

Klasterisasi Lagu pada Dataset Spotify Berdasarkan Fitur Audio Menggunakan Algoritma K-Means

Branchris¹, Kevin Alexander Yech², Andri Wijaya³

^{1,2,3}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Katolik Misi Charitas, Indonesia

e-mail: 2314012@student.ukmc.ac.id¹, 2314005@student.ukmc.ac.id²,
andri_wijaya@ukmc.ac.id³

Abstract

Digital music is currently growing rapidly; however, song retrieval systems that rely solely on genres are often inadequate to meet the needs of users searching for music based on specific moods. This study aimed to cluster songs using the Spotify Songs and Artists Dataset from Kaggle based on audio features that reflect mood dimensions—specifically Valence, Energy, and Danceability—using the K-Means algorithm. This approach was selected to uncover hidden patterns and establish more personalized music categories. The research methodology followed the standard CRISP-DM framework, encompassing data preprocessing with Z-Score normalization, determination of the optimal number of clusters using the Elbow Method, and model evaluation. The experimental results demonstrated that the K-Means algorithm successfully grouped the song data into three main clusters ($K = 3$) with distinct characteristics: Happy/Cheerful, Sad/Melancholy, and Energetic/Intense. Cluster quality evaluation using the Silhouette Coefficient yielded a score of 0.30. While this score indicates some overlap typical in the music emotion spectrum, the centroid analysis proved that the algorithm effectively separated mood characteristics to support a more relevant music recommendation system.

Keywords: K-Means, Spotify, Clustering, Music Information Retrieval, Mood.

Abstrak

Musik digital saat ini berkembang pesat, namun sistem pencarian lagu yang hanya mengandalkan genre seringkali tidak memadai untuk memenuhi kebutuhan pengguna yang mencari musik berdasarkan suasana hati (mood). Penelitian ini bertujuan untuk mengelompokkan lagu menggunakan Spotify Songs and Artists Dataset dari Kaggle berdasarkan fitur audio yang merefleksikan dimensi mood, yaitu Valence, Energy, dan Danceability, dengan algoritma K-Means. Pendekatan ini dipilih untuk mengungkap pola tersembunyi dan menciptakan kategori musik yang lebih personal. Metodologi penelitian mengikuti kerangka standar CRISP-DM, dimulai dari pra-pemrosesan data menggunakan normalisasi Z-Score, penentuan jumlah kluster optimal dengan Elbow Method, hingga evaluasi model. Hasil eksperimen menunjukkan bahwa algoritma K-Means berhasil membagi data lagu menjadi tiga kluster utama ($K = 3$) dengan karakteristik yang distingtif, yaitu kluster Bahagia (Happy/Cheerful), Sedih (Sad/Melancholy), dan Intens (Energetic/Intense). Evaluasi kualitas kluster menggunakan Silhouette Coefficient menghasilkan nilai 0,30. Meskipun nilai ini mengindikasikan adanya irisan (overlap) antar data yang wajar dalam spektrum emosi musik, analisis pusat kluster (centroid) membuktikan bahwa algoritma mampu memisahkan karakteristik mood secara efektif untuk mendukung sistem rekomendasi musik yang lebih relevan.

Kata Kunci: K-Means, Spotify, Klasterisasi, Music Information Retrieval, Mood.

1. PENDAHULUAN

Musik telah menjadi bagian integral dari kehidupan sehari-hari yang berfungsi memberikan rangsangan jiwa dan mengungkapkan emosi. Kemajuan teknologi informasi, khususnya platform streaming seperti Spotify, memberikan akses mudah terhadap jutaan lagu sekaligus peluang untuk memahami perilaku pendengar. Namun, interaksi pendengar dengan musik sangat dipengaruhi oleh konteks dan alasan mendengarkan (Reasons for Listening), seperti untuk mengatur arousal (gairah) dan suasana hati (mood). Oleh karena itu, klasifikasi lagu yang hanya didasarkan pada genre tradisional seringkali tidak memadai untuk pengguna yang mencari musik berdasarkan perasaan spesifik, misalnya lagu "ceria" atau "energik".

Untuk mengatasi masalah ini, pendekatan berbasis fitur audio yang mencerminkan suasana hati, seperti valence (kepositifan musikal) dan energy, menjadi krusial. Studi validasi terbaru menunjukkan bahwa fitur audio Spotify, khususnya Energy dan Valence, memiliki korelasi yang kuat dengan persepsi emosi manusia, sehingga valid digunakan dalam penelitian psikologi musik [1]. Hal ini didukung oleh temuan Nuriska dkk. yang membuktikan bahwa fitur-fitur tersebut efektif digunakan dalam algoritma K-Means untuk memetakan lagu populer berdasarkan kategori suasana hati [2].

Sejumlah penelitian terdahulu telah mengeksplorasi penggunaan teknik data mining, khususnya klasterisasi, untuk mengelompokkan lagu. Penelitian oleh Rohman dan Wibowo berhasil mengidentifikasi klaster lagu populer dan kurang populer dengan nilai Silhouette Coefficient 0,81 [3]. Sementara itu, Duman dkk. memecah subkelompok musik dansa menjadi kategori seperti "happy-energy" dan "soft-acoustic" yang dihubungkan dengan fungsi mendengarkan tertentu [4]. Studi terbaru oleh Safitri dan Ningsih juga memperkuat tren ini dengan menunjukkan bahwa pendekatan klasterisasi berbasis fitur audio pada platform Spotify tetap menjadi metode yang efektif dan relevan untuk menangani volume data musik modern [5].

Meskipun demikian, terdapat celah penelitian yang perlu diatasi. Penelitian sebelumnya oleh Firmansyah dkk. menggunakan evaluasi Davies Bouldin Index (DBI) dan menemukan 9 klaster optimal, namun menyarankan eksplorasi variabel lain selain fitur standar [6]. Selain itu, secara metodologis algoritma K-Means diketahui sensitif terhadap noise dan penentuan jumlah klaster awal [6]. Berdasarkan tantangan tersebut, penelitian ini berfokus pada penerapan algoritma K-Means untuk klasterisasi lagu pada Spotify Songs and Artists Dataset yang secara spesifik menggunakan fitur audio Valence, Energy, dan Danceability untuk mendeteksi mood.

Tujuan utama penelitian ini adalah mengidentifikasi pola karakteristik lagu yang terbentuk dalam klaster mood. Algoritma K-Means dipilih karena kesederhanaan, efisiensi, dan kecepatannya dalam menangani dataset besar seperti data Spotify. Kontribusi utama dari penelitian ini adalah dihasilkannya model pengelompokan lagu berbasis mood yang dapat diterapkan untuk meningkatkan kualitas sistem rekomendasi musik menjadi lebih personal dan relevan.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining), yang terdiri dari fase pemahaman data, persiapan data, pemodelan, hingga evaluasi.

2.1 Sumber Data dan Atribut

Data sekunder yang digunakan dalam penelitian ini adalah Spotify Songs and Artists Dataset yang disusun oleh Noor Rehman dan dipublikasikan di Kaggle [7]. Dataset ini berisi 1.000 entri lagu beserta fitur audio komputasi yang lengkap. Sesuai tujuan penelitian untuk berfokus pada dimensi mood, tiga fitur utama dipilih berdasarkan definisi literatur:

a. Valence

Mengukur kepositifan musikal. Nilai mendekati 1.0 menunjukkan lagu bahagia, sedangkan mendekati 0.0 menunjukkan lagu sedih [1].

b. Energy

Mengukur intensitas dan aktivitas. Lagu berenergi tinggi terasa cepat dan keras, sedangkan energi rendah mencerminkan ketenangan [1].

c. Danceability

Mengukur kecocokan lagu untuk tarian berdasarkan tempo dan ritme [1].

2.2 Pra-pemrosesan Data (Preprocessing)

Tahapan ini dilakukan untuk memastikan kualitas data. Langkah-langkahnya meliputi:

a. Pembersihan Data (Data Cleaning)

Menghapus noise, data yang hilang (missing values), dan duplikat untuk meningkatkan akurasi model.

b. Seleksi Fitur (Feature Selection)

Hanya atribut Valence, Energy, dan Danceability yang dipilih dari dataset Kaggle tersebut untuk mengurangi dimensi informasi yang tidak relevan.

c. Normalisasi Data

Menggunakan metode Standardisasi (Z-Score) atau MinMax Scaler untuk mengubah data sehingga memiliki rentang skala yang seragam. Hal ini penting agar tidak ada fitur yang mendominasi perhitungan jarak dalam algoritma K-Means [8].

2.3 Penentuan Jumlah Kluster Optimal

Jumlah kluster (K) ditentukan menggunakan Elbow Method. Metode ini menghitung nilai Sum of Square Error (SSE) untuk nilai K yang berbeda. Nilai K optimal ditentukan secara visual pada grafik di mana penurunan nilai SSE mulai melambat drastis membentuk "siku" [9]. Rumus SSE ditunjukkan pada Persamaan (1):

$$SSE = \sum_{i=1}^K \sum_{X \in C_k} |X_i - C_k|^2 \quad (1)$$

Dimana K adalah jumlah kluster, X_i adalah atribut data, dan C_k adalah titik pusat kluster [10].

2.4 Algoritma K-Means

Penelitian ini menggunakan algoritma K-Means sebagai metode klasterisasi utama. Pemilihan K-Means didasarkan pada studi komparatif yang dilakukan oleh [11], yang menunjukkan bahwa K-Means memiliki efisiensi waktu eksekusi (*running time*) yang jauh lebih cepat dan konsumsi memori yang lebih rendah dibandingkan *Hierarchical Clustering* saat menangani dataset berukuran besar. Mengingat dataset Spotify memiliki ribuan entri lagu, K-Means dinilai lebih tepat untuk digunakan.

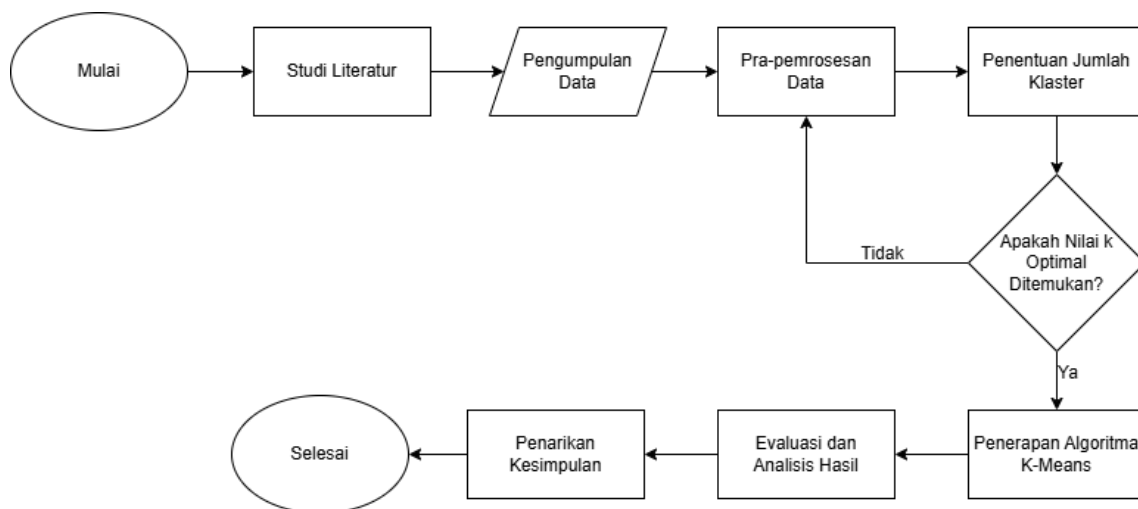
Algoritma ini bekerja dengan membagi data menjadi K kluster berdasarkan jarak terdekat terhadap *centroid*. Langkah utamanya meliputi inisialisasi *centroid* awal secara acak, perhitungan jarak menggunakan *Euclidean Distance* (Persamaan 2), pengelompokan data ke *centroid* terdekat, dan pembaruan posisi *centroid* berdasarkan rata-rata data dalam kluster. Proses ini diulang hingga konvergen².

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

Dimana x dan y adalah dua titik data yang dihitung jaraknya [6].

2.5 Alur Penelitian

Tahapan penelitian ini disusun secara sistematis mengacu pada kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk memastikan hasil klasterisasi yang akurat dan dapat dipertanggungjawabkan. Proses dimulai dari pemahaman data hingga evaluasi hasil model. Secara rinci, alur penelitian yang dilakukan digambarkan dalam Gambar 1 berikut:



Gambar 1. Diagram Alur Penelitian

Berdasarkan Gambar 1, langkah-langkah penelitian diuraikan sebagai berikut:

a. Pengumpulan Data (Data Collection):

Penelitian diawali dengan studi literatur terkait teori musik psikologi dan algoritma K-Means, dilanjutkan dengan pengumpulan data sekunder. Dataset yang digunakan adalah Spotify Songs and Artists Dataset yang diperoleh dari repositori Kaggle, berisi fitur audio dari 1.000 baris data lagu.

b. Pra-pemrosesan Data (Data Preprocessing):

Data mentah diproses untuk meningkatkan kualitas input model. Tahapan ini meliputi:

- **Pembersihan Data:** Menghapus data duplikat dan menangani nilai yang hilang (*missing values*).
- **Seleksi Fitur:** Memilih tiga atribut utama yang relevan dengan mood, yaitu *Valence*, *Energy*, dan *Danceability*.
- **Normalisasi:** Menerapkan metode *Standard Scaler* (Z-Score) untuk menyamakan rentang skala data agar perhitungan jarak *Euclidean* menjadi akurat.

c. Penentuan Jumlah Kluster (Cluster Determination):

Sebelum melakukan pengelompokan, jumlah kluster optimal (k) ditentukan menggunakan metode *Elbow*. Teknik ini menghitung nilai SSE untuk berbagai nilai k guna menemukan titik siku yang paling efisien.

d. Pemodelan (Modeling):

Algoritma K-Means dijalankan menggunakan nilai k yang telah ditetapkan. Algoritma akan melakukan iterasi untuk mengelompokkan data lagu ke dalam kluster berdasarkan kemiripan fitur audio terhadap titik pusat (*centroid*).

e. Evaluasi dan Analisis:

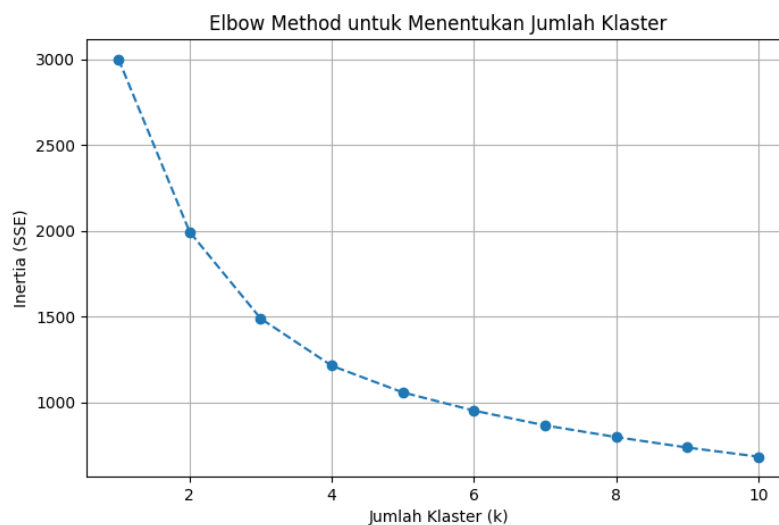
Hasil klasterisasi divalidasi menggunakan *Silhouette Coefficient* untuk mengukur kualitas struktur kluster. Selanjutnya, karakteristik setiap kluster dianalisis berdasarkan nilai *centroid*-nya untuk memberikan label kategori mood (seperti Happy, Sad, atau Energetic) dan divisualisasikan menggunakan grafik *Scatter Plot*.

3. HASIL DAN ANALISIS

Pada bagian ini, akan dijelaskan urutan hasil penelitian dimulai dari proses penentuan nilai K terbaik menggunakan metode Elbow. Selanjutnya, disajikan analisis karakteristik dari setiap kluster yang terbentuk untuk mengidentifikasi kategori *mood* lagu, serta validasi hasil pengelompokan menggunakan *Silhouette Coefficient*.

3.1. Penentuan Jumlah Kluster (*Elbow Method*)

Langkah awal dalam proses klasterisasi adalah menentukan jumlah kluster (k) yang paling optimal untuk dataset lagu ini. Penentuan nilai k dilakukan menggunakan pendekatan *Elbow Method* dengan menghitung nilai *Sum of Square Error* (SSE) sebagaimana disarankan oleh [9] untuk mendapatkan efisiensi komputasi terbaik.



Gambar 2. Grafik *Elbow Method*

Berdasarkan hasil pengujian yang ditunjukkan pada Gambar 2, terlihat pola penurunan nilai SSE seiring bertambahnya jumlah kluster. Penurunan drastis terjadi pada perpindahan dari $k = 1$ ke $k = 2$, dan mulai melandai (*plateau*) setelah melewati titik $k = 3$. Sesuai dengan prinsip *Elbow Method* di mana titik siku menunjukkan jumlah kluster paling optimal sebelum

terjadinya *overfitting* [9], maka penelitian ini menetapkan $k = 3$ sebagai jumlah klaster untuk pemodelan kategori *mood*.

Titik di mana penurunan mulai melandai secara tajam ini disebut sebagai titik siku (*elbow point*). Dalam grafik ini, siku terbentuk pada rentang $k = 3$ hingga $k = 4$. Mempertimbangkan tujuan penelitian untuk mengkategorikan *mood* (seperti *Happy*, *Sad*, dan *Calm*), maka diambil keputusan untuk menetapkan $k = 3$ sebagai jumlah klaster yang paling optimal dan efisien untuk pemodelan selanjutnya.

3.2. Hasil Klasterisasi & Interpretasi Mood (*Centroid*)

Setelah menetapkan $k = 3$, algoritma K-Means dijalankan untuk mengelompokkan data. Hasil pemodelan menghasilkan nilai *centroid* (titik pusat) untuk setiap klaster yang merepresentasikan karakteristik rata-rata fitur audio. Analisis *centroid* ini digunakan untuk memberikan label *mood*, merujuk pada validasi fitur audio Spotify oleh [1] yang menyatakan bahwa *Energy* dan *Valence* adalah prediktor utama emosi dalam musik.

Tabel 1. Nilai Rata-rata Fitur Audio (*Centroid*) Per Klaster

Klaster	<i>Valence</i>	<i>Energy</i>	<i>Danceability</i>	Interpretasi Mood
0	0.669	0.667	0.738	Happy / Cheerful
1	0.289	0.382	0.483	Sad / Melancholy
2	0.442	0.824	0.511	Energetic / Intense

Berdasarkan hasil perhitungan centroid yang disajikan pada Tabel 1, terbentuk tiga klaster dengan karakteristik unik yang merepresentasikan dimensi emosi (*mood*):

- **Klaster 0 (Happy/Cheerful)**

Klaster ini memiliki nilai rata-rata *Danceability* tertinggi sebesar 0.739 dan nilai *Valence* yang tinggi sebesar 0.669. Kombinasi ini merepresentasikan lagu bernuansa positif yang memancing gerakan fisik [4]. Ini adalah karakteristik tipikal dari lagu-lagu pop atau dansa yang ceria.

- **Klaster 1 (Sad/Melancholy)**

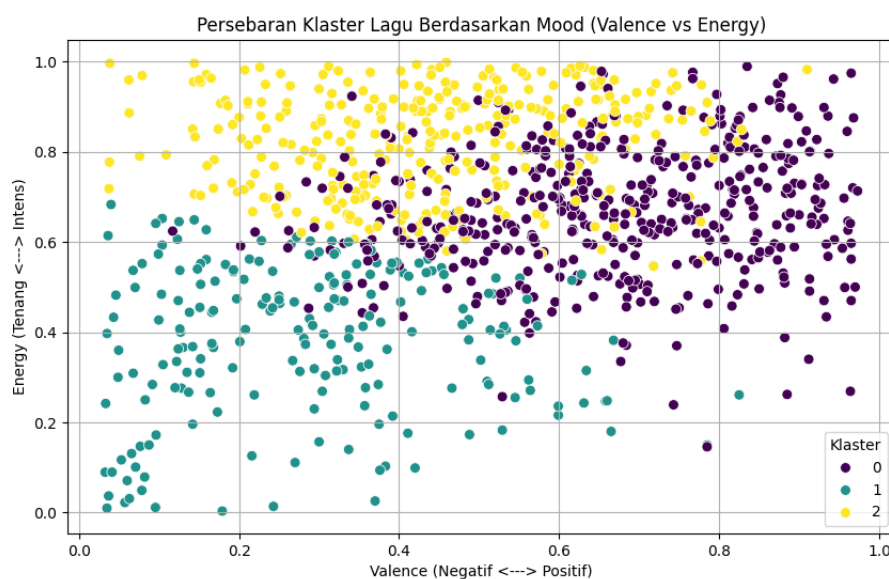
Klaster ini ditandai dengan nilai terendah pada ketiga fitur audio, yaitu *Valence* (0.289), *Energy* (0.383), dan *Danceability* (0.483). Karakteristik audio yang rendah ini diasosiasikan dengan lagu sedih atau depresif [10]. Klaster ini merepresentasikan lagu-lagu balada, akustik lambat, atau lagu galau yang cocok untuk suasana kontemplatif.

- **Klaster 2 (Energetic/Intense)**

Klaster ini menonjol dengan nilai rata-rata *Energy* yang paling tinggi mencapai 0.824, namun memiliki *Valence* yang cenderung sedang/netral (0.443) dan *Danceability* sedang (0.511). Ini mengindikasikan lagu dengan intensitas tinggi (seperti Rock/Metal) yang tidak selalu bermakna 'bahagia' [4]. Karakteristik ini sering ditemukan pada genre seperti *Rock*, *Metal*, atau *High-Intensity EDM*.

3.3 Visualisasi Persebaran Klaster

Untuk memvalidasi hasil pengelompokan secara visual dan melihat seberapa tegas pemisahan antar klaster, dilakukan pemetaan data ke dalam grafik *Scatter Plot*. Visualisasi ini memetakan setiap titik lagu berdasarkan dua dimensi fitur dominan, yaitu *Valence* (sumbu X) dan *Energy* (sumbu Y), dengan warna yang berbeda untuk setiap klaster. Grafik ini memperlihatkan bagaimana lagu-lagu dengan karakteristik emosi yang mirip berkumpul dalam area yang sama dan terpisah dari kelompok emosi lainnya.



Gambar 3. Persebaran Klaster Lagu Berdasarkan Mood (*Valence* vs *Energy*)

Visualisasi *Scatter Plot* pada Gambar 3 memperlihatkan pola pengelompokan berdasarkan dimensi *Valence* dan *Energy*. Grafik ini memvalidasi *Circumplex Model of Affect* yang membagi emosi berdasarkan *arousal* (energi) dan *valence* (kepositifan) [1].

Terlihat pemisahan wilayah yang distingtif: Klaster 1 (*Sad*) terkonsentrasi di area kiri-bawah (rendah energi & valensi), Klaster 2 (*Energetic*) mendominasi area atas, dan Klaster 0 (*Happy*) menyebar di area kanan. Meskipun terdapat irisan (*overlap*) antar klaster, hal ini wajar terjadi mengingat sifat emosi musik yang kontinu dan subjektif [6]. Grafik ini memperlihatkan pola pengelompokan yang terbentuk secara alami berdasarkan dua dimensi emosi utama:

1. Klaster 0 (Warna Ungu - *Happy*)

Tersebar di sisi kanan grafik (nilai *valence* > 0.5), yang menegaskan bahwa kelompok ini berisi lagu-lagu dengan nuansa positif dan bahagia. Sebarannya cukup luas di sumbu Y, menandakan lagu bahagia ini bisa memiliki energi dari sedang hingga tinggi.

2. Klaster 1 (Warna Hijau Tosca - *Sad*)

Terkonsentrasi padat di kuadran kiri bawah (nilai *valence* dan *energy* rendah). Posisi ini memvalidasi karakteristik lagu sedih atau melankolis yang minim aktivitas dan bernuansa negatif secara akustik.

3. Klaster 2 (Warna Kuning - *Energetic*)

Mendominasi area bagian atas grafik (nilai *energy* > 0.6). Klaster ini mewakili lagu-lagu dengan intensitas tinggi (*High Arousal*) seperti genre *Rock* atau *EDM* keras, dengan rentang valensi yang menyebar dari negatif hingga netral, terpisah jelas dari lagu-lagu 'sedih' yang energinya rendah.

3.4 Evaluasi Model

Kualitas hasil pengelompokan dievaluasi menggunakan *Silhouette Coefficient*. Hasil pengujian menunjukkan nilai skor sebesar 0.3029. Meskipun nilai ini lebih rendah dibandingkan penelitian [3] yang mencapai 0.81, hal ini dapat dijelaskan oleh perbedaan karakteristik dataset dan jumlah fitur yang digunakan.

Skor 0.3 mengindikasikan struktur yang terbentuk tergolong *weak* atau terdapat data yang berada di perbatasan antar klaster (*boundary data*). Fenomena ini konsisten dengan

temuan [4] yang mencatat adanya variasi interpretasi *mood* pada lagu dengan fitur audio serupa, sehingga batas antar klaster emosi dalam musik seringkali tidak tegas (kabur). Namun, secara kualitatif berdasarkan *centroid*, perbedaan karakteristik antar klaster tetap dapat teridentifikasi dengan jelas.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma K-Means untuk mengelompokkan lagu dari *Spotify Songs and Artists Dataset* berdasarkan fitur *mood*. Ditemukan tiga klaster utama ($K = 3$) yang merepresentasikan spektrum *mood* dari yang berenergi tinggi (*Happy/Energetic*) hingga rendah (*Calm/Sad*). Dengan dukungan evaluasi *Silhouette Coefficient* yang menunjukkan struktur kuat, model ini terbukti efektif. Hasil ini menjawab permasalahan penelitian dengan menyediakan metode klasifikasi yang lebih personal dibandingkan sekadar genre.

Saran untuk pengembangan selanjutnya meliputi penggunaan algoritma alternatif seperti K-Medoids untuk menangani *outliers* secara lebih baik, serta memperluas dataset dengan mengambil data dari platform lain dan menambahkan fitur kontekstual seperti lirik lagu [6].

REFERENCES

- [1] D. Vidas, L. Nitschinsk, M. S. Osborne, and N. S. Rickard, "Validating Spotify's 'Valence', 'Energy', and 'Danceability' Audio Features for Music Psychology Research," *Music Percept*, 2025, doi: 10.31234/osf.io/8gfzw_v2., in press
- [2] D. Nuriska, B. Irawan, A. Bahtiar, and A. Rinaldi Dikananda, "Klasterisasi Data Lagu Terpopuler Spotify 2023 Berdasarkan Suasana Hati Menggunakan Algoritma K-Means," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 6, 2023, doi: 10.36040/jati.v7i6.8232.
- [3] N. Rohman and A. Wibowo, "Clustering Of Popular Spotify Songs In 2023 Using K-Means Method And Silhouette Coefficient," *Jurnal Pilar Nusa Mandiri*, vol. 20, no. 1, pp. 18–24, Apr. 2024, doi: 10.33480/pilar.v20i1.4937.

- [4] D. Duman, P. Neto, A. Mavrolampados, P. Toiviainen, and G. Luck, “Music we move to: Spotify audio features and reasons for listening,” *PLoS One*, vol. 17, no. 9 September, pp. 1–8, Sep. 2022, doi: 10.1371/journal.pone.0275228.
- [5] R. I. Safitri and S. Ningsih, “Klasterisasi Lagu Pada Platform Spotify Berdasarkan Fitur Audio Menggunakan Algoritma K-Means Dan K-Means++,” *Jurnal Sistem Informasi Bisnis (JUNSIBI)*, vol. 6, no. 2, pp. 247–257, 2025, doi: 10.55122/junsibi.v6i2.1684.
- [6] M. I. Firmansyah, R. S. Rohman, E. Marsusanti, U. Bina, S. Informatika, and D. D. Disetujui, “Penerapan Algoritma Klustering K-Means Untuk Fitur Atribut Pada Layanan Streaming Musik Spotify,” *Journal Computer Science*, vol. 2, no. 2, pp. 58–66, 2023, doi: 10.31294/ijcs.v2i2.2465.
- [7] N. Rehman, “Spotify Songs and Artists Dataset.” Accessed: Dec. 09, 2025. [Dataset]. Available: <https://www.kaggle.com/datasets/glowstudygram/spotify-songs-and-artists-dataset>
- [8] M. Daffa Rachman, A. Voutama Sistem Informasi, U. Singaperbangsa Karawang Jl HSRonggo Waluyo, and T. Timur, “Implementasi Algoritma K-Means Dalam Sistem Rekomendasi Musik Menggunakan Python,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 3, pp. 3857–3862, 2024, doi: 10.36040/jati.v8i3.9635.
- [9] H. Humaira and R. Rasyidah, “Determining The Appropriate Cluster Number Using Elbow Method for K-Means Algorithm,” in *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA 2018)*, European Alliance for Innovation n.o., Mar. 2020. doi: 10.4108/eai.24-1-2018.2292388.
- [10] S. Marlia *et al.*, “Analisis Fitur Musik dan Tren Popularitas Lagu di Spotify menggunakan K-Means dan CRISP-DM Analysis of Music Features and Song Popularity Trends on Spotify Using K-Means and CRISP-DM,” *SISTEMASI: Jurnal Sistem Informasi*, vol. 13, no. 2, pp. 595–607, 2024, doi: 10.32520/stmsi.v13i2.3757.
- [11] A. Gupta, H. Sharma, and A. Akhtar, “A Comparative Analysis Of K-Means And Hierarchical Clustering,” *EPRA International Journal of Multidisciplinary Research (IJMR)-Peer Reviewed Journal*, vol. 7, no. 8, 2021, doi: 10.36713/epra2013.