Perbandingan K-Means dan K-Medoids Untuk Clustering Lagu Setipe di Spotify Berdasarkan Karakteristik Audio

Syalomiele Pratama Agustinus Susanto[1], Suastika Yulia Riska[2]
[1,2]Prodi Teknik Informatika, Fakultas Teknologi dan Desain, Institut Teknologi dan Bisnis Asia Malang
Email: [1]pratamasyalomiele@gmail.com, [2]riska.suastika@asia.ac.id

Abstract. *The growth of music streaming services like Spotify has created a demand for song clustering systems that can enhance the user experience through more accurate recommendations. To improve user experience, a clustering system is needed to group songs based on audio feature similarities such as danceability, energy, acousticness, instrumentalness, liveness, speechiness, and valence. This study compares two clustering algorithms, K-Means and K-Medoids, for grouping Spotify songs based on these audio features. The K-Means algorithm is known for its computational efficiency, while K-Medoids is more robust to outliers. Evaluation was carried out using the Davies-Bouldin Index (DBI) to measure the quality of cluster separation. The results of the study show that K-Means performs best with k = 3, yielding a DBI of 0.857, while K-Medoids performs best with k = 9, yielding a DBI of 0.844. Although K-Medoids slightly outperforms K-Means in terms of cluster quality, K-Means excels in computational efficiency. This research provides valuable insights into the effectiveness of both algorithms in music recommendation systems based on clustering and contributes to the literature on song clustering in streaming platforms.*
Keywords: *Clustering, K-Means, K-Medoids, Audio Characteristic, Spotify.*

Abstrak. *Pertumbuhan layanan streaming musik seperti Spotify menghadirkan kebutuhan akan sistem pengelompokan lagu yang mampu meningkatkan pengalaman pengguna melalui rekomendasi yang lebih akurat. Untuk meningkatkan pengalaman pengguna, diperlukan sistem pengelompokan lagu berdasarkan kemiripan fitur audio seperti danceability, energy, acousticness, instrumentalness, liveness, speechiness, dan valence. Penelitian ini membandingkan dua algoritma clustering, yaitu K-Means dan K-Medoids, dalam mengelompokkan lagu-lagu Spotify berdasarkan fitur audio tersebut. Algoritma K-Means dikenal efisien dalam komputasi, sementara K-Medoids lebih robust terhadap outlier. Evaluasi dilakukan menggunakan Davies-Bouldin Index (DBI) untuk mengukur kualitas pemisahan antar-kluster. Hasil penelitian menunjukkan bahwa K-Means memberikan hasil terbaik pada k = 3 dengan DBI 0,857, sedangkan K-Medoids memberikan hasil terbaik pada k = 9 dengan DBI 0,844. Meskipun K-Medoids sedikit lebih baik dalam hal kualitas klaster, K-Means lebih unggul dalam efisiensi waktu komputasi. Penelitian ini memberikan wawasan penting mengenai efektivitas kedua algoritma dalam sistem rekomendasi musik berbasis clustering dan dapat memperkaya literatur tentang pengelompokan lagu di platform streaming*
Kata Kunci: *Clustering, K-Means, K-Medoids, Karakteristik Audio, Spotify.*

1. Pendahuluan

Perkembangan teknologi digital dan arus globalisasi telah mengubah secara signifikan pola konsumsi musik, dengan *platform streaming* menjadi medium utama dalam kehidupan sehari-hari. Layanan seperti *Spotify* menyediakan akses ke jutaan lagu dengan beragam karakteristik audio yang terstruktur dalam bentuk fitur-fitur numerik. Untuk meningkatkan

kualitas pengalaman pengguna, khususnya dalam proses rekomendasi dan pencarian lagu dengan kemiripan karakteristik, diperlukan mekanisme komputasional yang cerdas untuk mengelompokkan koleksi lagu tersebut secara efektif. Metode *Clustering* (pengelompokan) dapat diterapkan dengan efektif untuk tujuan ini, yaitu mengelompokkan lagu berdasarkan kemiripan pada ciri-ciri audio seperti *danceability*, *energy*, *acousticness*, *instrumentalness*, *liveness*, *speechiness*, dan *valence*.Secara industri, proses *clustering* berperan penting tidak hanya dalam meningkatkan sistem rekomendasi, tetapi juga dalam mendukung industri musik dalam analisis tren musik digital. Dengan pengelompokan yang akurat akan dapat membantu mengenali pola preferensi pengguna secara *real time*, mengotomatiskan playlist, menargetkan audiens dengan lebih tepat, serta memanfaatkan hasil analisis big data untuk pengambilan keputusan yang lebih efisien. Dua algoritma yang sering dipakai untuk tujuan ini adalah K-Means dan K-Medoids, yang masing-masing memberikan pendekatan berbeda dalam menetapkan pusat dari setiap kelompok. Evaluasi internal seperti *Indeks Davies-Bouldin (DBI)* menjadi tolak ukur penting untuk mengevaluasi seberapa baik algoritma tersebut dalam membentuk kelompok-kelompok lagu (Pangestu et al., 2024).

Penelitian sebelumnya umumnya menggunakan K-Means untuk mengelompokkan lagu-lagu populer yang ada di *Spotify* dan menetapkan jumlah klaster dengan menggunakan *DBI* atau *Silhouette*. Namun, hingga saat ini, belum ada analisis mendalam yang membandingkan K-Means dengan K-Medoids pada kumpulan data yang lengkap mengenai lagu. Misalnya, dalam penelitian yang dilakukan oleh (Ikhsan Firmansyah et al., 2023), mereka menerapkan K-Means dengan optimalitas DBI pada sembilan klaster. Di sisi lain, (Privandhani, 2022) sudah menggunakan kedua algoritma dalam studinya, tetapi hanya terbatas pada lagu *pop* Indonesia tanpa eksplorasi mendalam mengenai efisiensi dan validasi internal. Padahal, kedua algoritma tersebut layak untuk dibandingkan karena K-Means dikenal efisien dalam proses komputasi dan mampu bekerja baik pada data berukuran besar, namun sensitif terhadap *outlier* dan data ekstrem, sedangkan K-Medoids lebih *robust* terhadap *noise* dan *anomali* karena menggunakan titik data aktual sebagai pusat klaster, bukan nilai rata-rata(Pangestu et al., 2024). Perbedaan karakteristik tersebut menjadikan penelitian komparatif antara K-Means dan K-Medoids penting untuk memahami algoritma mana yang lebih optimal dalam mengelompokkan lagu berdasarkan karakteristik audio secara global.

Literatur yang relevan menunjukkan bahwa K-Means dapat menghasilkan nilai *DBI* yang rendah, seperti 0,299 untuk tiga klaster pada lagu-lagu populer di *Spotify* tahun 2023 (Nuriska et al., 2024). Sementara itu, K-Medoids juga menunjukkan performa yang baik dalam klasifikasi audio dan video di berbagai domain. Namun, banyak penelitian yang hanya menganalisis satu algoritma tanpa melakukan perbandingan. Contohnya, (Rohmah Zaidah et al., 2021) membandingkan tiga algoritma dan menemukan bahwa metode Agglomerative lebih unggul dibandingkan K-Means dan K-Medoids. Penelitian ini dibedakan dengan membandingkan secara langsung K-Means dan K-Medoids pada dataset Spotify, dan menggunakan *DBI* sebagai ukuran validasi untuk menentukan mana yang lebih baik dan efisien dalam mengelompokkan lagu berdasarkan kemiripan karakteristik audio.

Penelitian ini mengimplementasikan algoritma K-Means dan K-Medoids pada data audio lagu *Spotify* dengan fitur seperti *danceability, energy, acousticness, instrumentalness, liveness, speechiness,* dan *valence* serta mengevaluasi performanya menggunakan *indeks Davies-Bouldin (DBI)* untuk menentukan algoritma yang paling efisien dalam meminimalkan *scatter* dan memaksimalkan *separation* antar-cluster. Hasil penelitian diharapkan membantu pengambangan sistem rekomendasi musik yang lebih personalisasi dan akurat, sekaligus Memberikan evaluasi komprehensif terhadap perbandingan K-Means dan K-Medoids menggunakan metrik *DBI* pada data *Spotify*, serta memperkaya literatur mengenai komparasi K-Means dan K-Medoids dalam konteks data Spotify yang masih jarang dieksplorasi.

2. Tinjauan Pustaka

2.1 Clustering

Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan atau kemiripan. *Clustering* juga disebut sebagai segmentasi data dalam beberapa aplikasi. *Clustering* bersifat unsupervised learning karena tidak membutuhkan proses training dalam pengelompokan kelasnya, tidak seperti klasifikasi yang bersifat *supervised learning* (Prianto & Bunyamin, 2020). *Clustering* dipahami sebagai suatu teknik pengelompokan objek-objek data ke dalam himpunan-himpunan *(cluster)* sedemikian rupa sehingga objek di dalam satu cluster memiliki kemiripan tinggi satu sama lain, tujuan utamanya adalah menemukan struktur, pola, atau kelompok homogen dalam kumpulan data yang belum berlabel sehingga mempermudah analisis dan pengambilan keputusan (Evandi et al., 2025; Zahra et al., 2024).

2.2 K-Means

Algoritma K-Means merupakan salah satu algoritma *clustering* yang masuk dalam kelompok *unsupervised learning* yang digunakan untuk membagi data menjadi beberapa kelompok dengan sistem partisi. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada algoritma K-Means, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dahulu target kelasnya. Masukan yang diterima adalah data atau objek dan k kelompok *(cluster)* yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut (Sari, 2023; Wanto et al., 2020).Langkah – langkah *clustering* dengan metode K-Means adalah sebagai berikut (Mardiansa et al., 2023):

1. Tentukan nilai k sebagai jumlah *cluster* yang ingin dibentuk.
2. Inisialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara, namun yang paling sering dilakukan adalah dengan cara acak yang diambil dari data yang ada
3. Menghitung jarak setiap data input terhadap masing – masing centroid menggunakan rumus jarak *Euclidean (Euclidean Distance)* hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Persamaan *Euclidean Distance* ditunjukan pada persamaan 1 :

$$d(x-y) = \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2} \quad (1)$$

dimana: d = titik dokumen; x = pusat cluster; y = data

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centorid* (jarak terkecil).
5. Memperbaharui nilai *centorid*. Nilai *centorid* baru di peroleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan persamaan 2:

$$C_k = (\frac{1}{n_k}) \sum \quad d_1 \quad (2)$$

dimana n_k = jumlah data dokumen dalam cluster; k dan d_1 = dokumen dalam *cluster* k

6. Melakukan perulangan dari langkah 2 hingga 5,sampai anggota tiap *cluster* tidak ada yang berubah.Jika langkah 6 telah terpenuhi, maka nilai pusat *cluster* pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

2.2 K-Medoids

Algoritma K-Medoids merupakan salah satu metode clustering *partisi (non-hierarki)* yang mirip dengan K-Means, namun berbeda dalam pemilihan "pusat" klaster (centroid) yang digunakan. Alih-alih menggunakan rata-rata vektor sebagai pusat klaster, K-Medoids memilih suatu objek nyata dari himpunan data sebagai *medoid* (perwakilan klaster) sehingga titik pusat *cluster* selalu berasal dari dataset (Alodia et al., 2021).Langkah-langkah dalam perhitungan algoritma K-Medoids adalah sebagai berikut (Ningrum et al., 2021):

1. Inisialisasi pusat cluster sebanyak k (jumlah cluster)

2. Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:

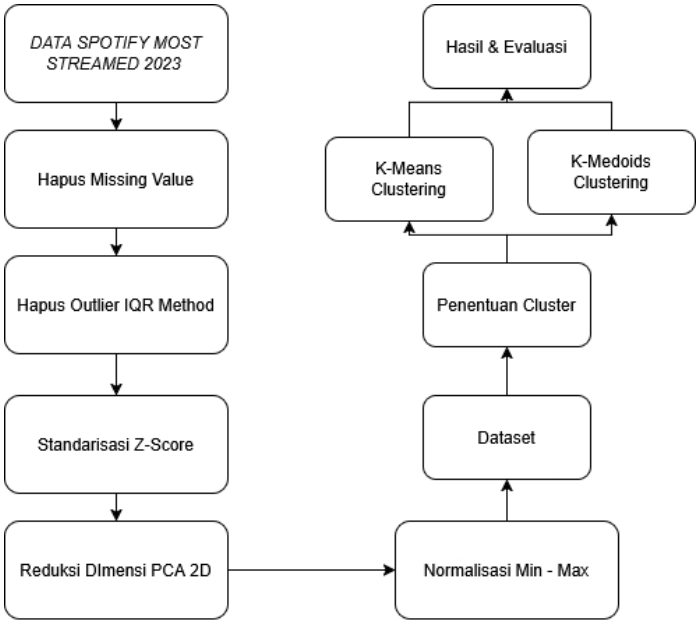
$$d_{ij} = \sqrt{\sum_{a=1}^{p} (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)} \quad (3)$$

dimana i=1,…..,n; j=1,…..,n dan p adalah banyak variabel, serta V adalah matrik varian kovarian.

3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoids* baru.
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru.
5. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total distance lama. Jika S < 0, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai *medoids*
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoids, sehingga didapatkan *cluster* beserta anggota cluster masing-masing

3. Metode Penelitian

Data yang digunakan dalam penelitian ini diperoleh dari *Kaggle* (Elgiriyewithana, 2023) mengenai kompresi data dari lagu yang paling banyak diputar di *Spotify 2023*, yang mencakup atribut serta popularitas setiap lagu. Atribut yang dianalisis dibatasi pada fitur karakteristik audio, seperti *danceability, energy, acousticness, instrumentalness, liveness, speechiness, dan valence*. Penelitian ini menggunakan algoritma K-Means dan K-Medoids, yang kemudian dibandingkan satu sama lain. Selain itu, penelitian ini juga melakukan proses *clustering* untuk membentuk beberapa jumlah *cluster*, yaitu 2, 3, 5, 7, dan 9 *cluster*. Kerangka penelitian dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

Berdasarkan Gambar 1, Kerangka penelitian ini menggambarkan alur proses yang dilakukan dalam penelitian mulai dari pengolahan data hingga evaluasi hasil *clustering*. Tahap awal dilakukan pembersihan data dengan menghapus nilai kosong (*missing values*) dan mendeteksi *outlier* menggunakan metode *Interquartile Range (IQR)* untuk menjaga kualitas data.

Selanjutnya dilakukan standarisasi menggunakan metode Z-Score agar setiap fitur memiliki skala distribusi yang sebanding, diikuti dengan reduksi dimensi menggunakan *Principal Component Analysis (PCA)* menjadi dua komponen utama untuk memudahkan visualisasi data. Setelah itu, data dinormalisasi kembali menggunakan metode *Min–Max* agar semua fitur berada dalam rentang [0,1]. Tahap berikutnya adalah proses pengelompokan menggunakan dua algoritma, yaitu K-Means dan K-Medoids. Tahap akhir penelitian adalah evaluasi hasil *clustering* menggunakan metrik *Davies-Bouldin Index (DBI)* untuk menilai kualitas pembentukan *cluster* serta menentukan algoritma yang paling optimal dalam mengelompokkan lagu berdasarkan kemiripan karakteristik audio. Kemudian untuk fitur karakteristik audio yang digunakan ditunjukan pada Tabel 1.

Tabel 1. Fitur audio spotify

Fitur Audio	***Keterangan***
Danceability	Seberapa cocok lagu untuk dijadikan sebagai musik untuk menari
Valence	Positivitas dari konten musik lagu
Energy	Tingkat energi yang dirasakan dari lagu
Acousticness	Jumlah suara akustik yang ada dalam lagu
Instrumentalness	Jumlah konten instrumental dalam lagu
Liveness	Kehadiran elemen pertunjukan langsung dalam lagu
Speechiness	Jumlah kata-kata yang diucapkan dalam lagu

Pada Tabel 1 dijelaskan untuk pengertian setiap fitur audio yang digunakan dalam penelitian ini berdasarkan sumber data dari (Elgiriyewithana, 2023) mengenai lagu-lagu *most streamed Spotify 2023*, dengan fokus pada beberapa fitur audio. *Danceability* mengukur sejauh mana lagu cocok untuk menari, sementara *valence* menggambarkan tingkat keceriaan lagu. *Energy* mengukur intensitas lagu, *acousticness* menilai jumlah suara akustik dalam lagu, dan *instrumentalness* menunjukkan proporsi musik tanpa vokal. *Liveness* mencatat elemen pertunjukan langsung, sementara *speechiness* mengukur banyaknya ucapan atau kata dalam lagu. Setiap fitur ini digunakan untuk mengklaster lagu-lagu berdasarkan karakteristik audio mereka dalam penelitian ini.

4. Hasil dan Pembahasan

Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Means dan K-Medoids dalam mengelompokkan lagu berdasarkan karakteristik audio dari dataset *Spotify* 2023. Dataset awal berisi 953 lagu, namun setelah dilakukan pembersihan data dan penghapusan nilai yang hilang, hanya tersisa 694 lagu yang digunakan dalam analisis. Proses reduksi dimensi dilakukan dengan menggunakan *Principal Component Analysis (PCA)*, yang menghasilkan dua komponen utama yang dapat menjelaskan 52,73% dari total variansi data. Langkah ini dilakukan untuk mempermudah visualisasi dan mempercepat proses *clustering* tanpa kehilangan informasi penting dalam data. Evaluasi dilakukan pada nilai k = 2, 3, 5, 7, dan 9 dengan menggunakan metrik *Davies-Bouldin Index (DBI)* sebagai ukuran kualitas *cluster*. Semakin kecil nilai *DBI*, semakin baik pemisahan antar-*cluster* dan semakin kompak distribusi data di dalamnya. Nilai perbandingan evaluasi *davies bouldin index* terhadap jumlah *cluster* menggunakan algoritma K-Means dan K-Medoids ditunjukkan pada Tabel 2.

Tabel 2. Hasil Evaluasi perbandingan K-Means dan K-Medoids

Nilai k	***DBI (K-Means)***	***Waktu (detik)***	***DBI (K-Medoids)***	***Waktu (detik)***
2	1.074	0.01	1.183	0.18
3	0.857	0.01	0.864	0.22
5	0.929	0.01	0.984	0.97
7	0.946	0.01	1.024	1.04
9	0.885	0.01	0.844	1.21

Berdasarkan tabel 2, K-Means memperoleh hasil terbaik pada k = 3 dengan nilai *DBI* sebesar 0,857, sedangkan K-Medoids memberikan hasil terbaik pada k = 9 dengan nilai *DBI* 0,844. Performa ini menunjukkan bahwa secara kualitas pemisahan, K-Medoids sedikit lebih

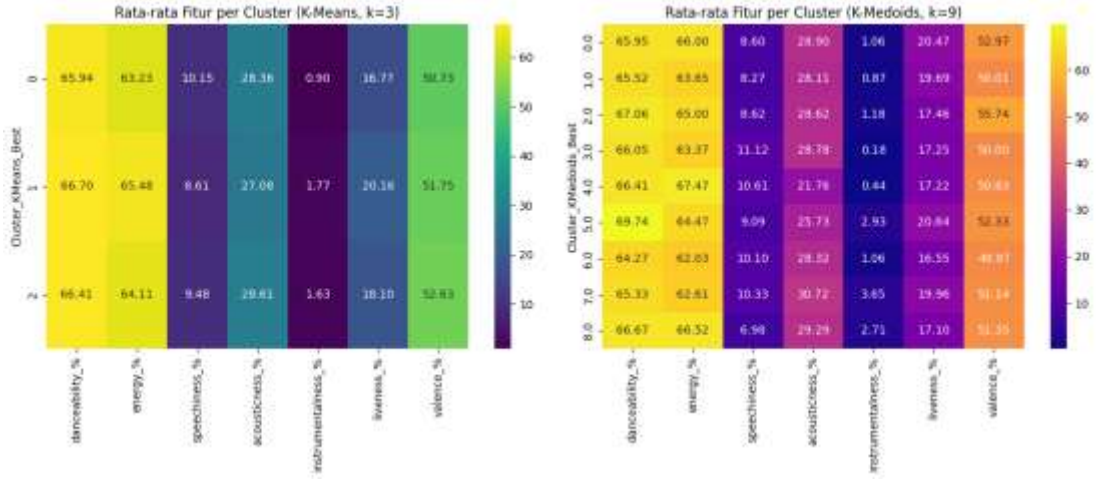
ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

baik, namun K-Means jauh lebih unggul dalam efisiensi waktu komputasi. Rata-rata waktu K-Means adalah 0,01 detik, sedangkan K-Medoids membutuhkan rata-rata 0,72 detik per proses. Kemudian Untuk melihat keseimbangan pembagian data pada setiap *cluster*, distribusi jumlah lagu per *cluster* disajikan pada Tabel 3.

Tabel 3. Distribusi lagu per cluster

Algoritma	***k***	***Distribusi per cluster***
K-Means	2	k1 = 212, k2 = 482
K-Means	3	k1= 146, k2 = 315, k3 = 233
K-Means	5	k1= 128, k2 = 174, k3 = 140, k4 = 75, k5 = 177
K-Means	7	k1= 81, k2 = 163, k3 = 121, k4 = 52, k5 = 94, k6 = 139 ,k7 = 44
K-Means	9	k1= 83, k2 = 135, k3 = 63, k4 = 128, k5 = 55, k6 = 36 ,k7 = 69,k8 = 56 ,k9 = 69
K-Medoids	2	k1 = 361, k2 = 333
K-Medoids	3	k1= 303, k2 = 244, k3 = 147
K-Medoids	5	k1= 141, k2 = 175, k3 = 164, k4 = 99, k5 = 115
K-Medoids	7	k1= 100, k2 = 90, k3 = 106, k4 = 127, k5 = 78, k6 = 109 ,k7 = 84
K-Medoids	9	k1= 104, k2 = 97, k3 = 96, k4 = 65, k5 = 59, k6 = 94 ,k7 = 62,k8 = 69 ,k9 = 48

Berdasarkan tabel 3, distribusi pada K-Means relatif seimbang, sedangkan K-Medoids memberikan pembagian yang lebih spesifik dan mendetail terhadap variasi karakteristik lagu. Hal ini menunjukkan bahwa K-Means lebih cenderung menghasilkan *cluster* yang lebih besar dan homogen, sementara K-Medoids mampu menangkap perbedaan yang lebih halus antara lagu-lagu dalam dataset. Dengan demikian, K-Medoids dapat memberikan gambaran yang lebih rinci tentang struktur data, meskipun dengan distribusi *cluster* yang lebih bervariasi. Kemudian visualisasi *heatmap* untuk memperlihatkan rata-rata nilai setiap fitur audio di masing-masing *cluster* untuk K-Means (k=3) dan K-Medoids (k=9) ditujukan pada Gambar 2.



Gambar 2. Heatmap rata rata fitur

Berdasarkan gambar 2 dapat disimpulkan bahwa pada K-Means dengan k=3, fitur *danceability* dan *energy* mendominasi semua *cluster* dengan nilai sekitar 63–66%, yang menunjukkan bahwa lagu-lagu yang terkelompok memiliki ritme dan dinamika yang tinggi. Sementara itu, fitur *instrumentalness* menunjukkan nilai sangat rendah (sekitar 1%), menandakan bahwa mayoritas lagu dalam *cluster* ini memiliki unsur vokal yang dominan. Di sisi lain, K-Medoids dengan k=9 memberikan variasi yang lebih besar pada fitur *acousticness*, *valence,* dan *liveness,* yang menunjukkan bahwa *cluster* terbentuk berdasarkan nuansa lagu mulai dari lagu akustik yang santai hingga lagu *pop* yang energik. Untuk contoh lagu setipe dari *cluster* terbaik berdasarkan *DBI*, yaitu K-Means (k=3) ditunjukan pada Tabel 4.

Tabel 4. Contoh lagu setipe cluster terbaik K-Means

Nilai k	*Contoth lagu*	*Artist*
1	*Un x100to*	*Bad Bunny, Grupo Frontera*
1	*See you again*	*Tyler, The Creator, Kali Uchis*
2	Seven *(feat. Latto)*	*Jung Kook, Latto*
2	*LALA*	*Myke Towers*
3	*Vampire*	*Olivia Rodrigo*
3	*Cruel samuel*	*Taylor Swift*

Berdasarkan Tabel 4 *cluster* 1 didominasi oleh lagu-lagu yang memiliki elemen ritmis dan melodius, menggabungkan seperti *reggaeton*, *trap*, dan *hip hop* dengan pengaruh *Latin*. Lagu-lagu dalam *cluster* ini cenderung memiliki tempo cepat dan penuh energi, dengan vokal yang kuat dan instrumen elektronik atau akustik yang khas, seperti yang terlihat pada *un x100to (Bad Bunny, Grupo Frontera)* dan *See You Again (Tyler, The Creator, Kali Uchis)*. *Cluster* 2, di sisi lain, lebih fokus pada vokal dominan yang menggabungkan *rap* dan *pop*. Lagu-lagu seperti *Seven feat. Latto (Jung Kook, Latto)* dan *LALA (Myke Towers)* memiliki ritme yang *upbeat* dengan pengaruh *trap* dan *hip hop* yang jelas, namun tetap *catchy*, cocok untuk pendengar yang menikmati perpaduan antara *pop* dan *rap*. Sedangkan *cluster* 3 lebih dominan dengan lagu *pop emosional* dan *indie*, seperti *vampire (Olivia Rodrigo)* dan *Cruel Summer (Taylor Swift)*, yang menyajikan tema lirik yang mendalam tentang cinta dan hubungan, dengan melodi yang lebih lembut dan atmosfer yang lebih introspektif atau gelap. Secara keseluruhan, *cluster* ini menunjukkan keragaman tipe lagu, mulai dari yang enerjik dan ritmis di k1, lebih kombinatif dan *upbeat* di k2, hingga *pop* yang emosional dengan lirik yang dalam di k3. Kemudian untuk contoh lagu setipe dari *cluster* terbaik berdasarkan *DBI,* yaitu K-Medoids (k=9) ditujukan pada Tabel 5.

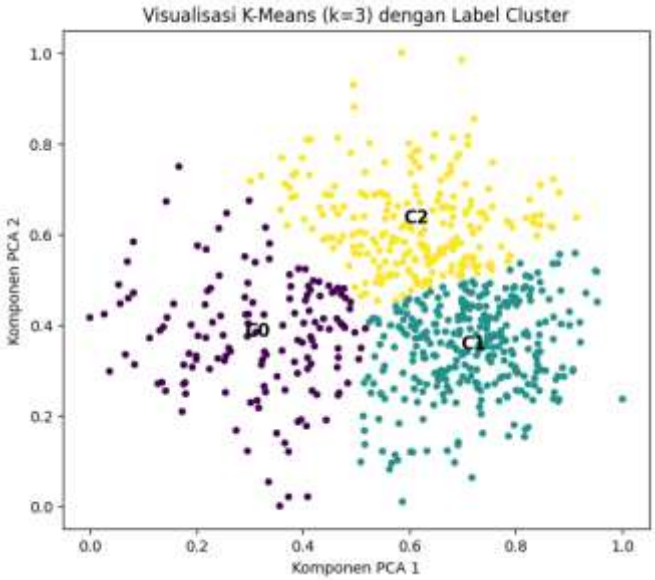
Tabel 5. Contoh lagu setipe cluster terbaik K-Medoids

Nilai k	*Contoth lagu*	*Artist*
1	*Seven (feat Latto)*	*Jung Kook, Latto*
1	*Cupid*	*Fifty Fifty*
2	*Fukumea*	*Gunna*
2	*Classy 101*	*Feid, Young Miko*
3	*Super Shy*	*NewJeans*
3	*Cruel samuel*	*Taylor Swift*
4	*Un x100to*	*Bad Bunny, Grupo Frontera*
4	*See you again*	*Tyler, The Creator, Kali Uchis*
5	*Vampire*	*Olivia Rodrigo*
5	*Amargura*	*Karol G*
6	*LALA*	*Myke Towers*
6	*Where She Goes*	*Bad Bunny*
7	*Sprinter*	*Dave, Central Cee*
7	*Last Night*	*Morgan Wallen*
8	*Kill Bill*	*SZA*
8	*I Wanna Be Yours*	*Arctic Monkeys*
9	*Anti-Hero*	*Taylor Swift*
9	*Baby Don't Hurt Me*	*David Guetta, Anne-Marie, Coi-Leray*

Berdasarkan Tabel 5 pada *cluster k=1*, lagu-lagu seperti *Seven (feat. Latto)* dan *Cupid* memiliki nuansa *pop modern* yang ceria, dengan ritme cepat dan elemen *rap* yang *catchy*. *Cluster k=2* menonjolkan elemen *trap* dan *hip hop* dengan pengaruh *Latin*, seperti yang terlihat pada *Fukumea (Gunna)* dan *Classy 101 (Feid, Young Miko),* yang menawarkan *beat* berat dan vokal yang kuat. *Cluster k=3* terdiri dari lagu-lagu *pop* yang ringan dan *upbeat*, namun tetap emosional,

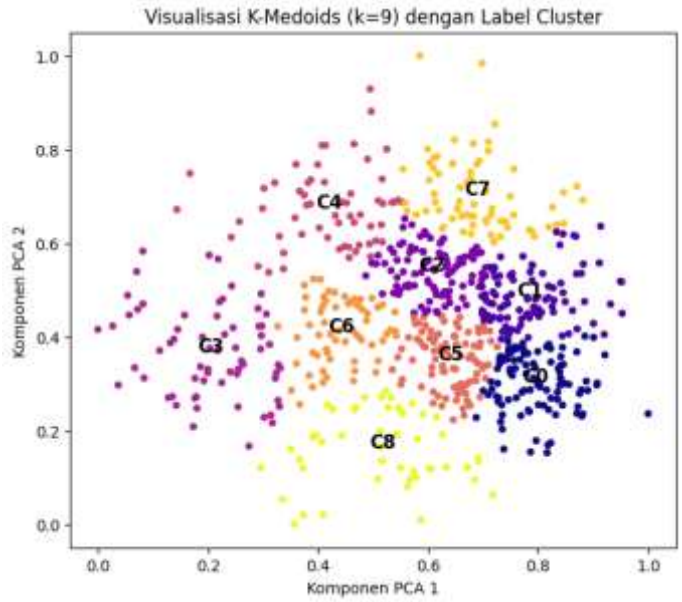
ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

seperti *Super Shy (New Jeans)* dan *Cruel Summer (Taylor Swift)*, dengan lirik yang lebih mendalam. Lagu-lagu di *cluster k=4*, seperti *Un x100to (Bad Bunny)* dan *See You Again (Tyler, The Creator)*, menggabungkan *reggaeton* dan *hip hop* dengan tempo cepat dan penuh energi. Pada *cluster k=5*, lagu seperti *Vampire (Olivia Rodrigo)* dan *Amargura (Karol G)* menghadirkan tema cinta dan hubungan yang kompleks, dengan melodi yang lebih emosional dan melankolis. *Cluster k=6* berisi lagu-lagu seperti *LALA (Myke Towers)* dan *Where She Goes (Bad Bunny)*, yang memiliki pengaruh *trap* dan *reggaeton* dengan nuansa *urban* yang santai namun energik. *Cluster k=7* menggabungkan *hip hop* dan *country*, terlihat pada *Sprinter (Dave, Central Cee)* dan *Last Night (Morgan Wallen)*, dengan perpaduan yang unik dan lirik yang kuat. *Cluster k=8* menampilkan lagu-lagu *R&B* dan *indie rock* seperti *Kill Bill (SZA)* dan *I Wanna Be Yours (Arctic Monkeys)*, yang lebih introspektif dengan lirik dan melodi yang melankolis. Terakhir, *cluster k=9* berisi lagu-lagu *pop* dan elektronik yang energik seperti *Anti-Hero (Taylor Swift)* dan *Baby Don't Hurt Me (David Guetta, Anne-Marie, Coi-Leray)*, yang menawarkan melodi *catchy* dan ritme cepat. Secara keseluruhan, *cluster* ini mencerminkan keragaman tipe dan gaya musik, dari ceria hingga emosional dan kombinasi yang eksperimental. Kemudian untuk visualisasi hasil *clustering* menggunakan K-Means dengan k terbaik yaitu k =3 ditunjukan pada Gambar 3.



Gambar 3. Visualisasi Hasil Clustering Menggunakan K-Means

Berdasarkan Gambar 3 berikut menunjukkan hasil visualisasi proses pengelompokan lagu menggunakan algoritma K-Means dengan jumlah *cluster* k=3. Visualisasi dilakukan dengan metode *Principal Component Analysis (PCA)* untuk mereduksi dimensi data ke dua komponen utama agar hasil kluster dapat divisualisasikan secara dua dimensi. Terlihat bahwa setiap titik merepresentasikan satu lagu dengan karakteristik audio tertentu, seperti *danceability, energy, valence*, dan lain-lain. Warna berbeda menunjukkan pembagian ke dalam tiga *cluster* utama. *Cluster* 1, 2, dan 3 terlihat memiliki batasan area yang cukup jelas, menunjukkan bahwa algoritma K-Means berhasil memisahkan lagu berdasarkan kemiripan fitur audio. Kemudian untuk visualisasi hasil *clustering* menggunakan K-Medoids dengan k terbaik yaitu k = 9 ditunjukan pada Gambar 4.



Gambar 4. Visualisasi Hasil Clustering Menggunakan K-Medoids

Sementara itu, Gambar 4 memperlihatkan hasil pengelompokan menggunakan algoritma K-Medoids dengan jumlah *cluster* k = 9. Terlihat bahwa pembagian K-Medoids menghasilkan penyebaran data yang lebih halus dengan sembilan kelompok berbeda. Algoritma ini cenderung lebih tahan terhadap *outlier* dibandingkan K-Means, karena medoid dipilih berdasarkan titik aktual dari dataset, bukan nilai rata-rata. Dengan demikian, meskipun hasil *cluste*r K-Medoids tampak lebih banyak dan sedikit tumpang tindih, pendekatan ini memberikan pemisahan yang lebih realistis terhadap distribusi lagu yang memiliki karakteristik unik.

5. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma K-Means dan K-Medoids memiliki keunggulan masing-masing dalam proses pengelompokan lagu berdasarkan karakteristik audio dari dataset *Spotify 2023*. Algoritma K-Means menunjukkan kinerja terbaik pada k=3 dengan nilai *Davies-Bouldin Index (DBI)* sebesar 0,857 dan waktu komputasi rata-rata 0,01 detik, sehingga lebih efisien dan menghasilkan pembagian *cluster* yang seimbang serta mudah divisualisasikan. Sementara itu, K-Medoids memberikan hasil terbaik pada k=9 dengan *DBI* sebesar 0,844 dan waktu komputasi rata-rata 0,72 detik, yang menunjukkan kualitas pemisahan cluster sedikit lebih baik dan lebih tahan terhadap outlier, meskipun dengan waktu eksekusi yang lebih lama. Hasil heatmap memperlihatkan bahwa K-Means cenderung mengelompokkan lagu dengan karakteristik ritmis dan energik tinggi, sedangkan K-Medoids mampu menangkap variasi fitur seperti *acousticness, valence,* dan *liveness* yang menggambarkan keragaman nuansa lagu. Secara keseluruhan, K-Means lebih unggul dalam efisiensi dan keseimbangan distribusi data, sedangkan K-Medoids lebih unggul dalam ketelitian dan kemampuan mengidentifikasi variasi halus antar lagu, sehingga pemilihan algoritma sebaiknya disesuaikan dengan tujuan analisis apakah menekankan pada kecepatan atau ketepatan dalam pemodelan cluster.

Referensi

Alodia, D. A., Fialine, A. P., Endriani, D., & Widodo, E. (2021). Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan. *Sepren*, *2*(2), 1–13. https://doi.org/10.36655/sepren.v2i2.606

Elgiriyewithana, N. (2023). *Most Streamed Spotify Songs 2023*. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/6367938

Evandi, R., Sapri, S., & Elfianty, L. (2025). Clustering Data Siswa Menggunakan Metode K-

Means Untuk Mengetahui Tingkat Prestasi Akademik Di SMK Negeri 2 Kota Bengkulu. *JURNAL MEDIA INFOTAMA*, *21*(2 SE-Articles). https://doi.org/10.37676/jmi.v21i2.9601

Ikhsan Firmansyah, M., Saepul Rohman, R., & Marsusanti, E. (2023). Penerapan Algoritma Klastering K-Means Untuk Fitur Atribut Pada Layanan Streaming Musik Spotify. *Indonesian Journal Computer Science*, *2*(2), 58–66. https://doi.org/10.31294/ijcs.v2i2.2465

Mardiansa, M., Sari, H. L., & Prahasti, P. (2023). Penerapan Data Mining Untuk Mengetahui Minat Siswa Pada Pelajaran IPA Menggunakan Metode K-Means Clustering. *Jurnal Multidisiplin Dehasen (MUDE)*, *2*(4), 693–702.

Ningrum, H., Irawan, E., & Lubis, M. R. (2021). Implementasi Metode K-Medoids Clustering Dalam Pengelompokan Data Penyakit Alergi Pada Anak. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, *6*(1), 130. https://doi.org/10.30645/jurasik.v6i1.277

Nuriska, D., Irawan, B., Bahtiar, A., & Rinaldi Dikananda, A. (2024). Klasterisasi Data Lagu Terpopuler Spotify 2023 Berdasarkan Suasana Hati Menggunakan Algoritma K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3843–3850. https://doi.org/10.36040/jati.v7i6.8232

Pangestu, C., Shaufiah, S., & Wijaya, R. (2024). X Spotify Cares Clustering Analysis using K-Means and K-Medoids. *Jurnal Media Informatika Budidarma*, *8*(1), 497. https://doi.org/10.30865/mib.v8i1.7279

Prianto, C., & Bunyamin, S. (2020). *Pembuatan aplikasi clustering gangguan jaringan menggunakan metode K-Means clustering* (Vol. 1). Kreatif.

Privandhani, N. A. (2022). Clustering pop songs based on spotify data using K-Means And K-Medoids algorithm. *Jurnal Mantik*, *6*(2), 1542–1550.

Rohmah Zaidah, A., Indira Septiarani, C., Sholikhatun Nisa, M., Yusuf, A., & Wahyudi, N. (2021). Komparasi Algoritma K-Means, K-Medoid, Agglomerative Clustering Terhadap Genre Spotify. *Jurnal Ilmiah Ilmu Komputer*, *7*(1), 49–54. https://doi.org/10.35329/jiik.v7i1.186

Sari, H. L. (2023). *Penerapan Data Mining Untuk Mengetahui Minat Siswa Pada Pelajaran IPA Menggunakan Metode K-Means Clustering*. *2*(4), 693–702.

Wanto, A., Kom, M., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Kom, M., Napitupulu, D., Negara, E. S., & Kom, M. (2020). *Data Mining: Algoritma dan Implementasi*. Yayasan kita menulis.

Zahra, F., Khalif, A., Sari, B. N., Karawang, U. S., Timur, T., & Barat, J. (2024). *PROVINSI INDONESIA MENGGUNAKAN ALGORITMA K-MEDOIDS*. *12*(2).