Prosiding Seminar Nasional Sains dan Teknologi Seri 02
Fakultas Sains dan Teknologi, Universitas Terbuka
Vol. 1 No. 2 (2024)
e-ISSN: 3047-6569

PERBANDINGAN KINERJA ALGORITMA CLUSTERING K-MEANS DAN K-MEDOIDS PADA POPULARITAS LINE WEBTOON

Fairuz Salsabila[1*], Nur Azise[2] , Muhammad Ali Ridla[3]
[123]Program Studi Sistem Informasi, Universitas Ibrahimy, Situbondo

**Penulis korespondensi: limeilin65@gmail.com*

ABSTRAK

Menerapkan algoritma klustering untuk mengelompokkan popularitas line webtoon dapat sangat berguna karena dapat mengetahui preferensi pembaca sehingga meningkatkan minat baca. Klustering merupakan teknik pengelompokan data yang mirip satu sama lain ke dalam kelompok yang berbeda. Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Means dan K-Medoids pada dataset Line Webtoon. Output dari penelitian ini adalah hasil evaluasi kinerja dari masing-masing algoritma. Data yang digunakan dalam penelitian ini berjumlah 718 record data. Evaluasi kinerja kedua algoritma menggunakan Davies-Bouldin Index (DBI). Hasil evaluasi menunjukkan nilai K-Means sebesar 0.809, sedangkan K-Medoids sebesar 1.1. Hasil tersebut menunjukkan bahwa K-Means menghasilkan klustering yang lebih baik dibandingkan K-Medoids berdasarkan metrik DBI.

Kata kunci: Klustering, K-Means, K-Medoids

1 PENDAHULUAN

Perkembangan teknologi informasi dan internet telah membawa perubahan besar dalam berbagai aspek kehidupan, termasuk dalam cara masyarakat mengonsumsi konten. Salah satu platform digital yang berkembang pesat adalah *Line Webtoon*, yang menyediakan komik daring dalam format visual yang menarik dan mudah diakses.(SEMBIRING & TRIANA, 2022) Webtoon telah menjadi populer di berbagai kalangan. Dilansir dari CNN Indonesia, pengguna aktif Line Webtoon mencapai 6 juta pengguna (Putra, 2020).

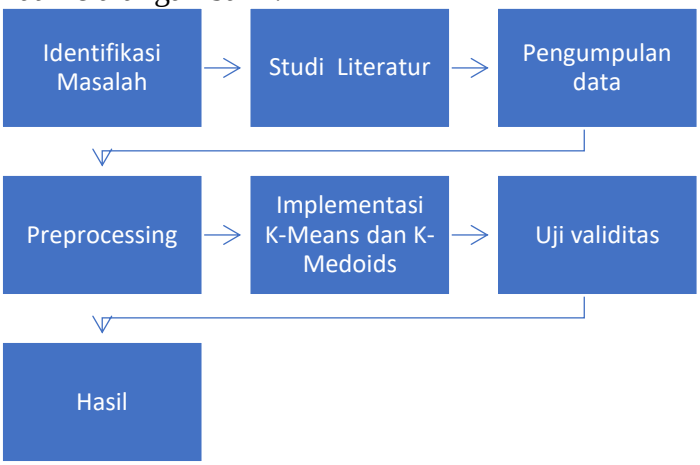
Dalam mengelompokkan data popularitas *Webtoon*, penggunaan algoritma klustering menjadi perlu untuk memahami preferensi pembaca sehingga dapat meningkatkan minat baca. Karena minat baca di Indonesia masih rendah. Hanya berkisar 0,001% menurut data UNESCO (Devega, 2017). Algoritma klustering K-Means dan K-Medoids adalah dua metode yang sering digunakan untuk mengelompokkan data. K-Means dikenal dengan kesederhanaannya dan efisiensinya, sedangkan K-Medoids lebih robust terhadap outliers dan memberikan hasil yang lebih stabil dalam beberapa kasus (Meiriza & Ali, 2023).

Penelitian yang dilakukan oleh (Supriyadi et al., 2021), menunjukkan bahwa klustering pada data armada kendaraan truk mendapatkan nilai (Davies-Bouldin Index) DBI untuk validasi dengan algoritma K-Means sebesar 0.67 dan untuk algoritma K-Medoids sebesar 1.78. Sehingga, algoritma K-Means dinilai lebih relevan karena nilai DBI yang lebih rendah. Penelitian lainnya yang dilakukan oleh (Marlina et al., 2018), menunjukkan bahwa klustering pada data wilayah sebaran cacat pada anak memperoleh tiga kluster. Validitas yang digunakan adalah *Silhoutte Coefficient*. Adapun nilai yang dihasilkan dari algoritma K-Means adalah 0.1443 sedangkan untuk K-Medoids adalah 0.5009. Sehingga K-Medoids dinilai lebih baik.

Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Means dan K-Medoids dalam mengelompokkan popularitas Webtoon di platform Line Webtoon. Perbandingan ini dilakukan untuk menentukan algoritma mana yang lebih efektif dan efisien dalam mengelompokkan data popularitas berdasarkan pada Davies-Bouldin Index (DBI). Dengan melakukan perbandingan kinerja antara dua algoritma, diharapkan dapat diperoleh wawasan yang lebih mendalam mengenai efektivitas masing-masing metode dalam konteks data popularitas Webtoon. Hasil penelitian ini diharapkan dapat membantu dalam memilih algoritma klustering yang paling sesuai untuk analisis data popularitas Line Webtoon.

2 METODE PENELITIAN

Pada prosedur penelitian, tahapan-tahapan harus dilakukan dengan rinci, terstruktur dan sistematis agar target penelitian dapat tercapai sesuai dengan yang diharapkan. Adapun tahapan penelitian dapat dilihat melalui gambar 1.



Gambar 1. Tahapan dalam Penelitian

2.1 Identifikasi Masalah

Terdapat perbedaan kinerja antara algoritma K-Means dan K-Medoids yang telah ditunjukkan dalam berbagai studi sebelumnya, dimana hasilnya dapat bervariasi bergantung pada konteks dan jenis data yang digunakan. Namun dalam konteks data popularitas *Webtoon*, belum ada penelitian yang secara langsung membandingkan algoritma K-Means dan K-Medoids untuk menentukan algoritma yang lebih efektif. Sehingga, diperlukan untuk mengkaji performa K-Means dan K-Medoids pada dataset *Webtoon.*

2.2 Studi Literatur

Pada tahapan ini, dilakukan pengumpulan data dan informasi untuk dapat membantu penelitian dengan memanfaatkan studi ilmiah maupun literatur yang berkaitan.

2.3 Pengumpulan Data

Penelitian dilakukan dengan menggunakan data sekunder pada dataset *Line Webtoon Originals* yang didapat dari Kaggle. Terdapat 718 *record* data dengan 8 atribut. Atribut tersebut meliputi, *genre, authors, weekdays, length, subscribers, ratings, views, likes.* Akan tetapi, dilakukan seleksi atribut sehingga hanya menyisakan 6 atribut yaitu *genre, length, subscribers, ratings, views, likes* karena dianggap kurang mempengaruhi popularitas. *Genre* dapat berupa komedi, fantasi, horor, romantis dan lainnya. *Length* merupakan jumlah episode yang sudah diterbitkan di setiap komik. Sedangkan *subscribers, ratings, views,* dan *likes,* merupakan jumlah *subscribers, ratings, views,* dan *likes* dari masing-masing komik. Data yang digunakan merupakan data pada tahun 2022.

2.4 *Preprocessing*

Menurut (Junaedi et al., 2011), pemrosesan data dapat dilakukan dengan melakukan *data cleaning, data integration, data selection,* dan *data transformation.* Pada tahap ini dilakukan *cleaning* dengan pengecekan *noisy* dan *missing value*. Selain itu, juga dilakukan *data selection* dan *data transformation* dengan menyeleksi atribut yang dianggap kurang mempengaruhi popularitas serta mentransform data dengan melakukan normalisasi data dan mengubah atribut genre yang awalannya nominal diubah menjadi numerik.

Tabel 1 Dataset Webtoon Sesudah *Preprocessing*

No	Title id	Genre	Length	Subscribers	Ratings	Views	Likes
1	4633	0.000	0.00229	0.01865	0.98476	0.00083	0.00179
2	4631	0.667	0.00229	0.00094	0.98286	0.00000	0.00010
3	4630	0.111	0.00229	0.00332	0.67048	0.00008	0.00009
4	4628	0.222	0.00686	0.00130	0.95619	0.00013	0.00028
5	4627	0.333	0.00343	0.00220	0.92762	0.00005	0.00000
6	4626	0.444	0.00343	0.00630	0.98476	0.00024	0.00064
7	4625	0.000	0.00343	0.01131	0.87810	0.00031	0.00088
…	…	…	…	…	…	…	…
718	431	0.556	0.13501	0.02962	0.73333	0.02694	0.01838

2.5 Implementasi K-Means dan K-Medoids

Klustering merupakan metode untuk menemukan struktur kluster dalam sebuah dataset. Jika mirip maka dimasukkan ke dalam kluster yang sama sedangkan yang ketidak miripannya jauh maka dimasukkan ke dalam kluster yang berbeda.(Sinaga & Yang, 2020) K-Means dan K-Medoids merupakan algoritma yang populer dalam klustering.(Kamila et al., 2019) Tujuan utama dari klustering adalah menemukan struktur dalam data tanpa menggunakan label atau kategori yang sudah ditentukan sebelumnya.

2.5.1 K-Means

Menurut buku *Data Clustering: Algorithms and Its Application*, K-Means Clustering adalah salah satu algoritma klustering paling sederhana dan efisien yang diusulkan oleh literatur data klustering. K-Means Clustering juga merupakan partisi algoritma clustering yang paling banyak digunakan.(Oyelade et al., 2019) Dengan langkah-langkah sebagai berikut:

1. Tentukan jumlah Kluster (K)

Syarat untuk nilai *k* tidak boleh lebih dari jumlah data. Inisialisasi k pusat kluster (centroid)
Inisialisasi dilakukan secara random sebanyak nilai k, titik ini akan menjadi centroid dari masing-masing kluster.

2. Inisialisasi *k* pusat kluster (centroid)
   Inisialisasi dilakukan secara random sebanyak nilai k, titik ini akan menjadi centroid dari masing-masing kluster.
3. Hitung jarak ke centroid
   Alokasikan masing-masing data ke centroid terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Jarak paling dekat antara satu data dengan centroid tertentu akan memengaruhi data di dalam kluster.

$$\boldsymbol{d_{kc}}\sqrt{\sum_{\boldsymbol{j=1}}^{\boldsymbol{m}}(\boldsymbol{k_{ia}}-\boldsymbol{c_{ib}})^{\boldsymbol{2}}} \quad (1)$$

$\boldsymbol{d_{kc}}$: jarak antara titik data ke centroid
$\boldsymbol{m}$: jumlah dimensi (jumlah atribut)
$\boldsymbol{k_{ia}}$: koordinat ke-j dari pusat klaster
$\boldsymbol{c_{ib}}$: koordinat ke-j dari titik data

$\boldsymbol{k_{ia} - c_{ib}}$: perbedaan antara koordinat ke-j dari titik data dan koordinat ke-j dari pusat klaster

4. Tentukan centroid baru
   Centroid baru bisa didapat dari rata-rata data yang ada pada masing-masing kluster.

$$\boldsymbol{c_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_i} \quad (2)$$

$\boldsymbol{c_k}$: pusat klaster ke-k yang baru
$\boldsymbol{n_k}$: jumlah titik data dalam klaster ke-k
$\boldsymbol{x_i}$: vektor titik data ke-i dalam klaster ke-k

5. Lakukan iterasi
   Jika masih ada data yang berpindah kluster atau nilai centroid maka ulangi langkah 3.
6. Evaluasi
   Pada tahapan ini dapat dilakukan dengan menggunakan (Davises-Bouldin Index)DBI

2.5.2 K-Medoids

K-Medoids adalah algoritma klasterisasi yang mirip dengan K-Means, tetapi alih-alih menggunakan rata-rata dari titik data dalam klaster sebagai pusat (centroid), K-Medoids menggunakan titik data yang sebenarnya dalam klaster sebagai medoid. Dengan langkah-langkah sebagai berikut:

1. Inisialisasi
   Pilih *k* titik sebagai medoid awal dari sekumpulan data
2. Penetapan Klaster
   Tetapkan setiap titik dari data ke medoid terdekat. Jarak antara titik data x_i dan medoid m_j dapat dihitung dengan menggunakan jarak Euclidean.

$$\boldsymbol{d(x_i, m_j) = \sqrt{\sum_{i=1}^{n}(x_{il} - m_{jl})^2}}$$

$\boldsymbol{x_i}$: titik data ke-*i*
$\boldsymbol{m_j}$: medoid atau titik pusat ke-*j*
$\boldsymbol{l}$: indeks yang digunakan untuk menjumlahkan komponen-komponen vektor $\boldsymbol{x_i}$ dan $\boldsymbol{m_j}$
$\boldsymbol{x_{il}}$: komponen ke-*l* dari titik data $\boldsymbol{x_i}$
$\boldsymbol{m_{jl}}$: komponen ke-*l* dari medoid $\boldsymbol{m_j}$
$\boldsymbol{(x_{il} - m_{jl})^2}$: Hitung selisih antara komponen ke-*l* dari $\boldsymbol{x_i}$ dan $\boldsymbol{m_j}$ **,** lalu kuadratkan hasilnya

$\boldsymbol{\sum_{i=1}^{n}(x_{il} - m_{jl})^2}$: Jumlah semua kuadrat selisih komponen tersebut.

$\boldsymbol{\sqrt{\sum_{i=1}^{n}(x_{il} - m_{jl})^2}}$: Ambil akar kuadrat dari jumlah tesebut. Hasil akhirnya adalah jarak Euclidean antara titik $\boldsymbol{x_i}$ dan $\boldsymbol{m_j}$

3. Memperbarui Medoid
   Untuk setiap klaster, pilih medoid baru meminimalkan total jarak dari semua titik dalam klaster ke medoid. Total jarak dalam klaster C_j dihitung dengan rumus sebagai berikut:

$$\boldsymbol{(m_j) = \sum_{x_i \in C_j} d(x_{i,} m_j)}$$

$\boldsymbol{m_j}$**:** medoid (titik data aktual yang dipilih sebagai pusat klaster)

$\boldsymbol{C_j}$: himpunan titik data yang dikelompokkan berdasarkan kedekatan terhadap medoid

$\boldsymbol{d(x_i, m_j)}$: jarak antara titik data $\boldsymbol{x_i}$ dan calon medoid m

$\sum_{x_i \in C_j} \boldsymbol{d(x_i, m_j)}$: total jarak dari semua titik data $\boldsymbol{x_i}$ dalam klaster $\boldsymbol{C_j}$ ke calon medoid m.

4. Iterasi
   Ulangi langkah 2 dan 3 hingga medoid tidak berubah lagi.
5. Hasil
   Setelah iterasi konvergen, hasil akhir adalah klaster yang dibentuk dengan medoid sebagai pusatnya.

2.6 Uji Validitas

Pada tahap uji validitas, penelitian ini menggunakan *Davies Bouldin Index* (DBI) untuk mengevaluasi kinerja algoritma klusterisasi. DBI adalah metode evaluasi yang digunakan untuk mengukur seberapa baik kluster-kluster yang dihasilkan dari proses klusterisasi. Jika nilai DBI lebih rendah menunjukkan bahwa hasil klusterisasi lebih baik.

2.7 Hasil

Data yang telah dilakukan pemrosesan dan sudah diimplementasikan K-Means dan K-Medoids maka akan dibandingkan kinerjanya menggunakan DBI

3 HASIL DAN PEMBAHASAN

Setelah dilakukan *preprocessing* dan implementasi dari algoritma K-Means dan K-Medoids, maka dilakukan evaluasi menggunakan Davies-Bouldin Index (DBI) untuk mengevaluasi kinerja algoritma K-Means dan K-Medoids. DBI adalah met rik yang mengukur rata-rata kesamaan kluster, dengan nilai yang lebih rendah menunjukkan kluter yang lebih baik. Berikut adalah tabel yang menunjukkan hasil DBI untuk berbagai jumlah kluter untuk kedua algoritma:

Tabel 2. Hasil DBI

Nilai *k*	DBI K-Means	DBI K-Medoids
3	0.905	1.1311
4	0.866	1.592
5	0.892	1.905
6	0.809	1.1
7	0.812	1.738
8	0.843	1.238
9	0.914	1.171
10	0.846	1.248

Dari hasil diatas, terlihat bahwa nilai DBI terendah untuk K-Means adalah 0.809 ketika jumlah klaster (k) adalah 6, sedangkan nilai DBI terendah untuk K-Medoids adalah 1.1 juga pada *k* = 6. Ini menunjukkan bahwa K-Means lebih efektif dalam mengelompokkan data popularitas Webtoon dibandingkan K-Medoids.

4 KESIMPULAN

Penelitian ini membandingkan kinerja algoritma K-Means dan K-Medoids dalam mengelompokkan data popularitas *Webtoon*, dengan menggunakan *Davies-Bouldin Index* sebagai metrik evaluasi. Hasil menunjukkan bahwa nilai DBI pada K-Means lebih rendah dibandingkan dengan K-Medoids. Pengklusteran ini dapat membantu dalam memahami preferensi pembaca sehingga dapat meningkatkan minat baca.

DAFTAR PUSTAKA

Devega, E. (2017). *TEKNOLOGI Masyarakat Indonesia: Malas Baca Tapi Cerewet di Medsos*. Kominfo. https://www.kominfo.go.id/content/detail/10862/teknologi-masyarakat-indonesia-malas-baca-tapi-cerewet-di-medsos/0/sorotan_media#:~:text=Fakta pertama%2C UNESCO menyebutkan Indonesia,1 orang yang rajin membaca!

Junaedi, H., Budianto, H., Maryati, I., & Melani, Y. (2011). Data transformation pada data mining. *Prosiding Konferensi Nasional Inovasi Dalam Desain Dan Teknologi-IDeaTech*, *7*(3), 93–99.

Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, *5*(1), 119–125.

Marlina, D., Putri, N. F., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak. *J. CoreIT*, *4*(2), 64.

Meiriza, A., & Ali, E. (2023). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan. *Indonesian Journal of Computer Science*, *12*(2).

Oyelade, J., Isewon, I., Oladipupo, O., Emebo, O., Omogbadegun, Z., Aromolaran, O., Uwoghiren, E., Olaniyan, D., & Olawole, O. (2019). Data Clustering: Algorithms and Its Applications. *2019 19th International Conference on Computational Science and Its Applications (ICCSA)*, 71–81. https://doi.org/10.1109/ICCSA.2019.000-1

Putra, M. A. (2020). *Alasan Webtun Paling Laris di Indonesia*. CNN Indonesia. https://www.cnnindonesia.com/hiburan/20201002142816-241-553665/alasan-webtun-paling-laris-di-indonesia

SEMBIRING, B., & TRIANA, R. I. A. (2022). *Pengaruh Media Komik Line Webtoon Terhadap Kemampuan Menulis Cerpen Di SMK Swasta GBKP Kabanjahe Tp 2021/2022*.

Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727.

Supriyadi, A., Triayudi, A., & Sholihati, I. D. (2021). Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *6*(2), 229–240.