Jurnal TRANSFORMATIKA
Vol. 22, No.2, Januari 2025, pp. 115 - 120
P-ISSN: 1693-3656, E-ISSN: 2460-6731
https://journals.usm.ac.id/index.php/transformatika/

Optimasi Clustering *K-Means* Menggunakan *Algoritma Genetika* Dengan Data View Dan Like Di Tiktok

Galet Guntoro Setiaji[1*], Krida Pandu Gunata[2], Galih Setiarso[3]
[1]Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, Semarang, telp:024-6702757, e-mail: gallet@usm.ac.id
[2]Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, Semarang, telp:024-6702757, e-mail: galih@usm.ac.id
[3]Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, Semarang, telp:024-6702757, e-mail: kridapandu@usm.ac.id

ARTICLE INFO

History of the article :
Received 22 November 2024
Received in revised form 24 November 2024
Accepted 12 January 2025
Available online 29 January 2025

Keywords:

Clustering; K-Means; Algoritma Genetika; Davis Boulden Index

*** Correspondece:**

Telepon:
+62 85713050403

E-mail:
gallet@usm.ac.id

ABSTRACT

K-Means merupakan algoritma yang sering digunakan untuk melakukan pengelompokkan atau sering juga disebut clustering. Dengan menentukan pusat centroid awal secara random pada algoritma K-Means akan ditingkatkan performanya menggunakan *Algoritma Genetika* (*GA*). Menggunakan data set publik di Kaglle, berupa data set tiktok dimana jumlah view dan like dengan record data sebanyak ***19.084*** setelah dilakukan pembersih data. Yang akan diuji dengan melakukan performa clustering K-Means dengan Algoritma Genetika. Dan untuk validitas nya nanti menggunakan *Davis Boulden Index*, dimana hasil validitas *DBI* ini nanti akan meningkatkan performance *K-Means* dengan menambahkan Algoritma Genetika.
Dengan pengujian *K-Means* dengan jumlah *k*=3, *k*=4 dan *k*=5 menghasilkan masing-masing validitas DBI ***0,64*** ; ***0,79*** dan ***0,72***. Sedangkan untuk algoritma *K-Means* dengan peningkatan performa menggunakan *GA* didapatkan validitas dengan masing-masing *DBI* sebagai berikut ***0,45 ; 0,40*** dan ***0,60***. Dengan hasil penelitian menghasilkan bahwa peningkatan performa *K-Means* dengan menggunakan *GA* memberikan hasil validitas lebih kecil dari pada hanya menggunakan perhitungan *K-Means* saja.

1. PENDAHULUAN

Clustering merupakan metode pengelompokkan yang sering digunakan dalam pengolahan data mining, dimana algoritma yang sering digunakan untuk clustering adalah metode K-Means. Penelitian terkait kenapa K-Means lebih unggul diantaranya penelitian terkait komparasi K-Means dengan Fuzzy C-Means dimana validitas DBI K-Means lebih unggul [1]. Selanjutnya dipenilitan terkait komparasi metode K-Means dengan K-Medoids

dimana validitas DBI K-Means lebih unggul dari K-Medoids [2]. Sesuai dengan roadmap penelitian sebelumnya, akan meneruskan dengan optimasi metode *K-Means* menggunakan *Algoritma Genetika*.

Dimana kelemahan pada metode *K-Means* salah satunya adalah penentuan centroid awal yang masih random [3]. Dengan begitu kelemahan pada *K-Means* ini akan kita coba mengoptimasikan dengan menambahkan *Algoritma Genetika*. Dimana penelitian terkait optimasi menggunakan Algoritma Genetika untuk optimasi pelayanan kependudukan [4].Untuk mendukung penelitian kali ini, data publik yang kita gunakan merupakan data view tiktok sejumlah *19.084* yang ada di *https://www.kaggle.com/datasets*.

2. METODE PENELITIAN

Untuk metode pengelompokkan atau cluster dalam penelitian kali ini menggunakan *K-Means* dan optimasi menggunakan *Algoritma Genetika*.

2.1. *K-Means*

K-Means merupakan algoritma yang sering digunakan untuk pengelompokkan atau cluster. Berikut ini tahapan untuk algoritma K-Means [5].

1. Melakukan inisialisasi jumlah cluster (k), menentukan pusat centroid awal secara acak.
2. Mencari pusat cluster terdekat dengan rumus Euclidean Distance.

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - c_{jk})^2} \quad (1)$$

3. Melakukan perhitungan ulang pusat cluster.

$$c_j = \frac{1}{|S_j|}\sum_{x_i \in S_j} x_i \quad (2)$$

4. Melakukan pengulangan tahap 2 dan 3, hingga tidak ada perubahan pusat cluster atau jumlah interasi maksimum sudah tercapai.

2.2. *Genetika Algoritma*

Genetika Algoritma (GA) digunakan untuk melakukan perhitungan optimasi [6], berikut ini tahapannya.

1. Inisialisasi populasi awal dengan menentukan beberapa parameter.
2. Melakukan perhitungan evaluasi fitnes populasi.

Maksimalisasi: $f(x) = Objektif\ Function$ (3)

Normalisasi probabilitas fitnes: $P_i = \frac{f(x_i)}{\sum_{j=1}^{N} f(x_j)}$ (4)

3. Memilih nilai fitnes individu tertingi.
4. Melakukan crossover atau pengabungan 2 individu dari fitnes tertingi tadi untuk menghasilkan nilai individu baru.

$$\text{Child } 1 = \alpha.x_1 + (1-\alpha).x_2 \quad (5)$$

5. Mutasi dengan memodifikasi individu secara acak.

$$x^1 = x + \in, \in \sim N(0, \sigma^2) \quad (6)$$

6. Melakukan perhitungan ulang fitnes baru.
7. Mengecek apakah telah mencapai kondisi yang telah ditentukan
8. Output dari solusi yang dihasilkan.

2.3. *Python* dengan *PyGad*

Peneliti dalam melakukan perhitungan algoritma *K-Means* dan optimasi menggunakan *Algoritma Genetika*. Peneliti menggunakan bahasa pemrograman *Python* sebagai tools untuk membantu perhitungan *machine learning* [7], dan menggunakan library *PyGad* untuk melakukan perhitungan *Genetika Algoritma*.

2.4. Optimasi *K-Means* dengan *Algoritma Genetika*

Tahapan akhir dari penelitian ini adalah menggunakan optimasi *Algoritma Genetika* pada *K-Means* untuk menentukan perhitungan awal yaitu menentukan pusat *centroid* awal pada *K-Means*. Tahap selanjutnya menghitung masing-masing validasi dari perhitungan *K-Means* dan *optimasi K-Means GA* dengan menggunakan validitas *DBI*.

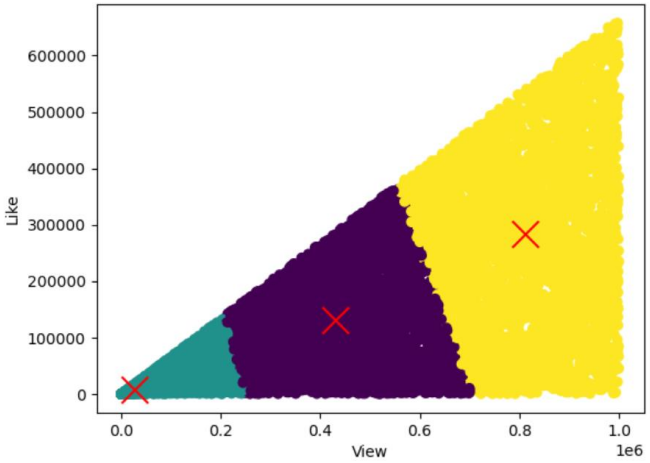
3. HASIL DAN PEMBAHASAN

Berikut ini hasil dan pembahasan penelitian menggunakan *Algoritma K-Means* dan melakukan *optimasi K-Means* dengan *GA*.

3.1. Hasil K-Means

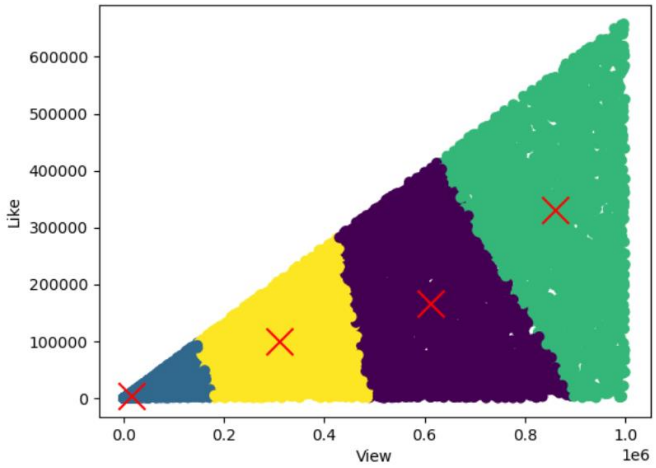
Berikut ini hasil dari penelitian menggunakan algoritma *K-Means* dengan masing-masing *k*=3, *k*=4 dan *k*=5.

a. Untuk *k*=3 menghasilkan nilai validitas *DBI* sebesar ***0,64***, sedangkan untuk jumlah kluster didapatkan jumlah 3.746 untuk *k*=1, untuk *k*=2 berjumlah 3.645 sedangkan *k*=3 berjumlah 11.693.



Gambar 1. Grafik *K-Means* dengan 3 Kluster

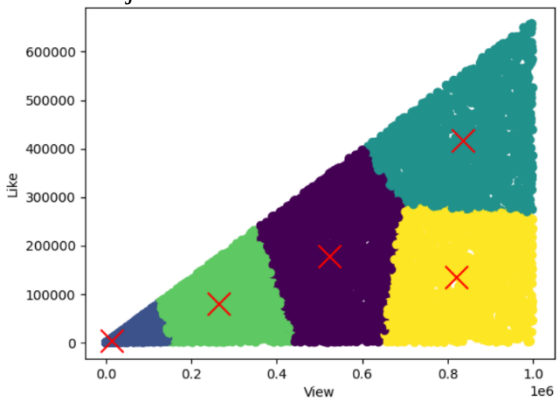
b. Untuk *k*=4 menghasilkan nilai validitas *DBI* sebesar ***0,79***, untuk kluster *k*=1 berjumlah 2.758, *k*=2 berjumlah 11.098, *k*=3 berjumlah 2.487 dan *k*=4 berjumlah 2.741.



Gambar 2. Grafik *K-Means* dengan 4 Kluster

DOI : https://doi.org/ 10.26623/y2tedy77

c. Untuk *k*=5 menghasilkan nilai validitas *DBI* sebesar ***0,72***, didapatkan jumlah kluster untuk *k*=1 berjumlah 2.520, *k*=2 berjumlah 10.835, *k*=3 berjumlah 1.691, *k*=4 berjumlah 2.388 dan untuk *k*=5 berjumlah 1.650.

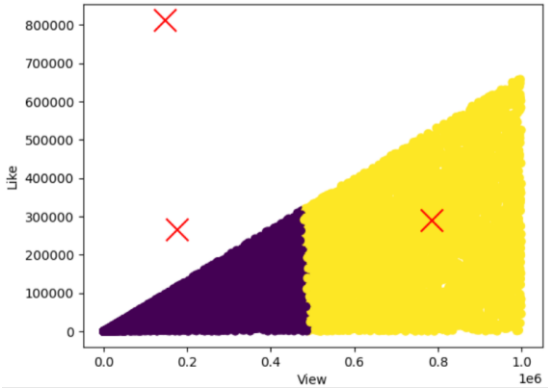


Gambar 3. Grafik K-Means dengan 5 Kluster

3.2. Optimasi *K-Means* menggunakan *GA*

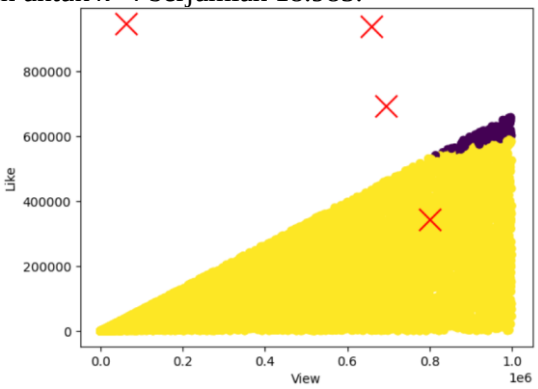
Dan untuk optimasi *K-Means* menggunakan *Genetika Algoritma* didapatkan hasil sebagai berikut.

a. Untuk *k=3* optimasi *K-Means* dengan *GA* menghasilkan nilai validitas *DBI* sebesar ***0,45***. Dengan hasil cluster *0 = 14092* dan *2= 4992*, dan tidak ada data di kluster 1.



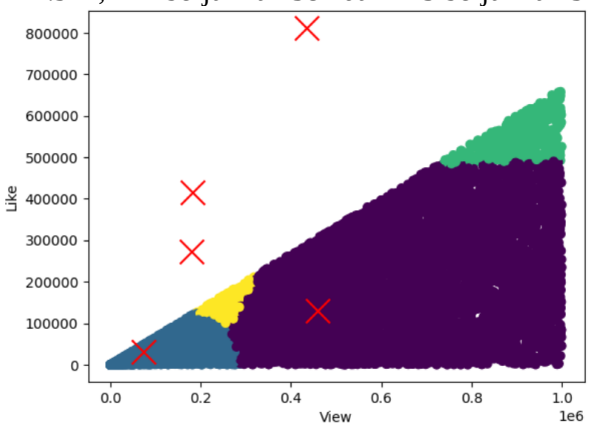
Gambar 4. Grafik optimasi *K-Means GA* dengan 3 kluster

b. Untuk *k*=4 optimasi *K-Means* dengan *GA* menghasilkan nilai validitas *DBI* sebesar ***0,40***. Untuk kluster 2 dan 3 menghasilkan jumlah data kosong, sedangkan kluster *k*=1 berjumlah 99 dan untuk *k*=4 berjumlah 18.985.



Gambar 5. Grafik optimasi *K-Means GA* dengan 4 kluster

c. Untuk *k=5* optimasi *K-Means* dengan *GA* menghasilkan nilai validitas *DBI* sebesar ***0,60***. Hasil kluster 1 menghasilkan nilai kosong, namun untuk kluster *k*=2 berjumlah 6.480, *k*=3 berjumlah 11.911, *k*=4 berjumlah 387 dan *k*=5 berjumlah 306.



Gambar 6. Grafik optimasi *K-Means GA* dengan 5 kluster

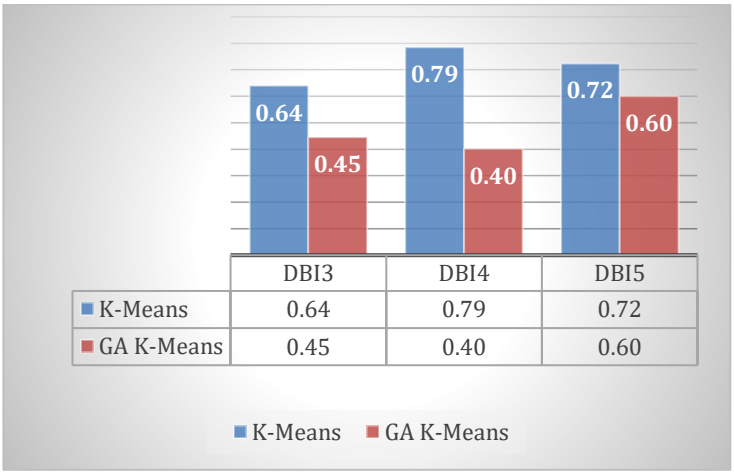
3.3. Pembahasan

Berikut ini hasil perbandingan *validasi DBI* metode *K-Means* dengan menggunakan optimasi *GA* pada *K-Means*.

Tabel 1. Hasil validitas *DBI* dari perhitungan *K-Means* dan *Optimasi K-Means GA*

Penelitian	***DBI k3***	***DBI k4***	***DBI k5***
K-Means	0,64	0,79	0,72
GA K-Means	0,45	0,40	0,60

Dibawah ini grafik dari hasil tabel 3.1, dimana terlihat validitas DBI optimasi K-Means dengan GA lebih kecil.



Gambar 7. Grafik perbandingan *DBI K-Means* dan *Optimasi K-Means GA*

DOI : https://doi.org/ 10.26623/y2tedy77

4. KESIMPULAN

Dari hasil penelitian terhadap peningkatan performa *K-Means* menggunakan *Genetika Algoritma,* didapatkan bahwa hasil mengolah data view tiktok menggunakan *K-Means* sederhana kurang bagus nilai validitasnya. Dimana nilai *DBI* untuk *K-Means* saja menghasilkan DBI ***0,64*** untuk *k=3,* ***0,79*** dengan *k=4* dan ***0,72*** dengan *k=5*. Sedangkan menggunakan performa *Genetika Algoritma* untuk mengoptimasi *K-Means*, didapatkan validitas *DBI* lebih bagus nilainya yaitu *DBI* ***0,45*** untuk *k=3,* ***0,40*** untuk *k=4* dan ***0,60*** untuk *k=5*.

REFERENCES

[1] G. Guntoro Setiaji and V. Vydia, "KOMPARASI METODE CLUSTERING K-MEANS DAN FUZZY C-MEANS UNTUK MEMPREDEKSI KETEPATAN WAKTU LULUS," May 2019. [Online]. Available: http://journals.usm.ac.id/index.php/jprt/index

[2] G. Guntoro Setiaji, A. Novita Putri, and D. Anggit Wicaksana, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Clustering Harga Beras di Provinsi Jawa Tengah," *Transformatika*, vol. 22, no. 1, pp. 39–45, 2024, doi: 10.26623/transformatika.v.

[3] A. Fauzi Sistem Informasi, F. H. Universitas Buana Perjuangan Karawang Jl Ronggowaluyo, T. Timur, and K. priati, *Data Mining dengan Teknik Clustering Menggunakan Algoritma K-Means pada Data Transaksi Superstore*. Seminar Nasional Informatika dan Aplikasinya (SNIA), 2017. [Online]. Available: http://community.tableau.com.

[4] S. F. Pane, R. Maulana Awangga, E. V. Rahcmadani, and S. Permana, "IMPLEMENTASI ALGORITMA GENETIKA UNTUK OPTIMALISASI PELAYANAN KEPENDUDUKAN," *Jurnal Tekno Insentif*, vol. 13, no. 2, pp. 36–43, Oct. 2019, doi: 10.36787/jti.v13i2.130.

[5] A. Rifa'i, G. Guntoro Setiaji, and V. Vydia, "PENGGUNAAN METODE K-MEANS PADA ANALISA DAN KLASIFIKASI CAPRES 2019 DI TWITTER," *Pengembangan Rekayasa dan Teknologi*, vol. 15, no. 1, pp. 43–47, 2019, [Online]. Available: http://journals.usm.ac.id/index.php/jprt/index

[6] S. Mauluddin, I. Ikbal, and A. Nursikuwagus, "Optimasi Aplikasi Penjadwalan Kuliah Menggunakan Algoritma Genetik," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 2, no. 3, pp. 792–799, 2018, [Online]. Available: http://jurnal.iaii.or.id

[7] M. Riziq sirfatullah Alfarizi, M. Zidan Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, "PENGGUNAAN PYTHON SEBAGAI BAHASA PEMROGRAMAN UNTUK MACHINE LEARNING DAN DEEP LEARNING," 2023.