Jurnal Informatika dan Tekonologi Komputer
E-ISSN: 2809-9230
P-ISSN: 2809-9249

Analisis Perbandingan Metode K-Means dan Gaussian Mixture Model dalam Pengelompokan Playlist Musik Berbasis Fitur Audio

Daniel Prasetiyo Dodi Darmawan [1]*, Christopher Mathew Putra [2], Samuel Yahya [3], Darren Jusman [4], Alfi Syahrian[5], Vitri Tundjungsari [6]

[1] Universitas Esa Unggul; Kota Bekasi, Provinsi Jawa Barat; e-mail : danielprasetyo053@gmail.com
[2] Universitas Esa Unggul; Kota Bekasi, Provinsi Jawa Barat; e-mail : cmp.mp14@gmail.com
[3] Universitas Esa Unggul; Kota Bekasi, Provinsi Jawa Barat; e-mail : samuelyahya25@gmail.com
[4] Universitas Esa Unggul; Kota Bekasi, Provinsi Jawa Barat; e-mail : darrenjusman@gmail.com
[5] Universitas Esa Unggul; Kota Bekasi, Provinsi Jawa Barat; e-mail : fyysyahrian10@gmail.com
[6] Universitas Esa Unggul; Kota Bekasi, Provinsi Jawa Barat; e-mail : vitri.tundjungsari@esaunggul.ac.id
* Corresponding Author : Daniel Prasetiyo Dodi Darmawan

Abstract: *Music streaming platforms rely on playlists as medium for users to store musical preferences and receive music recommendations based on the music stored. However, representing playlists as meaningful groups remains a major challenge due to the high diversity characteristics of music. In addition, the distribution of musical characteristics within playlists can vary significantly. This study aims to compare two clustering models with different approaches hard clustering using the K-Means method and soft clustering using the Gaussian Mixture Model (GMM). Playlists are represented as statistical aggregations of audio feature data from songs, such as energy, acousticness, and danceability. The hard clustering approach using K-Means produces compact and clearly separated clusters, while the Gaussian Mixture Model (GMM) generates clusters that capture playlist ambiguity, resulting in overlapping clusters due to its probabilistic nature. These differences have a direct impact on the implementation of the clustering results in downstream applications. This study emphasizes the importance of selecting an appropriate clustering method for further implementations, such as music recommendation systems, and provides insights into the trade-offs between interpretability and flexibility offered by both models.*

Keywords: *Audio Features; Clustering; Music; Spotify; Unsupervised Learning*

Abstrak: Platform streaming musik sangat bergantung pada playlist sebagai media utama untuk pengguna menyimpan selera musik dan rekomendasi musik. Namun representasi playlist sebagai intensitas yang bermakna menjadi tantangan utama karena karakteristik musik yang sangat beragam. Belum lagi mengenai distribusi karakteristik musik dalam playlist yang sangat bervariasi. Penelitian ini bertujuan untuk membandingkan dua model yang memiliki ciri khas berbeda yaitu hard clustering untuk metode K-Means dan soft clustering untuk metode Gaussian Mixture Model (GMM). Playlist direpresentasikan sebagai agregasi statistik dari data fitur audio lagu seperti energy, acousticness, danceability. Pendekatan hard clustering menggunakan metode K-Means membuat cluster yang kompak dan jelas sementara Gaussian Mixture Model (GMM) membuat cluster yang mewakili ambiguitas playlist sehingga clusternya menjadi tercampur karena berbasiskan probabilitas. Perbedaan ini berdampak langsung pada implementasi dari hasil analisis ini. Penelitian ini menegaskan pentingnya pemilihan metode clustering untuk implementasi lebih lanjut lagi seperti sistem rekomendasi musik serta memberikan wawasan mengenai interpretabilitas dan fleksibilitas kedua model.

Kata kunci: Clustering; Fitur Audio; Musik; Spotify; Unsupervised Learning

Naskah Masuk: 29 Januari 2026
Revisi: 1 Februari 2026
Diterima: 26 Februari 2026
Terbit: 17 Maret 2026
Ver. Skrg.: 17 Maret 2026

1. Pendahuluan

Musik menjadi sebuah media hiburan bagi banyak orang di masa modern ini. Adanya platform streaming musik menjadi media utama dalam konsumsi musik digital [1]. Dalam platform tersebut terdapat fitur playlist yang dapat digunakan sebagai sarana pengguna

mengatur, mengekspresikan, dan menemukan preferensi musik mereka. Kumpulan musik tersebut yang dikumpulkan dalam sebuah playlist dapat merepresentasikan perasaan pendengar karena tercermin juga dari perasaan pembuat lagu [2]. Playlist tersebut dapat digunakan oleh platform streaming musik online sebagai kunci dari memberikan rekomendasi lagu kepada pengguna sesuai dengan playlist yang telah dibuat pengguna.

Platform streaming musik yang paling populer di dunia yaitu Spotify telah mencapai 500 juta pengguna dalam satu dekade terakhir dan algoritma untuk rekomendasi musik berdasarkan playlist menjadi peran penting dalam konteks ini. Dengan algoritma APC (Automatic Playlist Continuation), maka APC dapat memberikan eksplorasi musik bagi pengguna dengan memberikan lagu atau artis baru untuk playlist tersebut [3]. Untuk membantu pengembangan algoritma tersebut, Spotify membuat data numerik khusus untuk setiap lagu yang dapat merepresentasikan lagu tersebut yaitu beberapa yang utama ada acousticness, danceability, duration, energy, instrumentalness, liveness, loudness, speechiness, tempo, dan valence [4]. Berdasarkan fitur audio yang ada, Spotify dapat menggunakannya untuk membuat algoritma yang tangguh dan relevan terhadap selera pengguna.

2. Kajian Pustaka atau Penelitian Terkait

Penggunaan fitur audio sangatlah beragam seperti mengetahui tingkat kematangan dari suatu buah kelapa [5], analisis bukti manipulasi suara untuk kebutuhan forensik [6], dan membuat sistem kunci cerdas IoT berbasis fitur audio sebagai media interaksi [7]. Hal ini membuktikan bahwa fitur audio dapat digunakan untuk keperluan yang sangat luas.

Beberapa penelitian menggunakan fitur audio seperti sistem rekomendasi musik sudah dilakukan, seperti berdasarkan kemiripan antar user menggunakan collaborative filtering [8], [9], dan penelitian [10] menggunakan playlist sebagai data user atau pengguna daripada riwayat pendengaran. Adapun penelitian mengenai klasifikasi genre musik berdasarkan dari genre [11], [12] yang membedah fitur audio dari masing-masing musik. Penelitian [13] menggunakan fitur audio Spotify untuk prediksi emosi menggunakan algoritma supervised learning yaitu Random Forest.

Berdasarkan penelitian [14] algoritma unsupervised learning K-Means dapat digunakan untuk klasterisasi lagu atau musik, artinya klasterisasi playlist juga dapat dilakukan dengan agregasi data dari fitur audio yang terdapat pada musik di dalam playlist. Penelitian [15] menggunakan Gaussian Mixture Model untuk klasifikasi musik sesuai genre digabungkan dengan MFCC dapat mencapai skor yang sangat tinggi hingga 100%. Maka dari itu, penelitian ini akan membandingkan performa klasterisasi dari model K-Means dan Gaussian Mixture Model. Manakah model yang paling sesuai untuk merepresentasikan emosi, suasana, genre atau selera pengguna dari playlist.

Analisis dan interpretasi khusus harus dilakukan karena penelitian ini berfokus pada bagaimana K-Means dan Gaussian Mixture Model melakukan clustering atau pengelompokkan. Maka dari itu, diperlukan analisis lebih lanjut mengenai isi cluster seperti makna hasil pengelompokkan yang dilakukan oleh K-Means dan Gaussian Mixture Model dan apakah interpretasi makna tersebut sesuai pada distribusi data yang ada.

3. Metode yang Diusulkan

Metode yang diusulkan dalam penelitian ini yang bertujuan untuk melakukan pengelompokan data dengan membandingkan dua pendekatan unsupervised learning, yaitu K-Means Clustering dan Gaussian Mixture Model (GMM) Clustering. Kedua metode ini digunakan untuk mengidentifikasi pola dan struktur tersembunyi dalam data berdasarkan tingkat kemiripan karakteristiknya. Dengan membandingkan kedua metode ini maka dapat dihasilkan metode yang paling optimal untuk melakukan pengelompokkan playlist berdasarkan rata-rata fitur audio yang ada.

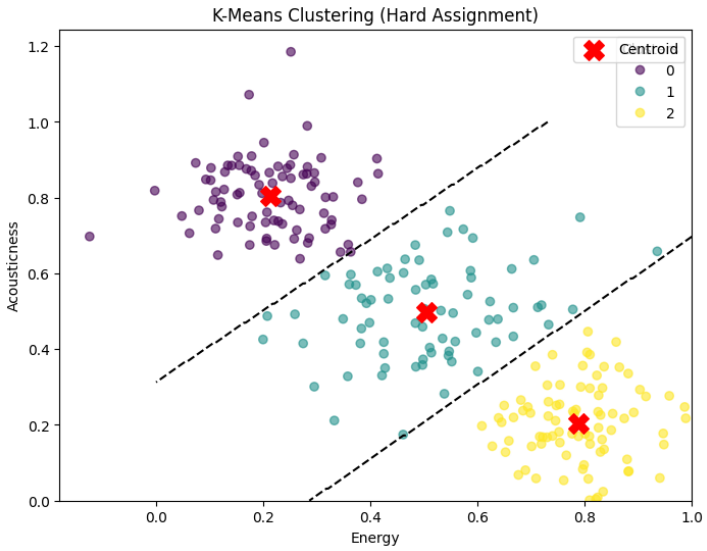
3.1. Alat dan Library Utama

Proses eksperimen ini menggunakan bahasa pemrograman Python karena terdapat berbagai library pendukung yang cukup lengkap. Library yang digunakan untuk implementasi

metode K-Means dan Gaussian Mixture Model (GMM) adalah scikit-learn. Visualisasi dan pengolahan data menggunakan library tambahan yaitu NumPy, Pandas, dan Matplotlib. Adanya proses Principal Component Analysis (PCA) untuk visualisasi menggunakan modul yang tersedia pada scikit-learn.

3.2. K-Means Clustering

K-Means Clustering merupakan salah satu metode unsupervised learning yang digunakan untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan tingkat kemiripan antar data. K-Means Clustering bertujuan untuk meminimalkan variasi intra-klaster (within-cluster variance) dan memaksimalkan perbedaan antar klaster (between-cluster variance). Metode ini bekerja dengan membagi data ke dalam k klaster yang sudah ditentukan, dimana setiap data akan dikelompokkan ke klaster dengan jarak terdekat terhadap pusat klaster (centroid).



Gambar 1. K-Means Clustering

Gambar 1 merupakan contoh gambar bagaimana K-Means membentuk cluster menggunakan data dummy. Terdapat separasi keputusan yang jelas antar cluster, jika terdapat data di dalam garis separasi tersebut maka data tersebut menjadi milik cluster tersebut tanpa pengecualian.

Proses kerja K-Means dimulai dengan menentukan jumlah klaster k, lalu menginisialisasi centroid secara acak atau menggunakan metode tertentu. Selanjutnya, setiap data akan dialokasikan ke klaster terdekat berdasarkan perhitungan jarak. Setelah seluruh data terklasifikasi, posisi centroid diperbarui dengan menghitung rata-rata dari seluruh data dalam klaster tersebut. Proses penugasan data dan pembaruan centroid dilakukan secara iteratif hingga konvergen, yang artinya ketika tidak terjadi perubahan signifikan pada posisi centroid [16].

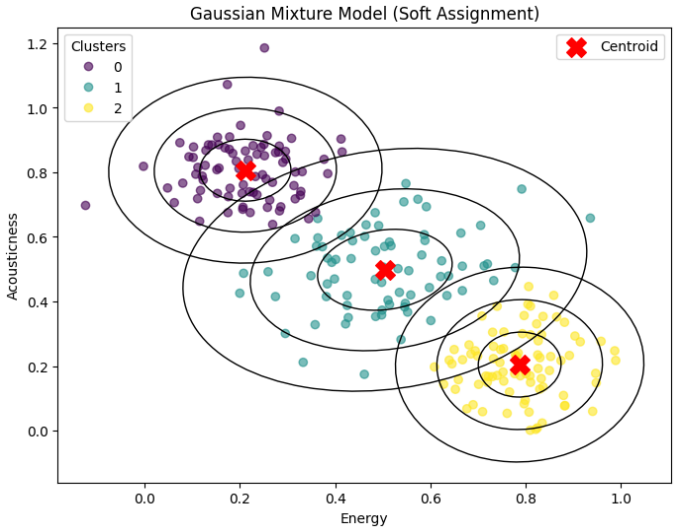
K-Means dipilih dalam penelitian ini karena kelebihannya seperti algoritma cluster yang paling sederhana dan cukup efektif untuk pengelompokkan karena sifatnya yang hard atau tegas karena memaksa titik data untuk berada di suatu cluster tertentu berdasarkan letak centroid terdekat.

3.3. Gaussian Mixture Model (GMM) Clustering

Gaussian Mixture Model (GMM) merupakan metode unsupervised learning yang digunakan untuk melakukan pengelompokan data (clustering) dengan pendekatan probabilistik. yang memodelkan data sebagai gabungan (mixture) dari beberapa distribusi Gaussian, di mana setiap distribusi tersebut merepresentasikan satu kelompok dalam data.

Distribusi komponen ini memiliki sebuah puncak atau data dalam satu cluster yang ketat, maka pola dominan dalam data ditangkap oleh distribusi komponen dengan baik [16].

Proses pembentukan cluster pada GMM dilakukan menggunakan algoritma Expectation-Maximization (EM), yaitu algoritma iteratif yang digunakan untuk mengestimasi parameter model ketika keanggotaan data terhadap cluster belum diketahui secara pasti. Pada algoritma ini terdapat 2 tahap, yaitu tahap Expectation untuk menghitung probabilitas setiap data terhadap masing-masing cluster berdasarkan parameter sementara, dan tahap Maximization untuk memperbarui parameter model supaya nilai kemungkinan data menjadi maksimum. Proses ini dilakukan secara berulang sampai model mendapat kondisi stabil [17].



Gambar 2. Gaussian Mixture Model (GMM) Clustering

Gambar 2 memperlihatkan hasil dari proses Gaussian Mixture Model (GMM) menggunakan data dummy dimana tidak ada batasan tegas antar cluster melainkan bukit yang mewakili persebaran cluster. Terdapat satu titik data yang menjadi bagian cluster 1 sementara titik data tersebut menjadi bagian cluster 0 pada gambar 1. Berdasarkan sifat probabilistik GMM, titik data tersebut menjadi bagian dari cluster 1 walaupun berdekatan dengan titik data cluster 0.

Pada gambar 2, terlihat bahwa cluster berbentuk lonjong karena mengakomodasi probabilitas sesuai dengan data pada titik tersebut berbeda dengan K-Means yaitu cluster lebih padat dan memaksa suatu titik data untuk berada di cluster berdasarkan centroid terdekat dari titik tersebut. Pendekatan soft clustering pada Gaussian Mixture Model (GMM) membuatnya menjadi lebih fleksibel dibandingkan metode K-Means, terutama dalam menangani data dengan bentuk cluster yang tidak beraturan sehingga dalam project pengelompokan playlist ini, GMM dapat membuat hasil menjadi lebih realistis. Berdasarkan penelitian [17], GMM dapat dirancang untuk menangani data yang tidak lengkap atau fitur yang absen dan lebih unggul daripada algoritma K-Means dinamis. Namun pada penelitian ini, dataset yang digunakan sudah cukup lengkap dan data yang tidak lengkap atau tidak sesuai harus difilter demi kemudahan penelitian dan kesesuaian data.

3.4. Dataset

Dataset yang digunakan dalam penelitian ini berasal dari dua sumber data yang saling melengkapi. Dataset primer adalah Spotify Million Playlist Dataset yang dirilis secara resmi oleh Spotify melalui platform AICrowd. Dataset tersebut diambil secara acak dan tidak seragam untuk memastikan rahasia bisnis Spotify tetap terlindungi [18]. Untuk efisiensi pemrosesan dan batasan komputasi, penelitian ini hanya menggunakan 50.000 playlist pertama dari total dataset yang tersedia. Dataset ini mencakup informasi struktural mengenai daftar putar yang dibuat oleh pengguna.

Sebagai data pendukung, digunakan dataset sekunder dari platform Kaggle (Spotify Dataset 1921-2020 - 600k+ Tracks) yang menyediakan karakteristik fitur suara (audio features) secara mendetail untuk setiap lagu. Dataset utama dan pendukung disatukan menggunakan track_id dari lagu yang ada pada setiap playlist

Tabel 1. Tabel Fitur

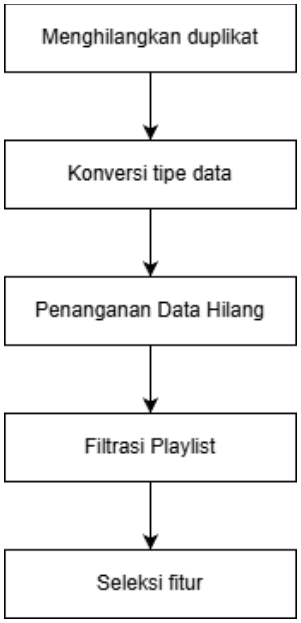
Fitur	Penjelasan
Acousticness	Ukuran keyakinan berkisar antara 0.0 hingga 1.0 digunakan untuk menentukan apakah sebuah trek memiliki sifat akustik.
Danceability	Menggambarkan seberapa sesuai sebuah lagu untuk ditarikan, berdasarkan kombinasi elemen musik seperti tempo, kestabilan ritme, kekuatan ketukan, dan keteraturan keseluruhan. Nilai 0.0 menunjukkan bahwa lagu tersebut paling tidak cocok untuk menari, sedangkan nilai 1.0 menunjukkan bahwa lagu tersebut paling cocok untuk menari.
Energy	Energy mengukur dari 0.0 hingga 1.0 dan mencerminkan persepsi intensitas dan aktivitas sebuah lagu.
Speechiness	Tingkat keberadaan kata-kata yang diucapkan dalam sebuah lagu, dengan nilai mendekati 1.0
Valence	Menggambarkan tingkat kepositifan emosi lagu pada skala 0.0-1.0 dari suasana negatif hingga ceria.
Tempo	Mempresentasikan kecepatan lagu dalam ketukan per menit (BPM)

3.5. Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kestabilan perhitungan fitur audio dan meningkatkan kualitas model. Proses ini dibagi menjadi tahap pembersihan dan tahap transformasi.

3.5.1. Pembersihan dan Penyaringan Data

Untuk menjaga stabilitas perhitungan dan relevansi analisis fitur audio, dilakukan beberapa langkah penyaringan sebagai berikut:

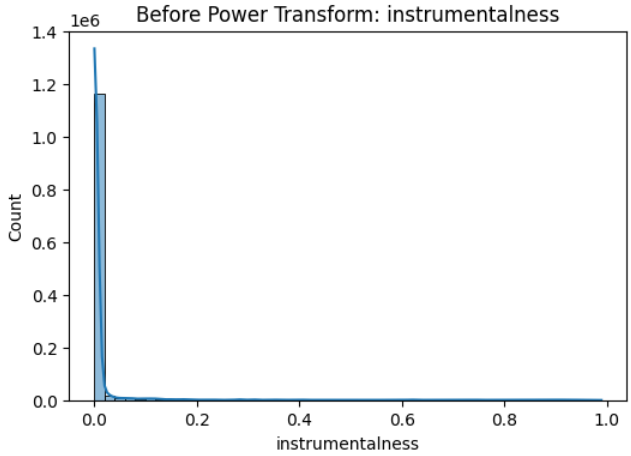


Gambar 3. Flowchart Data Preprocessing

Tahap awal dalam pra-pemrosesan dilakukan dengan membersihkan dataset dari redundansi dan inkonsistensi data seperti yang terdapat pada gambar 3. Proses ini dimulai dengan mengeliminasi entri yang terduplikasi serta memastikan seluruh fitur audio telah

dikonversi ke dalam format numerik untuk mendukung komputasi matematis. Baris yang tidak memiliki informasi fitur audio yang lengkap dihapus untuk menjaga integritas data selama fase perhitungan. Kemudian filtrasi playlist dengan minimal terdapat 15 lagu pada suatu playlist.

Fitur instrumentalness diputuskan untuk dihapus karena memiliki kemiringan yang tajam dan untuk menghilangkan noise dan bias serta mengoptimalkan clustering. Pada gambar 3, terlihat bahwa nilai instrumentalness hampir semua data bernilai mendekati 0. Hal ini tidak dapat memberikan kontribusi separasi yang jelas dan malah membebani komputasi. Musik pada Spotify jarang sekali terdapat musik dengan instrumental murni khususnya pada penelitian ini hanya mencakup 50.000 playlist yang mungkin tidak terdapat musik instrumental sama sekali.

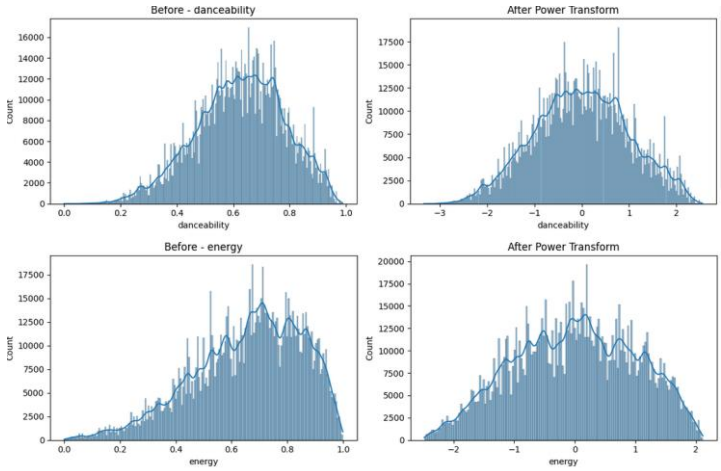


Gambar 4. Visual nilai Instrumentalness

3.5.2. Transformasi Data

Setelah data melalui tahap pembersihan, dilakukan transformasi fitur untuk menyiapkan dataset bagi algoritma clustering. Proses ini diawali dengan penerapan Power Transform Yeo-Johnson untuk mengubah distribusi fitur audio agar mendekati distribusi Gaussian atau normal [19]. Teknik ini berfungsi untuk menstabilkan varians dan meminimalisir kemiringan data sehingga menghasilkan bentuk distribusi yang lebih simetris.

Langkah selanjutnya adalah menerapkan Standard Scaler untuk mengubah skala fitur sehingga memiliki nilai rata-rata $\mu = 0$ dan standar deviasi $\sigma = 1$. Melalui standarisasi ini, seluruh fitur audio dipusatkan pada titik nol, yang secara signifikan mempermudah perhitungan jarak Euclidean pada algoritma K-Means.

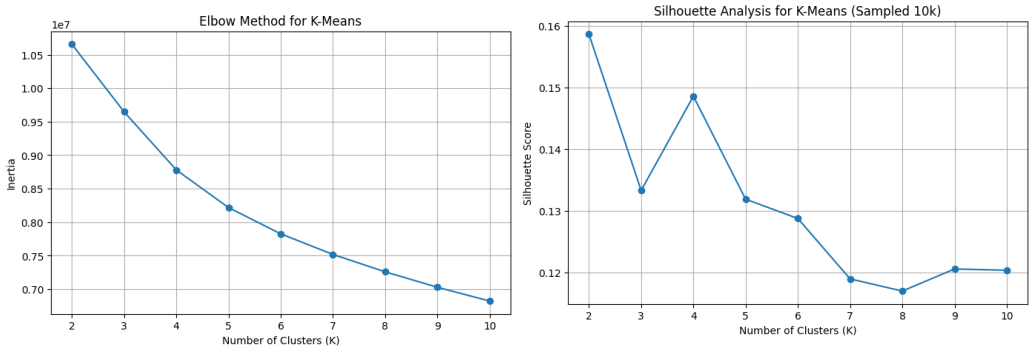


Gambar 5. Perbandingan Sebaran Distribusi Fitur Audio Sebelum dan Sesudah Tahap Transformasi

Secara keseluruhan, hasil dari tahap transformasi ini menunjukkan peningkatan kualitas data mentah yang signifikan. Meskipun masih terdapat beberapa fitur yang tidak sepenuhnya mengikuti distribusi normal sempurna karena tingkat kemiringan awal yang sangat tajam, secara umum data telah menjadi lebih baik dan siap digunakan sebagai input model clustering.

3.5.3. Persiapan Cluster

Sebelum melakukan clustering, penentuan jumlah cluster (K) sangat krusial. Dengan memilih K yang optimal untuk K-Means dan Gaussian Mixture Model (GMM) maka sumber daya yang dibutuhkan untuk melakukan clustering lebih efisien. Memilih K yang optimal juga mempermudah interpretasi dan menggali informasi dari cluster yang dibuat oleh model. Penentuan K dilakukan dengan kombinasi Elbow Method dan Silhoutte Score untuk memperoleh keputusan yang lebih objektif. Evaluasi dilakukan pada rentang K=2 hingga K=10 seperti yang ditunjukkan pada gambar 5.



Gambar 6. Elbow Method dan Silhouette Score

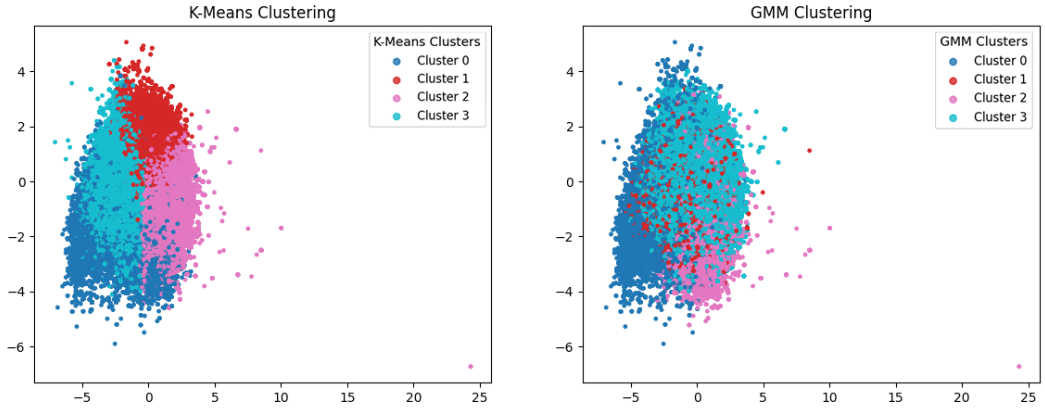
Berdasarkan grafik Elbow Method, penurunan terjadi signifikan dari K=2 hingga K=4. Setelah K=4, kurva menunjukkan pola penurunan yang lebih landai mengindikasikan kurangnya kontribusi tambahan dari penambahan cluster terhadap reduksi variasi intra-cluster. Maka dari itu K=4 merupakan titik transisi dimana kompleksitas model mulai memberikan keuntungan makin kecil. Akibat dari kompleksitas data, grafik Elbow Method tidak cukup memberikan deskripsi yang jelas untuk representasi penentuan K maka dari itu Silhouette Score dapat membantu untuk penentuan K lebih lanjut.

Pada analisis Silhouette Score, nilai tertinggi diperoleh adalah K=2. Namun K=4 menunjukkan nilai Silhouette tertinggi kedua. Meskipun K=2 memberikan separasi terbaik, jumlah cluster tersebut terlalu sederhana dan berpotensi mengaburkan variasi karakteristik playlist. K=4 memberikan keseimbangan kohesi cluster dan kemampuan model untuk menangkap struktur data yang lebih detail serta pengelompokan yang masih sesuai. Nilai K>4 pada Silhoutte Score mengalami penurunan signifikan khususnya pada K>7 mengindikasikan potensi over-segmentation pada data. Oleh karena itu, dengan mempertimbangkan dua metrik secara simultan maka K=4 dipilih sebagai jumlah cluster yang optimal dan representatif.

Jumlah cluster pada model Gaussian Mixture Model (GMM) akan mengikut jumlah cluster dari model K-Means untuk perbandingan yang adil dan lebih objektif antar hard clustering dengan soft clustering.

4. Hasil dan Pembahasan

Berdasarkan dari cluster yang dibuat oleh kedua model K-Means dan Gaussian Mixture Model (GMM) sangat berbeda. Cluster pada K-Means terlihat rapi dan terbagi dengan jelas bahwa terdapat 4 cluster berbeda dengan warna representatif masing-masing sementara cluster GMM terlihat cukup berantakan dan terlihat saling tindih satu sama lain. Cluster warna merah pada K-Means lebih jelas terlihat daripada GMM. Cluster warna merah pada GMM terlihat menyebar dan hanya nampak beberapa saja.

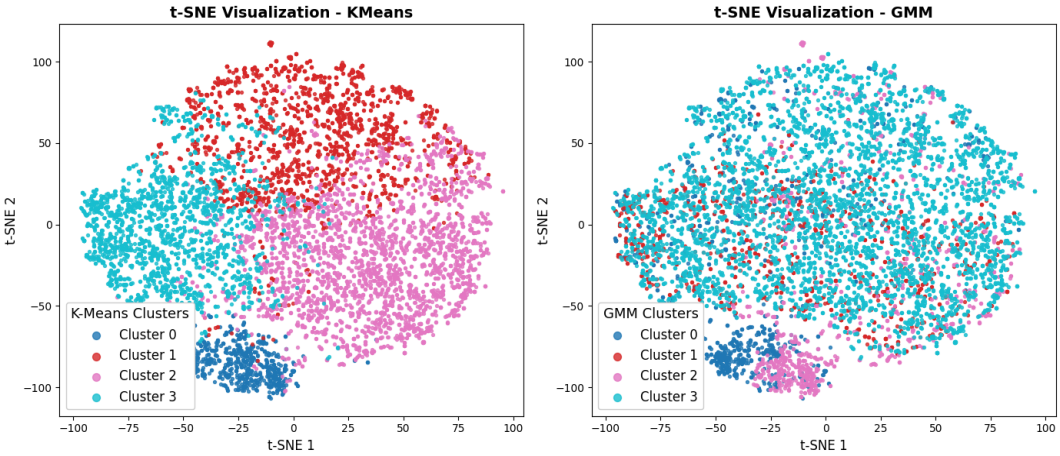


Gambar 7. Hasil Plot PCA antara K-Means dan Gaussian Mixture Model (GMM).

Hasil dari cluster merepresentasikan cara kerja dari kedua model dimana K-Means clustering memiliki ciri khas hard clustering sementara Gaussian Mixture Model (GMM) memiliki ciri khas soft clustering. Setiap data playlist pada K-Means dipetakan secara tegas berdasarkan jarak ke centroid sehingga terlihat jelas batasan dimana cluster terpisah antara satu dengan yang lain. Pola ini mencerminkan hard clustering dimana data playlist tersebut ditekan atau dipaksa masuk berdasarkan centroid terdekat dengan data. Sementara hasil cluster GMM terlihat tidak ada batasan jelas antar cluster, hal ini disebabkan karena GMM memberikan probabilitas terhadap setiap data yang ada. Akibatnya, data playlist yang memiliki karakteristik campuran berada di situasi tumpang tindih seperti terlihat pada cluster warna merah pada GMM.

4.1. Interpretasi Cluster

Hasil dari Gambar 8, cluster direpresentasikan dengan plot t-SNE untuk kejelasan visualisasi dan kemudahan interpretasi.

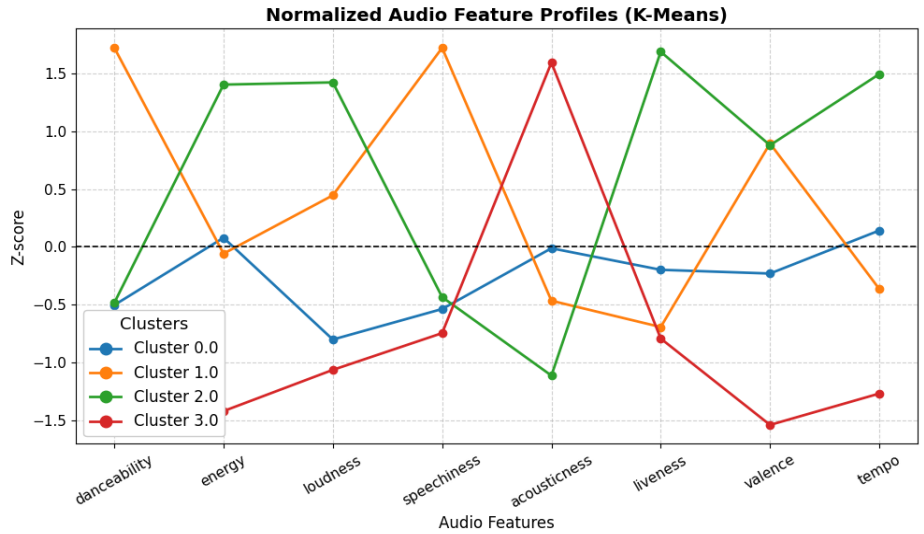


Gambar 8. Hasil Plot t-SNE antara K-Means dan Gaussian Mixture Model (GMM).

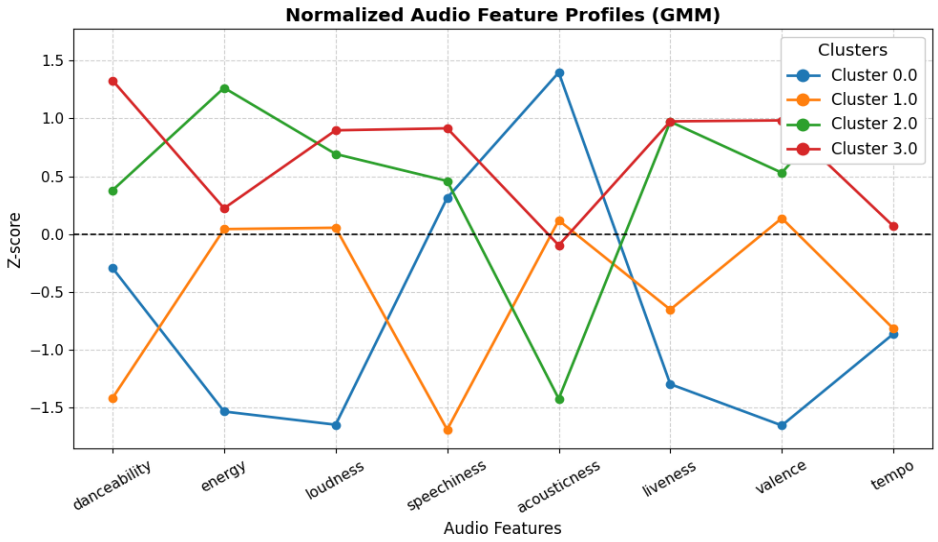
K-Means memiliki cluster yang lebih rapi di kedua plot PCA dan t-SNE sementara Gaussian Mixture Model (GMM) memiliki plot yang cukup berantakan dan didominasi oleh Cluster 3 atau cluster biru muda. Menggunakan plot t-SNE, cluster 1 atau cluster warna merah pada GMM terlihat lebih jelas bahwa data playlist ini tumpang tindih dengan cluster lain dan tersebar di sekitar tengah plot.

Dari visualisasi cluster ini, dapat disimpulkan sementara bahwa data fitur audio dari berbagai lagu pada sebuah playlist sangat beragam dan mirip-mirip terutama berdasarkan dari cluster yang tercipta oleh Gaussian Mixture Model (GMM). GMM lebih mempresentasikan kasus lagu dan kebiasaan manusia yang lebih nyata karena bercampur aduk. Fitur-fitur audio yang terkumpul di dalam suatu playlist tidak secara langsung mempresentasikan genre yang disukai namun kondisi alami dari sebuah lagu tersebut. Cluster K-Means mengelompokkan

data playlist ke dalam suatu cluster karena terdapat data dominan atau lebih dekat ke centroid clusternya. Untuk membuktikan hal ini maka diperlukan analisa lebih lanjut mengenai data setiap cluster yang ada untuk keduanya seperti representasi apa yang ingin ditunjukan oleh cluster yang dibuat oleh kedua model.



Gambar 9. Hasil Visualisasi rata-rata setiap fitur audio K-Means



Gambar 10. Hasil Visualisasi rata-rata setiap fitur audio Gaussian Mixture Model (GMM).

Gambar 9 dan 10 memvisualisasikan rata-rata setiap fitur audio tiap clusternya. Dengan visualisasi ini, maka interpretasi lebih mudah dilakukan karena data rata-rata fitur playlist sudah diskala. Salah satu cluster yang memiliki kemiripan yang sama namun berbeda cluster adalah Cluster 3 pada K-Means dan Cluster 0 pada Gaussian Mixture Model (GMM). Cluster 3 memiliki puncak di acousticness pada K-Means sementara acousticness paling tinggi dimiliki oleh Cluster 0 pada GMM. Jika dilihat berdasarkan plot t-SNE, tidak terdapat banyak playlist untuk cluster 0 karena terpisah di satu pulau yang sama dengan kebanyakan playlist cluster 2 atau cluster warna pink. Sementara playlist dengan acousticness tinggi berada di cluster 3 dan lebih banyak daripada milik model GMM.

Dari Gambar 8, plot t-SNE dalam Gaussian Mixture Model (GMM) didominasi oleh Cluster 3, karena pada grafik garis cluster 3 sekiranya berada diatas nilai setengah artinya rata-rata fitur playlist berada di nilai tengah. Artinya pada cluster ini merepresentasikan playlist dengan genre yang bercampur-campur khususnya lagu pop modern dengan fitur audio

ditengah-tengah (tidak ada dominasi fitur audio seperti cluster 0 yang tinggi akan karakter akustik).

Energy paling tinggi pada kedua cluster adalah cluster 2, dengan acoustic rendah cluster ini dapat mempresentasikan playlist dengan karakter yang asik/upbeat atau berenergi tinggi.

Tabel 2. Tabel Interpretasi Representasi Cluster

Cluster	K-Means	GMM
Cluster 0	Mewakili playlist dengan karakter campuran	Mewakili playlist dengan karakter lagu berakustik dan santai
Cluster 1	Mewakili playlist dengan karakter yang mudah didansa dan banyak kata-kata	Mewakili playlist dengan energi sedang biasa
Cluster 2	Mewakili playlist dengan karakter berenergi tinggi dan berkarakter elektronik	Mewakili playlist dengan karakter berenergi tinggi dan berkarakter elektronik
Cluster 3	Mewakili playlist dengan karakter lagu berakustik dan berkarakter santai	Mewakili playlist dengan karakter campuran

Perlu diingat bahwa interpretasi yang disajikan pada tabel 2 merupakan kesimpulan berbasis pola statistik internal tanpa validasi eksternal seperti label genre atau survei pengguna. Oleh karena itu, representasi karakter playlist pada masing-masing cluster masih bersifat dugaan awal dan memerlukan validasi lebih lanjut untuk memastikan kesesuaian dengan konteks musik yang sebenarnya.

4.2. Perbandingan Kuantitatif

Evaluasi kuantitatif berguna untuk melihat apakah cluster memiliki dasar yang kuat dan menjadi fondasi cluster yang baik. Silhouette Score membantu untuk melihat kualitas cluster berdasarkan jarak antara data di dalam clusternya sementara Davies-Bouldin Index menghitung kualitas cluster berdasarkan kepadatan dan jarak antar cluster [20].

Tabel 3. Tabel hasil Silhouette Score dan Davies-Bouldin Index

Cluster	Silhouette Score	Davies-Bouldin Index
K-Means	0.148	1.944
Gaussian Mixture Model	0.010	4.595

Nilai pada tabel 3 didapat dari sampel 25.000 data secara acak karena keterbatasan komputasi yang memakan waktu terlalu lama pada perhitungan Silhouette Score namun namun menggunakan setengah dari total data cukup untuk melakukan evaluasi secara garis besar. Evaluasi menunjukkan bahwa K-Means menghasilkan nilai 0.148 dan Davies-Bouldin Index sebesar 1.94. Sementara Gaussian Mixture Model (GMM) menghasilkan Silhouette Score sebesar 0.010 dan Davies-Bouldin Index sebesar 4.59. Nilai Silhouette yang lebih tinggi pada K-Means menunjukkan separasi cluster yang lebih baik dibandingkan GMM dimana terdapat selisih yang cukup besar yaitu 0.138. Hal ini diperkuat oleh nilai Davies-Bouldin Index yang lebih rendah sebesar 1.944 dibandingkan GMM yaitu 4.595. K-Means memiliki jarak antar cluster yang lebih jelas daripada GMM. Hasil sebaliknya berlaku pada GMM dimana adanya tumpang tindih cluster yang signifikan. Hal tersebut dikarenakan struktur data musik itu sendiri dan sifat probabilistik dari GMM yang memungkinkan terjadinya overlap antar cluster. Secara kuantitatif, K-Means menunjukkan clustering yang lebih baik untuk dataset yang digunakan.

4.3. Analisis Dispersi Relatif Cluster menggunakan Koefisien Variasi

Untuk mendukung atau menyanggah interpretasi yang sudah dibuat maka kita dapat membuat uji normalitas data untuk mengetahui apakah interpretasi hasil cluster dapat merepresentasikan data sesungguhnya yang ada menggunakan rumus Koefisien Variasi.

$$KV = \frac{s}{\bar{x}} \cdot 100\%, \quad (1)$$

Menggunakan koefisien variasi, kita dapat menguji keragaman data berdasarkan standar deviasi atau s dan rata-rata atau $\bar{x}$. Koefisien variasi berguna untuk melihat keragaman relatif data terhadap rata-ratanya, karena kita melihat data berdasarkan rata-rata maka koefisien variasi sangat cocok untuk pengujian ini. Perlu diingat bahwa Koefisien Variasi digunakan untuk menguji keragaman data dan bukan untuk menguji distribusi normal pada data. Berdasarkan Kvålseth [21], hasil koefisien variasi rendah adalah 0 sampai 0.20 atau 20% namun perlu diingat bahwa nilai tersebut adalah interpretasi heuristik untuk besaran variasi relatif sehingga tidak terdapat nilai baku universal. Sebagaimana yang telah dibahas oleh Kvålseth, interpretasi koefisien variasi harus dilakukan secara kontekstual maka dari itu koefisien variasi hanya digunakan untuk ukuran deskriptif yang harus dikombinasikan dengan analisis lain seperti yang sudah dilakukan yaitu analisis visualisasi cluster PCA dan t-SNE. Analisis lain juga harus dilakukan untuk mengakomodasi standarisasi nilai yang dilakukan seperti pada gambar 9 dan 10. Dengan kita melihat sebaran keragaman data terhadap rata-ratanya maka kita bisa membuat kesimpulan bahwa interpretasi yang dibuat berdasarkan rata-rata nilai fitur sesuai atau tidak sesuai.

Tabel 4. Tabel hasil rata-rata audio features K-Means

Cluster	Danceability	Speechiness	Acousticness	Energy
Cluster 0.0	0.576079	0.060822	0.228146	0.643930
Cluster 1.0	0.767042	0.183026	0.165939	0.627238
Cluster 2.0	0.577653	0.066394	0.077188	0.805025
Cluster 3.0	0.556549	0.049534	0.448272	0.461614

Tabel 5. Tabel hasil rata-rata audio features Gaussian Mixture Model (GMM)

Cluster	Danceability	Speechiness	Acousticness	Energy
Cluster 0.0	0.606409	0.090440	0.404440	0.536908
Cluster 1.0	0.585307	0.043679	0.229581	0.651513
Cluster 2.0	0.619032	0.093840	0.018604	0.740248
Cluster 3.0	0.636848	0.104474	0.200086	0.664564

Tabel 6. Tabel hasil standar deviasi audio features K-Means

Cluster	Danceability	Speechiness	Acousticness	Energy
Cluster 0.0	0.173951	0.293993	0.293993	0.236593
Cluster 1.0	0.109521	0.123587	0.174959	0.131749
Cluster 2.0	0.125008	0.062463	0.111410	0.101246
Cluster 3.0	0.136295	0.050284	0.266478	0.153262

Tabel 7. Tabel hasil standar deviasi audio features Gaussian Mixture Model (GMM)

Cluster	Danceability	Speechiness	Acousticness	Energy
Cluster 0.0	0.175293	0.100317	0.277316	0.225674
Cluster 1.0	0.152326	0.016210	0.265863	0.212395
Cluster 2.0	0.170099	0.084648	0.019821	0.161475
Cluster 3.0	0.149480	0.106306	0.227838	0.179471

Dari Tabel 4 dan 5, tabel tersebut adalah hasil rata-rata dari fitur audio yang dapat merepresentasikan gaya musik yang ada di sebuah playlist. Pada distribusi statistik playlist tersebut kita akan lihat apakah benar mempresentasikan playlist dominan akustik pada Cluster

3 K-Means dan Cluster 0 Gaussian Mixture Model (GMM). Cluster 3 K-Means memiliki rata-rata acoustic 0.44 sementara Cluster 0 Gaussian Mixture Model (GMM) memiliki acousticness 0.40. Melakukan perhitungan Koefisien variasi 0,26 / 0,44 = 0,59 atau 59% dan 0,27 / 0,40 = 0,67 atau 67%. Datanya sangat beragam sehingga melabelkan cluster tersebut memiliki ciri khas acoustic kurang tepat karena data tersebut cukup beragam.

Selanjutnya energy tertinggi artinya cluster tersebut merepresentasikan playlist yang tinggi energi yaitu Cluster 2 pada dua model. Cluster 2 k-means memiliki energy 0,80 dan Gaussian Mixture Model (GMM) memiliki 0,74 dibagi dengan standar deviasi masing-masing. Melakukan perhitungan Koefisien Variasi 0,1 / 0,8 = 0,125 atau 12,5% dan 0,16 / 0,74 = 0,21 atau 21%. Nilai KV yang didapat jauh lebih kecil. Kita dapat mendeskripsikan bahwa cluster 2 memiliki banyak playlist dengan dominan energy tinggi.

Berdasarkan dari hasil perhitungan koefisien variasi sederhana tersebut kita sudah bisa mendapatkan deskripsi mengenai cluster bahwa cluster dengan fitur acousticness tertinggi belum bisa merepresentasikan bahwa cluster tersebut didominasi playlist acoustic dan cluster dengan energy tinggi merepresentasikan bahwa cluster tersebut didominasi oleh playlist dengan energy tinggi.

Tabel 8. Tabel Data Kuartil dari Fitur audio K-Means

Cluster	Energy_q25	Energy_q50	Energy_q75	Acousticness_q_25	Acousticness_q50	Acousticness_q75
Cluster 0.0	0.501	0.688	0.830	0.009915	0.0815	0.3450
Cluster 1.0	0.534	0.632	0.720	0.032000	0.1030	0.2450
Cluster 2.0	0.729	0.814	0.884	0.006870	0.0319	0.0991
Cluster 3.0	0.359	0.475	0.574	0.224000	0.4210	0.6690

Tabel 9. Tabel Data Kuartil dari Fitur audio Gaussian Mixture Model (GMM)

Cluster	Energy_q25	Energy_q50	Energy_q75	Acousticness_q_25	Acousticness_q50	Acousticness_q75
Cluster 0.0	0.373	0.532	0.714	0.15900	0.3300	0.6150
Cluster 1.0	0.509	0.680	0.824	0.01640	0.1060	0.3690
Cluster 2.0	0.627	0.760	0.876	0.00189	0.0102	0.0311
Cluster 3.0	0.547	0.687	0.804	0.02710	0.1060	0.2950

Analisis selanjutnya adalah memecah data rata-rata menjadi beberapa bagian kuartil. Dengan memecah data fitur maka kita dapat mengetahui apakah interpretasi yang dihasilkan dapat mewakili data yang ada. Berdasarkan analisis sebelumnya acoustic tertinggi pada K-Means yaitu cluster 3 dan Gaussian Mixture Model (GMM) yaitu cluster 0, pada tabel 7 dan 8 terlihat bahwa acousticness pada K25 atau kuartil 25 bernilai 0.22 dan mengalami kenaikan menjadi 0.42 pada K50 dan menjadi 0.66 pada K75. Hal ini berarti bahwa terdapat selisih ~0.2 dimana distribusi data pada bagian kecil data tidak dapat merepresentasikan bahwa ia playlist akustik. Hal yang sama juga berlaku pada Gaussian Mixture Model (GMM).

Cluster energy tertinggi untuk K-Means dan Gaussian Mixture Model (GMM) yaitu cluster 2 memiliki nilai sangat tinggi di K25 bernilai 0.729 dan hal tersebut berlaku juga pada Gaussian Mixture Model (GMM) dengan nilai 0.627. Pada K50 keduanya juga hanya memiliki selisih ~0.1. Cluster K-Means lebih baik menangkap playlist dengan energy tinggi karena nilai K25 K-Means lebih tinggi daripada Gaussian Mixture Model (GMM). Ini menunjukkan bahwa cluster 2 merepresentasikan energy tinggi pada kedua model karena distribusi data yang cukup pada ketiga bagian K25, K50, dan K75.

Tabel 10. Tabel Lagu Playlist ID 34066 Cluster K-Means

track_name	artist_name	danceability	energy
Starships	Nicki Minaj	0.747	0.716

Firework	Katy Perry	0.638	0.832
Some Nights	fun.	0.672	0.738
Uptown Funk	Mark Ronson	0.856	0.609
Good Feeling	Flo Rida	0.706	0.890

Tabel 11. Tabel Lagu Playlist ID 4305 Cluster Gaussian Mixture Model (GMM)

track_name	artist_name	danceability	energy
Last Night	The Vamps	0.629	0.759
Hallelujah	Leonard Cohen	0.280	0.344
Rude	MAGIC!	0.774	0.756
Imagine - 2010 - Remaster	John Lennon	0.547	0.257
The Days	Avicii	0.590	0.724

Tabel 10 menunjukkan contoh lagu pada sebuah playlist acak yang diambil dari cluster 2. Cluster K-Means lebih baik menangkap energy tinggi karena semuanya diatas 0.6 sementara cluster Gaussian Mixture Model (GMM) terdapat 2 lagu yang memiliki energy rendah yaitu 0.344 dan 0.257. Perlu diingat bahwa daftar lagu tersebut hanyalah sebagian dari banyak lagu dari playlist tersebut. Namun berdasarkan analisis yang telah dilakukan hal ini sesuai dan berlaku karena analisis kuartil yang telah dilakukan.

5. Perbandingan

Perbandingan hasil menunjukkan bahwa metode K-Means unggul dalam mengelompokkan playlist dengan karakteristik fitur audio yang dominan dan jelas. Metode ini menghasilkan cluster yang mudah diinterpretasikan dan memiliki batasan yang tegas antar kelompok namun ada beberapa hal seperti pada pendekatan K-Means memiliki keterbatasan dalam merepresentasikan playlist dengan karakter campuran karena setiap playlist hanya dapat menjadi anggota satu cluster secara mutlak dikarenakan kurang sesuai dengan perilaku pengguna Spotify yang sering menyusun playlist bernuansa musik dalam satu daftar putar.

Disisi lain metode Gaussian Mixture Model (GMM) mampu menangkap kompleksitas tersebut melalui pendekatan probabilistik dan memungkinkan sebuah playlist memiliki keanggotaan pada lebih dari satu cluster,sehingga lebih mencerminkan pengguna yang memiliki keragaman karakter musik dalam satu playlist walaupun hasil clusternya terlihat kurang terstruktur.Hal ini sangat menggambarkan kondisi alami data musik yang tidak memiliki batasan yang tegas antar karakter.

6. Keterbatasan Penelitian

Terdapat beberapa keterbatasan penelitian karena seperti hanya menggunakan 50.000 ribu data playlist daripada dataset lengkap yaitu 1 juta playlist. Walaupun sudah menggunakan 50.000 data, komputasi untuk menghasilkan perhitungan kuantitatif seperti Silhouette Score memakan waktu yang cukup lama maka dari itu hanya menggunakan setengah yaitu 25.000 dan diambil secara acak. Dengan playlist yang terbatas tersebut, hasil cluster tidak lengkap dan dapat mempengaruhi hasil akhir yang mungkin dapat mengubah interpretasi.

Ketergantungan fitur audio pada Spotify yang mungkin bias karena fitur ini merupakan hasil ekstraksi algoritmik dari Spotify yang tidak diketahui proses lengkapnya. Dengan demikian, representasi karakter musik bergantung pada komputasi yang dilakukan oleh Spotify. Hal ini dapat menimbulkan bias terhadap interpretasi karakter playlist.

Evaluasi yang dilakukan penelitian ini hanya berasal dari metrik internal seperti Silhouette Score dan Davies-Bouldin Index. Interpretasi cluster yang dihasilkan berasal dari pola statistik rata-rata fitur tanpa validasi eksternal seperti genre label resmi playlist atau survei pengguna. Oleh karena itu, representasi playlist masih bersifat inferensial dan memerlukan pengujian lebih lanjut untuk memastikan kesesuaian yang sebenarnya.

7. Kesimpulan

Penelitian ini membandingkan pendekatan hard clustering K-Means dan soft clustering Gaussian Mixture Model (GMM) untuk representasi playlist musik pada Spotify. Setiap playlist memiliki nilai numerik yang merefleksikan karakteristik musik di dalamnya yang dibuat oleh Spotify.

Hasil eksperimen menunjukkan bahwa K-Means lebih baik mengelompokkan dengan nilai yang jelas dan dominan, clusternya terbentuk dengan jelas, memiliki nilai kuantitatif yang lebih baik, dan memiliki variasi internal yang cukup rendah khususnya untuk cluster 2. Cluster K-Means lebih baik merepresentasikan playlist dengan karakter musik yang homogen atau mirip. Sebaliknya GMM menghasilkan cluster yang lebih tumpang tindih daripada K-Means. Nilai variasi yang lebih besar mengartikan bahwa model ini menangkap ambiguitas dan variasi karakter lagu lebih baik daripada K-Means. Sifat probabilistik Gaussian Mixture Model GMM memungkinkan sebuah playlist terkait dengan cluster lain sehingga cluster terlihat tumpang tindih.

Perbedaan cluster K-Means dan Gaussian Mixture Model (GMM) bukan menunjukkan mana model yang lebih baik dan mana yang lebih buruk namun menegaskan bahwa perbedaan tersebut mencerminkan tujuan dari clustering. Analisis mendalam mengenai cluster juga penting untuk mengetahui kesesuaian data dengan cluster yang dihasilkan. Pemilihan metode harus disesuaikan dengan kebutuhan analisis khususnya interpretasi cluster yang tegas atau fleksibel. Perlu diingat bahwa setiap interpretasi dari clustering harus menyesuaikan konteks dan data yang diolah supaya interpretasi tidak ambigu. Dalam konteks ini, data yang digunakan kurang seragam dan ciri khas alami musik membuat data menjadi cukup kompleks sehingga dibutuhkan analisis lebih lanjut.

Cluster-cluster ini dapat berguna untuk membuat klasifikasi karakter playlist sehingga dapat meningkatkan performa sistem rekomendasi tanpa harus mengeksplorasi semua lagu dalam playlist. Pengelompokkan ini juga dapat digunakan untuk membuat garis besar selera pengguna berdasarkan playlist yang telah dibuat. Data tersebut bisa dimanfaatkan oleh platform musik digital untuk membuat algoritma khusus atau bisnis mereka tersendiri.

Ucapan Terima Kasih : Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya khususnya kepada penyedia data yang dibutuhkan dan dokumentasi fitur audio dari Spotify serta bimbingan dan arahan kepada dosen pembimbing dan rekan akademik sehingga penelitian ini dapat dijalankan pada sampai tahap akhir.

Daftar Pustaka

[1] A. Ramadhanti, M. Nursaif, and A. M. I. Taufik, "Motivasi Penggunaan Spotify Sebagai Media Penyebarluasan Karya Musik Musisi Indie Lokal," *Pros. Ind. Res. Workshop Natl. Semin.*, vol. 10, no. 1, pp. 904–916, Aug. 2019, doi: 10.35313/irwns.v10i1.1489.

[2] R. Fasti and D. Avianto, "Implementasi Web Klasifikasi Suasana Hati Berdasarkan Potongan Lagu dengan Memanfaatkan Convolutional Neural Network," *J. Indones. Manaj. Inform. Dan Komun.*, vol. 5, no. 1, pp. 881–892, Jan. 2024, doi: 10.35870/jimik.v5i1.573.

[3] T.-H. Ito and H. Shiokawa, "An Effective Graph-based Music Recommendation Algorithm for Automatic Playlist Continuation," in *Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, in ASONAM '23. New York, NY, USA: Association for Computing Machinery, Mar. 2024, pp. 459–463. doi: 10.1145/3625007.3627322.

[4] Spotify, "Web API Reference: Get Audio Features," Spotify for Developers, 2025. [Online]. Available: https://developer.spotify.com/documentation/web-api/reference/get-audio-features.

[5] P. R. Oktaviani, B. H. Iswanto, and H. Suhendar, "KARAKTERISTIK FITUR SUARA KETUKAN BUAH KELAPA BERDASARKAN DOMAIN WAKTU DAN DOMAIN FREKUENSI," *Jt. Pros. IPS Dan Semin. Nas. Fis.*, vol. 12, pp. 11–16, Jan. 2024, doi: 10.21009/03.1201.FA02.

[6] Y. Azzery, "ANALISIS STATISIK PERBANDINGAN MANIPULASI SUARA DAN SUARA ASLI MENGGUNAKAN TEKNIK AUDIO FORENSIK," *TEKNOKOM*, vol. 3, no. 1, pp. 29–33, Apr. 2020, doi: 10.31943/teknokom.v3i1.50.

[7] M. L. M. Akhlak, N. Istifarani, R. Sulistiyo, and S. Anggraeni, "Perancangan Sistem Kunci Cerdas Berbasis IoT Dan RFID Dengan Fitur Audio Interaktif," *KETIK J. Inform.*, vol. 3, no. 02, pp. 51–62, Nov. 2025, doi: 10.70404/ketik.v3i02.316.

[8] F. Yuniardini and T. Widiyaningtyas, "Analisis Perbandingan Pearson Correlation dan Cosine Similarity pada Rekomendasi Musik berbasis Collaborative Filtering," *Edumatic J. Pendidik. Inform.*, vol. 8, no. 2, pp. 555–564, Dec. 2024, doi: 10.29408/edumatic.v8i2.27781.

[9] I. Yoshua and H. Bunyamin, "Pengimplementasian Sistem Rekomendasi Musik Dengan Metode Collaborative Filtering," *J. STRATEGI - J. Maranatha*, vol. 3, no. 1, pp. 1–16, Apr. 2021, [Online]. Available: https://strategi.it.maranatha.edu/index.php/strategi/article/view/220.

[10] M. J. Yonathan, D. P. D. Darmawan, A. D. Rachmawanto, and A. Prabowo, "Sistem Rekomendasi Musik Berdasarkan Playlist dengan Collaborative Filtering," *J. Inform. Dan Tekonologi Komput. JITEK*, vol. 5, no. 3, pp. 175–184, Nov. 2025, doi: 10.55606/jitek.v5i3.8039.

[11] R. Abrah, M. A. M. Hayat, and L. Lukman, "Penerapan Machine Learning untuk Mengklasifikasikan Genre Musik Berdasarkan Fitur Audio," *Arus J. Sains Dan Teknol.*, vol. 3, no. 2, pp. 204–211, Oct. 2025, doi: 10.57250/ajst.v3i2.1699.

[12] C. N. Silla, A. L. Koerich, and C. A. A. Kaestner, "A Machine Learning Approach to Automatic Music Genre Classification," *J. Braz. Comput. Soc.*, vol. 14, no. 3, pp. 7–18, Sep. 2008, doi: 10.1007/BF03192561.

[13] N. D. Marsya, M. M. Munadhil, M. A. Dzakyananta, K. Amaliah, and D. Rofianto, "Penerapan Algoritma Random Forest dalam Prediksi Emosi Musik Berdasarkan Karakteristik Fitur Audio Spotify," *J. Sains Inform. Terap.*, vol. 4, no. 2, pp. 435–440, Aug. 2025, doi: 10.62357/jsit.v4i2.648.

[14] R. I. Safitri and S. Ningsih, "KLASTERISASI LAGU PADA PLATFORM SPOTIFY BERDASARKAN FITUR AUDIO MENGGUNAKAN ALGORITMA K-MEANS DAN K-MEANS++," *J. Sist. Inf. Bisnis JUNSIBI*, vol. 6, no. 2, pp. 247–257, Oct. 2025, doi: 10.55122/junsibi.v6i2.1684.

[15] M. D. Rumpumbo and I. M. W. Wirawan, "Optimasi Model Gaussian Mixture Model (GMM) untuk Klasifikasi Genre Musik Berbasis Mel-Frequency Cepstral Coefficients (MFCC)," *J. Nas. Teknol. Inf. Dan Apl.*, vol. 4, no. 1, pp. 153–160, Nov. 2025, doi: 10.24843/JNATIA.2025.v04.i01.p17.

[16] V. Tundjungsari, *Dasar Machine Learning (Edisi Revisi)*. Sleman, Yogyakarta: Deepublish, 2024.

[17] Y. Zhang *et al.*, "Gaussian Mixture Model Clustering with Incomplete Data," *ACM Trans Multimed. Comput Commun Appl*, vol. 17, no. 1s, p. 6:1-6:14, Mar. 2021, doi: 10.1145/3408318.

[18] C.-W. Chen, P. Lamere, M. Schedl, and H. Zamani, "Recsys challenge 2018: automatic music playlist continuation," in *Proceedings of the 12th ACM Conference on Recommender Systems*, in RecSys '18. New York, NY, USA: Association for Computing Machinery, Sep. 2018, pp. 527–528. doi: 10.1145/3240323.3240342.

[19] M. Riani, A. C. Atkinson, and A. Corbellini, "Automatic robust Box–Cox and extended Yeo–Johnson transformations in regression," *Stat. Methods Appl.*, vol. 32, no. 1, pp. 75–102, Mar. 2023, doi: 10.1007/s10260-022-00640-7.

[20] N. A. Yolandari, L. E. Butarbutar, G. C. H. Rajagukguk, M. F. Zulfi, Arnita, and F. Ramadhani, "ANALISIS PERBANDINGAN K-MEANS DAN DBSCAN DALAM PENGELOMPOKAN DATA TRAVEL REVIEW RATINGS MENGGUNAKAN EVALUASI SILHOUETTE INDEX DAN DAVIES-BOULDIN INDEX," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 3, Jul. 2025, doi: 10.23960/jitet.v13i3.6884.

[21] T. O. Kvålseth, "Coefficient of variation: the second-order alternative," *J. Appl. Stat.*, vol. 44, no. 3, pp. 402–415, Feb. 2017, doi: 10.1080/02664763.2016.1174195.