan barang. Teori *K-Means Clustering* digunakan sebagai bahan kajian ini. Metode penelitian deskriptif kuantitatif dengan Teknik pengumpulan data Observasi dan wawancara melalui tatap muka. Lokasi penelitian Toko Gunung Bumi jl. Babakan hantap No.540 Bandung Jawa barat. Hasil Penelitian ini adalah didapatkan 2 iterasi dengan nilai centroid C0 tidak laris 397.950 dan laris 1.248.300, dan centroid C1 tak laris 349.400 dan laris 1.272.700 dengan davies bolden 0.604.

**Kata kunci** : Pengelompokan, K-Means Clustering, Toko Gunung Bumi

## Abstract

The purpose of this research is to management and review of data quickly and accurately and helps groups, as well as companies to make critical decisions. The benefit of this research is to improve work efficiency in grouping goods. The K-Means Clustering theory is used as the material for this study. Descriptive quantitative research method with data collection techniques Observation and interviews through face to face. Research location Toko Gunung Bumi jl. Babakan hantap No. 540 Bandung, West Java. The results of this research are obtained 2 iterations with a value of C0 centroid not selling well 397.950 and selling 1.248.300, and centroid C1 not selling 349.400 and selling 1.272.700 with davies bolden 0.604.

**Keywords** : Clustering, K-Means Clustering, Earth Mountain Stores

## PENDAHULUAN

Perkembangan teknologi informasi pada saat ini menjadi salah satu faktor pendorong perkembangan dalam pembangunan dibidang bisnis ritel (Risdianto, 2019). Perkembangan teknologi informasi saat ini memudahkan para pekerja untuk melakukan pekerjaannya dengan efisien dan cepat. Teknologi sangat dibutuhkan dalam segala bidang, dikarenakan banyaknya data yang diolah secara konvensional akan memakan waktu yang lama dalam pengerjaannya (Desy & Sihombing, 2020). Oleh karena itu dibutuhkan sistem informasi yang terkomputerisasi dalam menunjang proses pengolahan data dan informasi sesuai kebutuhan dari proses-proses yang digunakan dalam bidang bisnis seperti ritel atau toko swalayan

Ritel merupakan salah satu bidang bisnis yang aktivitas utamanya adalah jual beli (Prasetyo, 2020). Bisnis ritel seperti mini market dan swalayan masih banyak yang melakukan proses bisnisnya secara konvensional, sehingga sering terjadi kesalahan dalam pencatatan data dan pengerjaan yang lambat (Chaniago, 2021). Toko Gunung Bumi merupakan salah satu perusahaan ritel yang saat ini masih menggunakan pemrosesan data

secara manual karena itu sering kesalahan dalam pencatatan data penjualan, sehingga kesulitan yang sering dialami di Toko Gunung Bumi sering sekali kekurangan stok barang dan menumpuknya stok barang yang ada di gudang

*K-Means Clustering* merupakan metode pengelompokan yang bersifat unsupervised learning (Mario et al., 2016). Unsupervised learning merupakan metode machine learning yang berguna untuk menarik kesimpulan dari data training (Nabila et al., 2021)[6]. *K-Means* merupakan metode analisis data atau metode data mining yang berfungsi untuk mengelompokkan data ke dalam cluster (Handoko, 2016).

## METODE

Pada tahapan ini menjelaskan gambaran metode penelitian secara menyeluruh. Berikut adalah langkah-langkah yang bisa di lihat pada gambar 1.



**Gambar 1 Tahapan Penelitian**

### **Tahapan penelitian**

1. Identifikasi masalah  
Bagaimana penerapan algoritma *K-Means Clustering* untuk identifikasi pengelompokan barang laris di Toko Gunung Bumi
2. Menganalisa masalah  
Toko Gunung Bumi sering sekali melakukan pembelian barang yang berlebih sehingga menyebabkan penumpukan barang dagang di Gudang.
3. Mempelajari Literatur  
Studi literatur suatu proses yang dilakukan untuk mengumpulkan artikel-artikel yang digunakan untuk penelitian ini.
4. Mengumpulkan data  
Data penjualan yang berasal dari Toko Gunung Bumi. Dalam penelitian ini bahan penelitian yang digunakan untuk kemudian diolah menjadi acuan adalah penjualan barang pada Toko Gunung Bumi Bandung dari Januari-juni 2022 sebagai objek penelitian
5. Melakukan Preprocessing dan Transformasi data  
Dalam tahap ini data akan diolah menggunakan Teknik *preprocessing* dan *transformasi* supaya data dapat digunakan pada penelitian ini. Data penjualan 30 barang dan akan diolah lagi untuk siap diolah ke dalam algoritma *K-Means Clustering*.
6. Penerapan *K-Means Clustering*  
Langkah yang dilakukan untuk membentuk clustering pada metode *K-Means*: (Indriyani & Irfiani, 2019)
  - a) Menentukan banyak nya cluster ( $k$ ) untuk jumlah cluster dari dataset yang ada.

- b) Menentukan k sebagai cetroid, biasanya dilakukan secara acak (random).
  - c) Hitung jarak data dengan cetroid menggunakan rumus jarak menggunakan rumus Euclidean Distance.
  - d) Kelompokan data berdasarkan kedekatan dengan centroid kemudian perbaruan nilai centroid menggunakan persamaan 2sj.
  - e) Lakukan langkah 2 sampai 4 sampai anggota tiap sluter tidak ada yang berubah.
7. Menguji Hasil  
Pengujian hasil menjelaskan implementasi k-means clustering software yang akan digunakan dipenelitian ini.
8. Penyimpulan hasil  
Menyimpulkan hasil menjelaskan tentang hasil keluar dari pemrosesan data menggunakan *k-means clustering*

### **Pengumpulan Data**

Data penelitian inin di kumpulkan dengan menggunakan metode (observasi) di Toko Gunung Bumi Bandung.

### **Algoritma K-Means**

Langkah yang dilakukan untuk membentuk clustering pada metode K-Means: (Indriyani & Irfiani, 2019)

- 1. Menentukan banyak nya cluster (k) untuk jumlah cluster dari dataset yang ada.
- 2. Menentukan k sebagai cetroid, biasanya dilakukan secara acak (random).
- 3. Hitung jarak data dengan cetroid menggunakan rumus jarak menggunakan rumus Euclidean Distance:

$$d(x_i, \mu_j) = \sqrt{\sum(x_i, \mu_j)^2}$$

**Gambar 2 Rumus Eulidean untuk hitun jarak data**

Dimana d merupakan titik dokumen, xi merupakan data kriteria dan  $\mu_j$  merupakan centroid pada cluster ke-j.

- 4) Kelompokan data berdasarkan kedekatan dengan centroid kemudian perbaruan nilai centroid menggunakan persamaan 2 :

$$\mu_j(t + 1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j$$

**Gambar 3 Rumus untuk menentukan jarak minimum**

Dimana  $\mu_j(t+1)$  merupakan centroid baru pada iterasi ke (t+1) dan  $N_{sj}$  merupakan banyak data pada cluster sj.

- 5) Lakukan langkah 2 sampai 4 sampai anggota tiap sluter tidak ada yang berubah.

## **HASIL DAN PEMBAHASAN**

### **Perhitungan manual algoritma K-Means**

Pada tahapan ini penetapan jumlah cluster k pada penelitian inin yaitu 2 cluster centroid awal yang di gunakan bisa di lihat di tabel 1.

C0	C1
516	747
579	985

**Tabel 1 Centroid Awal**

Setelah didapat kan centroid awal, maka Langkah selanjutnya yakni menghitung kemiripan setiap dokumen terhadap centroid menggunakan rumus Euclidean Distance sebagai berikut:

Perhitungan (C0) pada ABC KECAP ASIN KCL PLST 140/133ML dengan titik pusat sebagai berikut:

$$D1 = \sqrt{(516-516)^2 + (747-747)^2} = 0$$

$$D2 = \sqrt{(579-516)^2 + (985-747)^2} = 246,1971$$

Perhitungan (C1) pada ABC KECAP ASIN KCL PLST 140/133ML dengan titik pusat sebagai berikut:

$$D1 = \sqrt{(516-579)^2 + (747-985)^2} = 246,1971$$

$$D2 = \sqrt{(579-579)^2 + (985-985)^2} = 0$$

Kemudian Iterasi kembali untuk menentukan posisi centroid baru dengan cara hitung rata-rata data yang ada pada centroid yang sama. Dibawah ini perhitungan centroid baru:

$$C0(P1) = \frac{514+45+161+207+219+271+366+310+363+336+585+494+1037+620+469+533+642+821+353+20+336}{20} = 420.850$$

$$C0(P2) = \frac{747+445+299+141+19+141+301+306+259+605+310+711+640+440+301+133+222+391+18+284}{20} = 335.700$$

$$C1(P1) = \frac{579+1288+996+1010+1046+1433+2049+916+1268+1440}{10} = 1202.500$$

$$C1(P2) = \frac{985+1073+902+859+1448+1572+2569+1266+835+1492}{10} = 1300.100$$

Setelah melakukan perhitungan centroid didapatkan nilai centroid baru. karena itu kita akan melakukan kembali perhitungan dari langkan 2 sampai 4 seperti yang tertera dibab 2 dengan centroid baru untuk Iterasi 2 pada.

C1	C2
420.850	1202.500
335.700	1300.100

**Tabel 2 Tabel Centroid baru**

Lakukan penghitungan Euclidean distan dengan Centroid baru sampai menghasilkan pola yang sama dari iterasi sebelumnya dengan yang terakhir. Pada proses ini iterasi ke enam akan menjadi Nilai centroid baru terakhir dan menghasilkan jumlah data:

C0=(1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 21, 22, 23, 25, 27, 28, 29) sebanyak 20 anggota.

C1=(4, 15, 16, 17, 18, 19, 20, 24, 26, 30) sebanyak 10 anggota

C0	C1
397.950	1248.300
349.400	1272.700

**Tabel 3 Centroid akhir hasil Iterasi ke 6**

### **implementasi Software**

Berdasarkan validasi hasil pada *RapidMiner* menggunakan davis blouldin index, terbentuk dua cluster yang di anggap mampu mengelompokan data dengan baik. Seperti pada gambar 4 dengan performance vector, yang merupakan hasil evaluasi dari cluster yang di buat sebanyak 2 cluster dan memiliki nilai 0,604.



**Gambar 4 Performance vector Davis Boulding Index**

Dengan hasil cluster 0 20 barang kurang laris dan cluster 1 dengan 10 barang laris, seperti yang di tunjukkan pada gambar 5.

### Cluster Model

```
Cluster 0: 20 items  
Cluster 1: 10 items  
Total number of items: 30
```

**Gambar 5 Hasil Pemodelan Cluster**

### SIMPULAN

Berdasarkan hasil pemodelan menggunakan algoritma *K-Means Clustering* dengan menggunakan data yang sudah diolah menggunakan *Knowledge Discovery in Database(KDD)*. Hasil perhitungan mendapatkan 2 kategori penjualan sembako, yaitu tidak laris dan laris dengan penentuan nilai centroid secara random dengan rumus Euclidean Distance. Maka didapat kan 2 Iterasi dengan nilai rata-rata C0 397.950 dan 349.400, dan rata-rata C1 1.248.300 dan 1.272.700 dengan hasil C0 dengan 20 barang tidak laris dan C1 10 barang laris dan hasil dari evaluasi dari cluster yang dibuat sebanyak 2 cluster dan memiliki nilai 0,605. maka penelitian ini sangat cocok digunakan dalam penyetokan barang guna meningkatkan penjualan di Toko Gunung Bumi.

### DAFTAR PUSTAKA

- Risdianto, E. (2019). Analisis Pendidikan Indonesia di Era Revolusi Industri 4.0. *Research Gate, April*(January), 1–16.
- Desy, E., & Sihombing, C. (2020). *Vol . 4 No . 3 Agustus 2020 PENERAPAN TEKNOLOGI ACTIVEX DATA OBJECT . NET ( ADO . NET ) PADA SISTEM INFORMASI ADMINISTRASI PEMBAYARAN SPP SEKOLAH JISAMAR ( Journal of Information System , Applied , Management , Accounting and Researh ) p-ISSN : 2598-8700 . 4(3), 80–87.*
- Prasetyo, B. (2020). Penerapan Point Of Sales dalam Peningkatan Customer Satisfaction Pada Bisnis Ritel Modern. *Jesya (Jurnal Ekonomi & Ekonomi Syariah)*, 3(2), 389–395. <https://doi.org/10.36778/jesya.v3i2.219>
- Chaniago, H. (2021). Manajemen Ritel & Implementasinya. In *Bandung: Edukasi Riset Digital* (Issue November). [https://www.researchgate.net/profile/Harmon-Chaniago/publication/356192834\\_18\\_BUKU-MANAJEMEN\\_RITEL-HARMON-/links/6191291c3068c54fa5e55f33/18-BUKU-MANAJEMEN-RITEL-HARMON.pdf](https://www.researchgate.net/profile/Harmon-Chaniago/publication/356192834_18_BUKU-MANAJEMEN_RITEL-HARMON-/links/6191291c3068c54fa5e55f33/18-BUKU-MANAJEMEN-RITEL-HARMON.pdf)
- Mario, A., Herry, S., & Nasution, H. (2016). Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness. *Jurnal Sistem Dan Teknologi Informasi*, 1(1), 1–6
- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTISI>
- Handoko, K. (2016). Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di

Program Studi Tkj Akademi Komunitas Solok Selatan). *Jurnal Teknologi Dan Sistem Informasi*, 02(03), 31–40. <http://teknosi.fti.unand.id/index.php/teknosi/article/view/70>  
Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA: Jurnal Informatika*, 7(2), 109. <https://doi.org/10.30595/juita.v7i2.5529>.