**JURNAL INFORMATIKA (JUNIKA)**
ISSN xxxx - xxxx
Volume 1 Number 2, Oktober 2024
https://ejurnal.uniyos.ac.id/index.php/yij-server

# Review Paper
# Data Mining Metode Clustering Menggunakan Algoritma K-Means

**Muhammad Alfin Fikri** [(1)]
Teknik Informatika, Universitas Yos Soedarso, fikrialfin@gmail.com

**Adriana Male** [(2)]
Teknik Informatika, Universitas Yos Soedarso;adrianamale1996@gmail.com

***ABSTRACT***

*Clustering or clustering is a method of grouping data Clustering is the process of partitioning a set of data objects into subsets called clusters. The main aim of the cluster method is to group a number of data/objects into clusters (groups) so that each cluster contains data that is as similar as possible (Bahauddin, A., Fatmawati, A., & Sari, F. P., 2021). K-Means is a non-hierarchical data clustering method that attempts to partition existing data into one or more clusters/groups. This method partitions data into clusters/groups so that data that has the same characteristics is grouped into the same cluster and data that has different characteristics is grouped into another group.*

***Keywords****: Data mining; K-Means; Clustering.*

**ABSTRAK**

Clustering atau klasterisasi adalah metode pengelompokan data. Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Tujuan utama dari metode klaster adalah pengelompokkan sejumlah data/obyek keldalm klaster (grup) sehingga dalam setiap klaster akan berisi data yang semirip mungkin (Bahauddin, A., Fatmawati, A., & Sari, F. P., 2021). K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

**Kata kunci**: Data mining; K-Means; Clustering.

## PENDAHULUAN

Perkembangan kecanggihan teknologi yang semakin pesat merupakan aspek yang dapat dimanfaatkan untuk mencapai kemudahan-kemudahan, tidak terkecuali dalam arus informasi. Semakin besar suatu perusahaan, tentunya semakin besar data yang dimiliki. Semua data tersebut biasanya akan tersimpan dalam database center. Namun, banyaknya perusahaan yang tidak menyadari betapa berharganya tumpukan data-data lama yang dihasilkan perusahaan dalam bertransaksi. Clustering adalah teknik untuk mengelompokkan data berdasarkan kesamaan maupun perbedaan yang ada di antara data tersebut. Dimana data dalam satu kelompok memiliki karakteristik yang mirip dan sekaligus berbeda dengan data kelompok lain. Tujuan adanya clustering yaitu untuk membagi kumpulan data menjadi kelompok-kelompok (kluster). Data Clustering merupakan salah satu metode Data Mining yang bersifat tanpa arahan (unsupervised). K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan objective function yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalisasikan variasi antar cluster.

JURNAL INFORMATIKA (JUNIKA)
ISSN xxxx - xxxx
Volume 1 Number 2, Oktober 2024
https://ejurnal.uniyos.ac.id/index.php/yij-server

## METODE

Pada algoritma ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya Model yang digunakan untuk menyelesaikan penelitian ini adalah dengan menggunakan algoritma K-Means. K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

A. Algoritma K-Means

Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap cluster terdapat titik pusat (centroid) yang merepresentasikan cluster tersebut.

Data clustering menggunakan metode K-Means ini secara umum dilakukan dengan algoritma dasar sebagai berikut (Yudi Agusta, 2007) :

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung centroid/ rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke centroid/ rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atauapabila perubahan nilai pada objective function yang digunakan di atas nilai hreshold yang ditentukan.

$$D_{L_1}(x_2, x_1) = \|x_2 \updownarrow x_1\|_1 = \sum_{j=1}^{p} |x_{2j} \updownarrow x_{1j}|$$

dimana:

$p$ : Dimensi data

$|.|$ : Nilai absolut

Sedangkan untuk $L_2$ (*Euclidean*) *distance space*, jarak antara dua titik dihitung menggunakan rumus sebagai berikut[3]:

$$D_{L_2}(x_2, x_1) = \|x_2 \updownarrow x_1\|_2 = \sqrt{\sum_{j=1}^{p} (x_{2j} \updownarrow x_{1j})^2}$$

dimana:

p : Dimensi data

Lp (Minkowski) distance space yang merupakan generalisasi dari beberapa distance space yang ada seperti L1 (Manhattan/City Block) dan L2 (Euclidean), juga telah diimplementasikan

Pembaharuan suatu titik centroid dapat dilakukan dengan rumus berikut:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

Di mana:

μk = titik centroid dari cluster ke-K

Nk = banyaknya data pada cluster ke-K

xq = data ke-q pada cluster ke-K

**JURNAL INFORMATIKA (JUNIKA)**
ISSN xxxx - xxxx
Volume 1 Number 2, Oktober 2024
https://ejurnal.uniyos.ac.id/index.php/yij-server

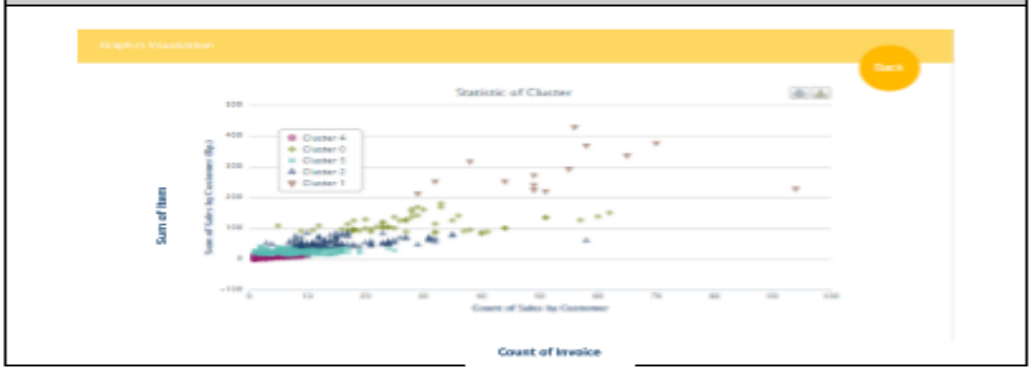
## HASIL

Hasil dari pengujian klastering yang telah diuji di program bahasa R pada swalayan. Hasil pengujian adalah sebagai berikut :

| Jenis Data | Data Customer | | | |
|---|---|---|---|---|
| Tahun Data | 2013 s/d 2014 | | | |
| Total Customer | 354 Record | | | |
| Jmlh Customer thdp Transaksi | 349 Record | | | |
| Jumlah Cluster | 5 | | | |
| Proses Clustering (Waktu) | ± 22 Menit | | | |
| Jumlah Iterasi | 29 Iterasi | | | |
| Detail Cluster | | | | |
| Urutan Cluster | Jumlah Customer | Centroid 1 | Centroid 2 | Urutan |
| Cluster 0 | 3 | 166.333 | 2,383,598,386.667 | 1 |
| Cluster 1 | 19 | 42.421 | 387,833,885.684 | 3 |
| Cluster 2 | 73 | 14.082 | 140,012,998.301 | 4 |
| Cluster 3 | 245 | 4.188 | 28,546,621.094 | 5 |

Gambar 1. Tabel Hasil uji Data Costumer

| Jenis Data | Data Product | | | |
|---|---|---|---|---|
| Tahun Data | 2013 s/d 2014 | | | |
| Total Product | 1.309 Record | | | |
| Jmlh Product thdp Transaksi | 1.285 Record | | | |
| Jumlah Cluster | 5 | | | |
| Proses Clustering (Waktu) | ± 41 Menit | | | |
| Jumlah Iterasi | 28 Iterasi | | | |
| Detail Cluster | | | | |
| Urutan Cluster | Jumlah Product | Centroid 1 | Centroid 2 | Urutan |
| Cluster 0 | 55 | 28.491 | 114.418 | 2 |
| Cluster 1 | 14 | 52.786 | 286.857 | 1 |
| Cluster 2 | 105 | 16.257 | 54.952 | 3 |
| Cluster 3 | 273 | 7.44 | 23.168 | 4 |
| Cluster 4 | 838 | 2.058 | 5.525 | 5 |
| Visualisasi | | | | |




Gambar 2. Gambar Penyebaran Cetroid

Dari sini kita bisa tahu bahwa penyebaran titik centroid berbeda beda setiap costumer. Semakin banyak costumer setiap waktunya maka semakin beda pula tiap titik tengah atau centroid dari visual tiap data yang akan di visualsisasikan. Dari gambar 2 kita bisa tau bahwa penyebaran terbanyak adalah diCluster 4 dimana cluster 4 mempunya penjualan tertinggi dalam periode tertentu

JURNAL INFORMATIKA (JUNIKA)
ISSN xxxx - xxxx
Volume 1 Number 2, Oktober 2024
https://ejurnal.uniyos.ac.id/index.php/yij-server

## KESIMPULAN

Konsep data mining mengunakan K-Means guna mengelompokkan data produk dan data pelanggan untuk mengetahui data yang memiliki potensi dengan melakukan proses perhitungan jumlah invoice terhadap produk sedangkan total penjualan terhadap pelanggan dan jumlah transaksi terhadap pelanggan untuk menggelompokkan data pelanggan.

Hasil ini dapat digunakan untuk memberikan saran pertimbangan dalam menentukan stratregi penjualan yaitu mengeliminasi produk dengan posisi cluster terbawah dan memberikan reward untuk pelanggan dengan posisi cluster teratas.

## DAFTAR PUSTAKA

1. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996).From data mining to knowledge discovery in databases.AI magazine
2. Freitas, A. A. (2013). Data mining and knowledge discovery with evolutionary algorithms.Springer Science & Business Media.
3. Liautaud, B., & Hammond, M. (2000).e-Business intelligence: turning information into knowledge into profit. McGraw-Hill, Inc..
4. Maimon, O., &Rokach, L. (Eds.).(2005). Data mining and knowledge discovery handbook New York: Springer
5. Ong, J. O. (2013). Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University.
6. Kusrini dan E.T. Luthfi., 2009, Algoritma Data Mining, Andi, Yogyakarta.