Jurnal Informatika
Vol. 25, No. 2, Desember 2025, pp.98-110
P-ISSN:1693-3877 / E-ISSN: 2407-1544
DOI:10.30873

Perbandingan K-Means dan GMM Untuk Analisis Popularitas Lagu Spotify Indonesia

K. Z. A. Vanefi[1], N. N. Afifah[2], C. Bella[3], M. S. Wisnubroto[*4], F. Farid[5]
[1,2,3,4,5]*Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Jl. Terusan Ryacudu, Way Hui, Kec. Jati Agung, Kab. Lampung Selatan, Lampung 35365*

E-mail: khaalishah.122450034@student.itera.ac.id, nisrina.122450052@student.itera.ac.id, cintya.122450066@student.itera.ac.id, syamsuddin.wisnubroto@sd.itera.ac.id, fajri.farid@sd.itera.ac.id

Abstract — ***The development of the music industry in Indonesia, especially through streaming platforms such as Spotify, creates challenges for record labels and musicians in making decisions regarding release timing and song promotion strategies. This study aims to compare two clustering algorithms, namely K-Means and Gaussian Mixture Model (GMM), in analyzing song popularity on Spotify Indonesia based on data such as total streams, peak streams, number of daily charts, and number of daily Top 10s. The results show that K-Means produces more accurate and easily interpretable clusters with a Silhouette Score of 0.731, while GMM has a Silhouette Score of 0.256, indicating less than optimal cluster separation. These findings indicate that K-Means is more suitable for data-driven decision making in the music industry, particularly for determining song release timing and promotional strategies. This research is useful for record labels, artist managers, and streaming platforms in designing more accurate and data-driven decisions.***

Keywords—Clustering; GMM; K-Means; Silhouette; Song Popularity

Abstrak — **Perkembangan industri musik di Indonesia, terutama melalui platform streaming seperti Spotify, menciptakan tantangan bagi label rekaman dan musisi dalam membuat keputusan terkait waktu rilis dan strategi promosi lagu. Penelitian ini bertujuan untuk membandingkan dua algoritma klasterisasi, yaitu K-Means dan Gaussian Mixture Model (GMM), dalam menganalisis popularitas lagu di Spotify Indonesia berdasarkan data seperti total streams, peak streams, jumlah chart harian, dan jumlah Top 10 harian. Hasil penelitian menunjukkan bahwa K-Means menghasilkan klaster yang lebih tepat dan mudah diinterpretasikan dengan Silhouette Score 0.731, sementara GMM memiliki Silhouette Score 0.256, menandakan pemisahan klaster yang kurang optimal. Temuan ini menunjukkan bahwa K-Means lebih cocok digunakan untuk pengambilan keputusan berbasis data dalam industri musik, khususnya untuk menentukan waktu rilis lagu dan strategi promosi. Penelitian ini bermanfaat bagi label rekaman, manajer artis, dan platform streaming dalam merancang keputusan yang lebih tepat dan berbasis data.**

***Kata kunci*— GMM; Klasterisasi; K-Means; Popularitas Lagu; Silhouette**

I. PENDAHULUAN

Perkembangan pesat industri musik di Indonesia selama satu dekade terakhir telah mengalami transformasi yang signifikan seiring dengan meningkatnya penggunaan *platform streaming*. Pesatnya perkembangan industri ini telah mengubah cara masyarakat menikmati musik dan membuka peluang baru bagi label rekaman dan musisi [1]. Media layanan *streaming* musik saat ini salah satunya merupakan aplikasi Spotify [2]. Dengan semakin banyaknya lagu yang dirilis setiap hari, tantangan utama yang dihadapi adalah bagaimana menentukan waktu rilis yang tepat dan strategi promosi yang efektif [3]. Keputusan-keputusan ini semakin kompleks karena berdasarkan data yang terbatas, sering kali hanya mengandalkan intuisi atau pengalaman sebelumnya. Oleh karena itu, penting untuk melakukan analisis berbasis data untuk memahami pola popularitas lagu secara lebih mendalam [4].

Dalam upaya melakukan analisis berbasis data tersebut, pemilihan metode klasterisasi menjadi aspek penting untuk memperoleh hasil yang akurat dan mudah diinterpretasikan dalam upaya melakukan analisis berbasis data tersebut, pemilihan metode klasterisasi menjadi aspek penting untuk memperoleh hasil yang akurat dan mudah diinterpretasikan [5]. Dari berbagai metode yang tersedia, K-Means dan *Gaussian Mixture Model* (GMM) merupakan dua pendekatan yang umum digunakan dalam pemodelan data kompleks. Metode K-Means memiliki keunggulan pada efisiensi komputasi serta kemampuannya

menghasilkan batas klaster yang tegas, sehingga sesuai digunakan untuk data dengan pemisahan yang relatif jelas antar kelompok [6]. Sementara itu, GMM menawarkan fleksibilitas lebih tinggi karena mampu memodelkan distribusi data yang tidak sepenuhnya linear melalui pendekatan probabilistik. Dengan mempertimbangkan karakteristik tersebut, kedua metode ini relevan untuk digunakan dalam analisis popularitas lagu yang memiliki pola bervariasi dan dinamis.

Penelitian ini bertujuan untuk membandingkan dua algoritma klasterisasi, yaitu K-Means dan *Gaussian Mixture Model* (GMM), dalam menganalisis data performa lagu di Spotify Indonesia. Dengan menggunakan indikator seperti total *streams*, *peak streams*, jumlah hari di *chart*, dan jumlah hari di Top 10, penelitian ini bertujuan untuk mengidentifikasi pola-pola yang mencirikan popularitas lagu. Keunggulan dari metode K-Means terletak pada kemampuannya untuk menghasilkan klaster yang lebih jelas dan mudah diinterpretasikan, sedangkan GMM memberikan pendekatan yang lebih fleksibel namun cenderung kurang terpisah dengan jelas.

Penelitian ini sangat penting bagi label rekaman, musisi independen, dan platform *streaming* seperti Spotify. Mereka dapat memanfaatkan hasil penelitian ini untuk merancang strategi rilis lagu yang lebih terarah dan berbasis data. Pemanfaatan dari penelitian ini terletak pada penggunaan algoritma klasterisasi dalam konteks analisis data *streaming* lagu, yang dapat memberikan wawasan baru dalam pengambilan keputusan strategis di industri musik digital Indonesia. Dengan memberikan pemahaman yang lebih baik mengenai segmen popularitas lagu, penelitian ini diharapkan dapat membantu para pelaku industri musik dalam mengoptimalkan keputusan mereka, serta meningkatkan efisiensi promosi dan perencanaan rilis lagu.

II. METODE PENELITIAN

Metode penelitian ini disusun berdasarkan landasan teori klasterisasi serta langkah-langkah teknis analisis data yang digunakan untuk mengidentifikasi pola popularitas lagu di Spotify Indonesia. Bagian ini menggabungkan kajian pustaka mengenai algoritma yang digunakan dengan metodologi pelaksanaan penelitian, sehingga memberikan gambaran menyeluruh mengenai dasar konsep, alasan pemilihan metode, serta prosedur analisis yang diterapkan secara objektif dan mendukung pengambilan keputusan strategis dalam industri musik digital.

A. Dasar Teoretis dan Pemilihan Metode

Penelitian ini memanfaatkan dua algoritma klasterisasi, yaitu K-Means dan *Gaussian Mixture Model* (GMM), yang secara luas digunakan dalam analisis data tanpa label. Berbagai penelitian terdahulu menunjukkan bahwa K-Means unggul dari sisi kecepatan, kesederhanaan, dan ketegasan batas klaster, sehingga efektif untuk data yang memiliki pemisahan kelompok yang jelas. Sementara itu, GMM menawarkan pendekatan probabilistik yang lebih fleksibel, mampu memodelkan klaster yang tumpang tindih dan berbentuk kompleks melalui distribusi Gaussian.

K-Means merupakan metode klasterisasi berbasis partisi yang bekerja secara iteratif untuk membagi data ke dalam sejumlah kelompok (klaster) dengan meminimalkan jarak antara data dan titik pusat klaster atau centroid. Proses pengelompokannya dilakukan dengan menghitung jarak Euclidean dari setiap data ke *centroid*, kemudian data akan masuk ke klaster dengan jarak terkecil.

Menurut peneliti [7], algoritma K-Means dimulai dengan penentuan centroid awal secara acak. Selanjutnya, setiap objek data diassign ke klaster berdasarkan jarak terdekat, dan proses perhitungan jarak serta pembaruan centroid dilakukan secara iteratif hingga perubahan posisi centroid menjadi sangat kecil atau algoritma mencapai konvergensi.

Jarak antar data pada K-Means dihitung menggunakan *Euclidean Distance*, dengan rumus:

$$D(X_1, X_2) = \sqrt{\sum_{i=1}^{p}(X_{1i} - X_{2i})^2} \quad (1)$$

Keterangan:

D : Lambang *Euclidean Distance*
X : Banyak objek
p : Jumlah data yang akan di *record*

K-Means memiliki beberapa keunggulan seperti waktu eksekusi yang cepat, mudah diimplementasikan, dan mampu mengurangi kompleksitas data [8]. Namun menurut peneliti [8], K-Means memiliki beberapa keterbatasan, yaitu:

a. Centroid awal ditentukan secara acak.
b. Jumlah klaster (K) harus ditentukan di awal dan harus tepat.
c. Tidak memiliki kemampuan untuk menginisialisasi centroid secara optimal.

Dalam konteks penelitian ini, teori K-Means menjadi dasar penting karena algoritma ini dianggap mampu menghasilkan klaster yang tegas, jelas, dan mudah diinterpretasikan pada data popularitas lagu Spotify yang memiliki batas pemisahan yang cukup kuat.

K-Means bekerja dengan mengelompokkan data berdasarkan jarak Euclidean terhadap centroid, menjadikannya metode yang cepat dan mudah diinterpretasikan. Namun, metode ini sensitif terhadap penentuan jumlah klaster (K) serta posisi centroid awal. Sebaliknya, GMM mengukur probabilitas keanggotaan suatu data dalam setiap klaster, memberikan soft assignment yang merepresentasikan kedekatan data terhadap distribusi Gaussian pada tiap klaster. Keunggulan ini membuat GMM cocok untuk data dengan struktur non-linear, meskipun membutuhkan komputasi lebih tinggi.

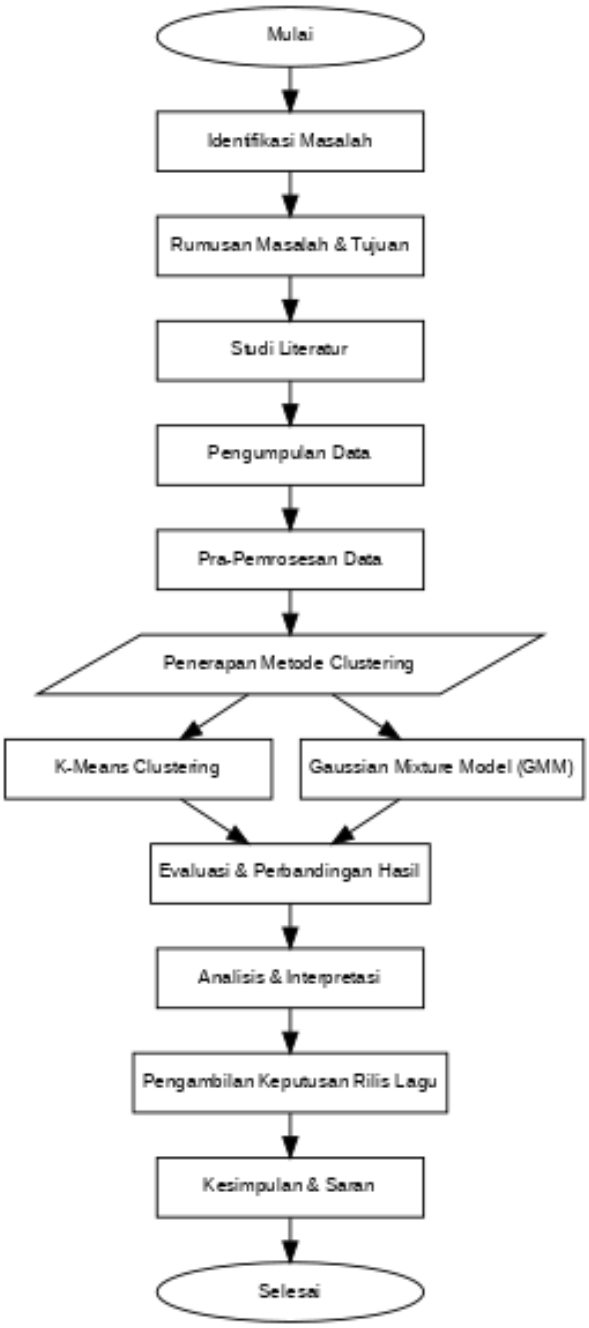
Berdasarkan karakteristik kedua metode, penelitian ini memilih K-Means dan GMM karena keduanya mampu memberikan perspektif yang berbeda dalam pemetaan popularitas lagu, sekaligus memungkinkan evaluasi komparatif yang relevan bagi pengambilan keputusan strategi rilis lagu dan flowchart promosi.

B. Flowchart Penelitian

Penelitian ini diawali dengan identifikasi masalah, studi literatur, serta pengumpulan dan pra pemrosesan data. Data kemudian dianalisis menggunakan dua metode clustering, yaitu K-Means dan *Gaussian Mixture Model* (GMM). Hasil kedua metode dievaluasi dan dibandingkan untuk menentukan performa terbaik, kemudian diinterpretasikan sebagai dasar pengambilan keputusan rilis lagu. Penelitian diakhiri dengan penyusunan kesimpulan dan saran.

Flowchart pada Gambar 1 menggambarkan alur penelitian yang dimulai dari proses identifikasi masalah hingga tahap penarikan kesimpulan. Penelitian diawali dengan mengidentifikasi isu terkait kebutuhan analisis popularitas lagu di Spotify Indonesia, kemudian dirumuskan masalah dan tujuan penelitian untuk membandingkan metode K-Means dan *Gaussian Mixture Model* (GMM). Selanjutnya dilakukan studi literatur untuk memperkuat landasan teori dan memilih metodologi yang tepat. Data sekunder dari Spotify kemudian dikumpulkan dan melalui tahapan pra-pemrosesan, meliputi pembersihan data, pemilihan fitur, penanganan nilai hilang, serta normalisasi. Setelah data siap, dua metode klasterisasi diterapkan, yaitu K-Means dan GMM, yang kemudian dievaluasi menggunakan *Silhouette Score*, *Davies–Bouldin Index* (DBI), dan *Calinski–Harabasz Index* (CHI) untuk menilai kualitas klaster. Hasil klasterisasi dianalisis dan diinterpretasikan guna memahami pola popularitas lagu. Temuan yang diperoleh selanjutnya digunakan sebagai dasar rekomendasi dalam pengambilan keputusan rilis lagu dan strategi promosi. Penelitian diakhiri dengan penyusunan kesimpulan serta saran untuk pengembangan penelitian selanjutnya.

P-ISSN:1693-3877 / E-ISSN: 2407-1544
DOI:10.30873



Gambar 1. Flowchart Penelitian

C. Sumber Data dan Teknik Pengumpulan

Data penelitian ini merupakan data sekunder yang diperoleh dari platform Spotify Indonesia melalui berkas *spotify_id_daily_totals.csv*, yang berisi informasi mengenai *Artist* dan *Title*, *Total Streams*, *Peak Streams*, jumlah hari berada di chart (*Days*), serta jumlah hari berada di Top 10 (*T10*). Teknik pengumpulan data dilakukan melalui pengunduhan data agregat serta studi literatur, kemudian seluruh data diolah menggunakan Python untuk mendukung proses analisis klaster.

D. Tahap Pra-Pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan kualitas data sebelum analisis, dimulai dari pemilihan fitur yang hanya menggunakan atribut numerik yang relevan, yaitu *Total Streams*, *Peak Streams*, *Days*, dan *T10*. Selanjutnya dilakukan pembersihan data dengan cara menghapus karakter non-

angka pada *Peak Streams* dan *Total Streams*, mengonversi seluruh nilai menjadi numerik, mengganti nilai hilang pada *T10* menggunakan median, serta menghapus atau memperbaiki data yang dianggap invalid. Setelah itu, seluruh fitur dinormalisasi menggunakan *StandardScaler* agar memiliki skala yang seragam ($\mu = 0, \sigma = 1$), sehingga tidak ada fitur yang mendominasi proses klasterisasi.

E. Implementasi Algoritma K-Means

Implementasi algoritma K-Means dilakukan melalui beberapa tahap, dimulai dari penentuan jumlah klaster optimal menggunakan *Elbow Method* yang menunjukkan bahwa $K = 4$ merupakan pilihan terbaik. Model kemudian dikonfigurasi dengan parameter *n_clusters* = 4, *random_state* = 42, dan *n_init* = 10, menghasilkan label klaster untuk setiap data yang selanjutnya dianalisis secara deskriptif. Validasi model dilakukan menggunakan beberapa metrik, yaitu *Silhouette Score* sebesar 0.731 yang menunjukkan kualitas klaster yang sangat baik, serta evaluasi tambahan melalui *Davies–Bouldin Index (DBI)* dan *Calinski–Harabasz Index (CHI)* untuk memperkuat keandalan hasil klasterisasi.

F. Implementasi *Gaussian Mixture Model* (GMM)

Implementasi *Gaussian Mixture Model* (GMM) diawali dengan pra-pemrosesan khusus, yaitu imputasi nilai hilang menggunakan mean serta normalisasi fitur seperti pada K-Means. Model kemudian dikonfigurasi dengan *n_components* = 4, *covariance_type* = "full", dan *random_state* = 42, menghasilkan *soft clustering* berupa probabilitas keanggotaan setiap data terhadap masing-masing klaster. Validasi menggunakan metrik yang sama dengan K-Means menunjukkan bahwa GMM memperoleh *Silhouette Score* sebesar 0.256, nilai *Davies–Bouldin Index* lebih tinggi, dan *Calinski–Harabasz Index* lebih rendah, yang mengindikasikan bahwa klaster yang dihasilkan GMM cenderung lebih tumpang tindih dibandingkan K-Means.

G. Perbandingan dan Implikasi Metode

Hasil perbandingan menunjukkan bahwa K-Means menghasilkan klaster yang lebih tegas dan mudah diinterpretasikan, sehingga lebih sesuai untuk decision making dalam konteks penentuan strategi promosi dan perilisan lagu. GMM tetap berguna sebagai model pembanding untuk memahami pola distribusi data yang tidak linear, namun kurang optimal untuk segmentasi populer berdasarkan data streaming Spotify.

III . HASIL DAN PEMBAHASAN

A. Hasil Pre-Processing DataFormat Teks

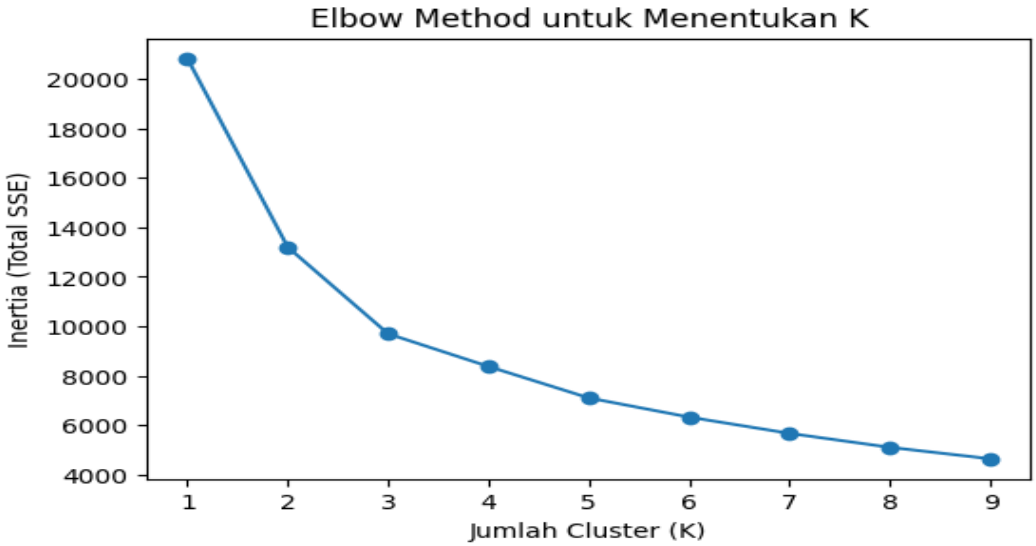
Setelah melalui tahap pre-processing, diperoleh dataset lagu Spotify Indonesia yang telah bersih, lengkap, dan siap digunakan untuk analisis klasterisasi. Seluruh variabel utama yaitu Total Streams, Peak Streams, Days, dan T10 telah dikonversi ke dalam bentuk numerik serta dinormalisasi menggunakan metode StandardScaler sehingga memiliki skala yang sebanding.

Dataset hasil pre-processing ini tidak mengandung nilai kosong, duplikat, maupun outlier ekstrim yang dapat mempengaruhi hasil analisis. Dengan demikian, data telah terstandarisasi dan siap digunakan dalam penerapan algoritma K-Means dan *Gaussian Mixture Model* (GMM) untuk mengidentifikasi tingkat popularitas lagu secara akurat.

B. Penentuan Jumlah Cluster Optimal

Jumlah klaster optimal dalam penelitian ini ditentukan menggunakan Metode Elbow yang bertujuan untuk menentukan jumlah klaster (K) optimal yang mencapai keseimbangan antara jumlah klaster dan nilai inersia (Total *Sum of Squared Errors*). Nilai inersia menunjukkan jarak kuadrat total antara setiap titik data dan pusat klasternya; semakin kecil nilai inersia, semakin baik kualitas pemisahan antar data dalam klaster tersebut.

P-ISSN:1693-3877 / E-ISSN: 2407-1544
DOI:10.30873



Gambar 2. Grafik Elbow

Berdasarkan Gambar 2, penurunan inertia terlihat sangat tajam dari $K = 1$ ke $K = 2$, kemudian masih menurun namun mulai melambat pada $K = 3$ dan $K = 4$, serta cenderung stabil setelahnya ($K > 4$). Titik 'siku' (elbow) tampak pada rentang $K = 3$ hingga $K = 4$, sehingga jumlah klaster optimal dipilih pada rentang tersebut karena penambahan klaster setelahnya tidak memberikan penurunan inertia yang signifikan.

Dari hasil analisis, teridentifikasi bahwa titik elbow berada pada kisaran $K = 3$ hingga $K = 4$. Pada rentang $K = 1$ hingga $K = 3$, kurva menunjukkan penurunan yang sangat tajam dengan gradien yang curam, mengindikasikan bahwa penambahan cluster pada fase ini memberikan peningkatan kualitas clustering yang substansial. Sebaliknya, setelah melewati $K = 4$, penurunan nilai WCSS menjadi relatif landai, menunjukkan bahwa penambahan cluster lebih lanjut hanya memberikan improvement yang marginal terhadap kualitas clustering. Fenomena ini mengindikasikan terjadinya diminishing returns, di mana manfaat penambahan cluster tidak sebanding dengan peningkatan kompleksitas model.

Berdasarkan pertimbangan trade-off antara kualitas clustering dan kompleksitas model, dipilih K =4 sebagai jumlah cluster optimal. Pemilihan ini didasarkan pada prinsip parsimoni dalam pemodelan, di mana model yang lebih sederhana namun tetap memberikan hasil yang memadai lebih diutamakan dibandingkan model yang kompleks dengan peningkatan performa yang tidak signifikan. Dengan menggunakan empat cluster, model mampu mengelompokkan data dengan tingkat homogenitas internal yang baik sambil menghindari *overfitting* yang dapat terjadi jika menggunakan terlalu banyak kluster.

C. Hasil Klasterisasi K-Means

Penerapan algoritma K-Means pada data lagu Spotify Indonesia menghasilkan empat klaster utama yang mampu membedakan karakteristik pola popularitas lagu berdasarkan empat indikator utama, yaitu Total Streams, Peak Streams, Days, dan T10. Proses klasterisasi dilakukan secara iteratif hingga mencapai kestabilan (konvergen), dengan hasil akhir berupa pembagian data lagu ke dalam empat kelompok dan posisi pusat klaster (*cluster centers*) yang merepresentasikan karakteristik masing-masing kelompok.

1. Pusat Klaster

Cluster Centers yang diperoleh dari proses K-Means ditunjukan pada Tabel 1. *Cluster* ini merepresentasikan hasil masing-masing lagu berdasarkan popularitas.

Tabel 1. Pusat Klaster Popularitas Lagu

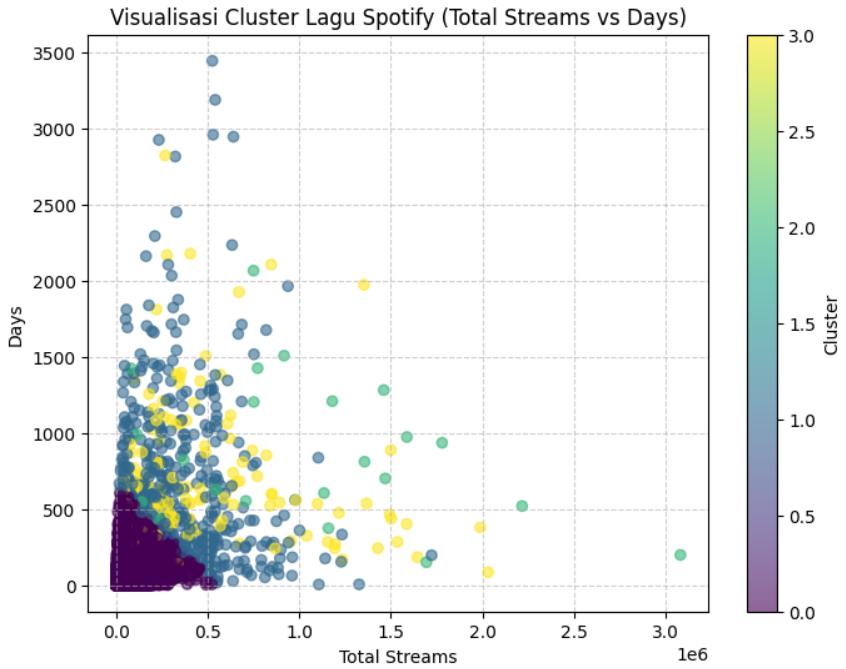
Klaster	Total Streams	Peak Streams	Days	T10	Interpretasi
0	59171.64	0.07	60.25	27.05	Lagu dengan popularitas rendah (*streams* dan *peak* rendah, jarang bertahan lama).
1	375698.51	0.75	724.5	39.05	Lagu dengan popularitas sedang (*streams* sedang, jarang masuk Top 10).
2	948984.6	76.4	822.6	227.96	Lagu dengan popularitas tertinggi (*streams* dan *peak* sangat tinggi, sering masuk Top 10).
3	523484.42	19.82	746.75	134.96	Lagu dengan popularitas tinggi (*streams* dan *days* tinggi, cukup sering masuk Top 10).

Interpretasi dari pusat klaster menyatakan bahwa Klaster 4 merepresentasikan lagu dengan popularitas tertinggi di Spotify Indonesia, sedangkan Klaster 0 mengindikasikan lagu dengan popularitas yang sangat rendah dan lagu jarang bertahan lama. Klaster 1 juga mengindikasikan lagu dengan popularitas rendah tetapi streams nya sedang dan terkadang masuk Top 10. Klaster 2 berperan sebagai kondisi yang stabil tetapi tidak tinggi dengan streams dan peak yang tinggi serta sering masuk Top 10. Validitas pemilihan empat klaster didukung oleh hasil evaluasi pada Gambar 2, dimana nilai *Silhouette Coefficient* menunjukkan bahwa $k = 4$ merupakan jumlah klaster yang paling optimal dalam memisahkan popularitas lagu Spotify di Indonesia.

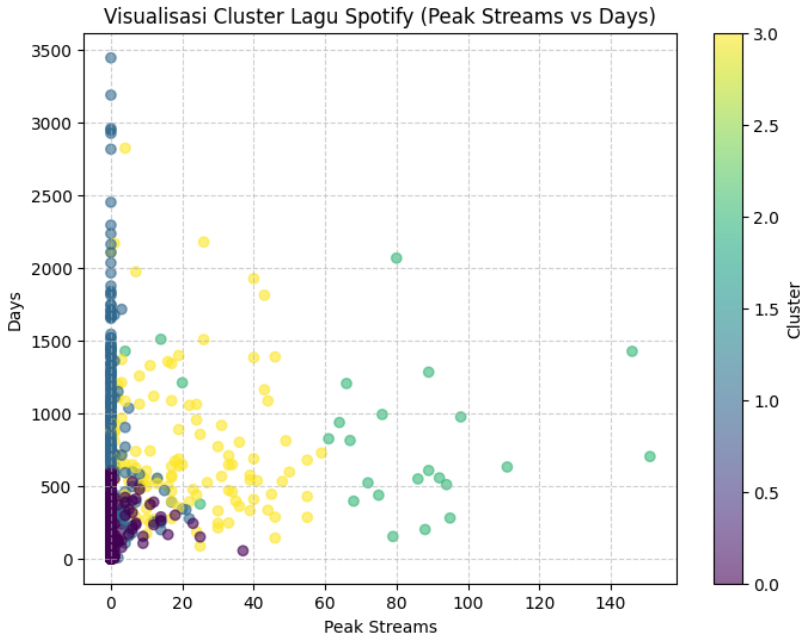
2. Hasil Klaster Popularitas Lagu

Hasil dari *streaming* lagu di Spotify digunakan untuk menentukan jumlah klaster popularitas lagu. Gambar 3 menampilkan hasil popularitas lagu berdasarkan streaming lagu di Spotify.

P-ISSN:1693-3877 / E-ISSN: 2407-1544
DOI:10.30873



***Gambar 3**. Scatter Plot Total Streams vs Days*



Gambar 4. *Scatter Plot Peak Streams vs Days*

Gambar 3 menampilkan *scatter plot* antara Total Streams dan Days yang menggambarkan bagaimana lagu-lagu Spotify Indonesia dikelompokkan ke dalam beberapa klaster. Titik-titik data tersebar dengan dominasi pada area Total Streams yang relatif rendah, sementara sebagian kecil lagu memiliki jumlah *streams* yang sangat tinggi. Warna pada plot menunjukkan klaster yang berbeda, di mana setiap klaster merepresentasikan pola karakteristik tertentu. Secara umum terlihat bahwa lagu dengan *Total Streams* rendah hingga sedang memiliki variasi lama bertahan di chart yang lebih luas, sedangkan lagu dengan *streams* sangat tinggi cenderung memiliki *Days* yang beragam namun jumlahnya lebih sedikit.

Visualisasi ini membantu memahami pola distribusi dan hubungan antara popularitas lagu (berdasarkan *streams*) dan durasi keberadaannya di chart Spotify.

Gambar 4 menunjukkan *scatter plot* antara Peak Streams dan Days yang menggambarkan bagaimana lagu-lagu Spotify dikelompokkan berdasarkan puncak jumlah *streams* harian dan lamanya bertahan di chart. Warna pada titik-titik menunjukkan klaster yang berbeda. Distribusi data tampak terpusat pada nilai *Peak Streams* yang rendah, menandakan bahwa sebagian besar lagu hanya mencapai puncak *streams* yang tidak terlalu tinggi. Meskipun demikian, terdapat beberapa lagu yang memiliki *Peak Streams* jauh lebih besar, meskipun jumlahnya sangat sedikit. Pola pada visualisasi ini memperlihatkan bahwa lagu dengan puncak *streams* tinggi tidak selalu berada di chart dalam waktu yang lama, sementara lagu dengan puncak rendah justru menunjukkan variasi *Days* yang lebih luas. Grafik ini membantu memahami hubungan antara performa puncak lagu dan daya tahannya di chart Spotify.

Dari Scatter Plot diatas, terlihat bahwa setiap cluster memiliki persebaran yang relatif terpisah. Klaster 2 berada di area dengan nilai Total Streams dan Peak Streams yang tinggi, sedangkan klaster 0 berada pada area dengan nilai rendah untuk kedua fitur tersebut. Hal ini menunjukkan bahwa algoritma K-Means berhasil mengelompokkan lagu berdasarkan tingkat popularitasnya dengan cukup baik.

D. Hasil Klasterisasi *Gaussian Mixture Model* (GMM)

Berbeda dengan *K-Means* yang menggunakan pendekatan jarak *Euclidean* untuk membentuk *cluster*, *GMM* menggunakan pendekatan probabilistik, di mana setiap data memiliki peluang (*probability*) untuk menjadi anggota dari setiap cluster. Dengan jumlah cluster yang sama ($K = 4$), *GMM* menghasilkan pusat distribusi dan karakteristik *cluster* yang serupa dengan hasil *K-Means*. Namun, pada *GMM* terdapat tumpang tindih antar *cluster* karena data dapat memiliki probabilitas keanggotaan di lebih dari satu cluster.

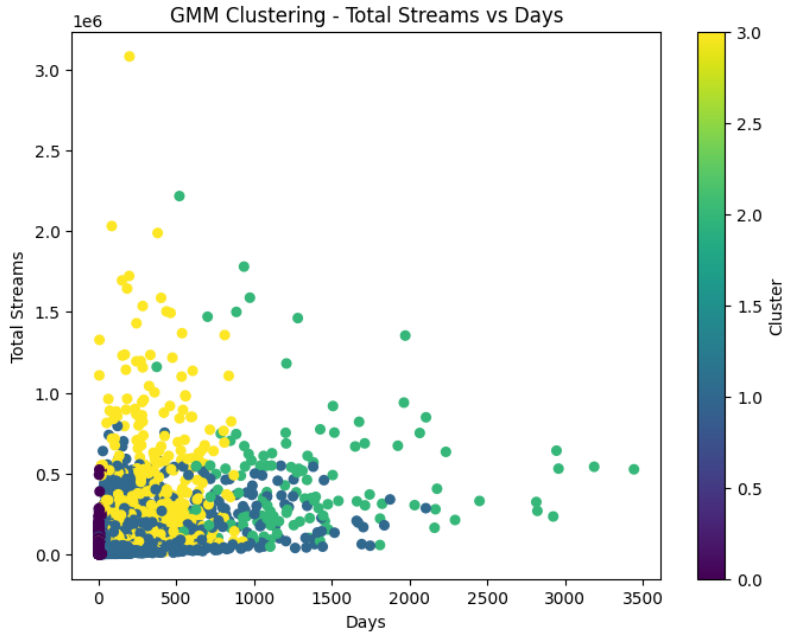
Tabel 2. *Pusat Klaster dengan Metode GMM*

Klaster	Total Streams	Peak Streams	Days	T10	Interpretasi
0	61240.52	0.12	65.88	28.47	Lagu dengan popularitas sangat rendah; jarang diputar dan tidak bertahan lama
1	401250.36	1.10	715.32	40.28	Lagu dengan popularitas sedang; cenderung stabil namun tidak viral
2	939852.11	74.98	828.45	225.63	Lagu dengan popularitas tertinggi; memiliki performa dan daya tarik tertinggi di Spotify
3	517615.72	18.70	752.18	131.22	Lagu dengan popularitas tinggi; populer tetapi tidak seintens klaster 2

Tabel 2 menampilkan pusat klaster hasil pemodelan *Gaussian Mixture Model* (GMM) beserta interpretasinya. Klaster 0 menggambarkan lagu dengan popularitas sangat rendah, ditunjukkan oleh nilai *Total Streams* dan *Peak Streams* yang kecil, serta waktu bertahan di chart yang singkat; lagu-lagu pada klaster ini umumnya kurang dikenal dan cepat menghilang dari chart. Klaster 1 berisi lagu dengan tingkat popularitas sedang, yang memiliki performa stabil namun tidak menunjukkan indikasi viral, terlihat dari nilai *Days* dan *T10* yang berada pada tingkat menengah. Klaster 2 mencerminkan lagu dengan popularitas tinggi, ditandai oleh *Total Streams* dan *Peak Streams* yang besar, serta durasi lama berada di chart; lagu dalam klaster ini biasanya memiliki daya tarik kuat dan bertahan lama di Spotify. Sementara itu, klaster 3 menunjukkan lagu yang cukup populer namun tidak sekuat klaster 2, dengan performa yang baik tetapi stabil pada tingkat yang sedikit lebih rendah.

P-ISSN:1693-3877 / E-ISSN: 2407-1544
DOI:10.30873

Hasil dari *GMM* menunjukkan pola yang konsisten dengan *K-Means*, namun batas antar cluster lebih halus dan tidak tegas karena adanya probabilitas keanggotaan yang berbeda pada setiap data. Hal ini menjadikan *GMM* lebih fleksibel dalam mengelompokkan data dengan distribusi yang tidak seragam.



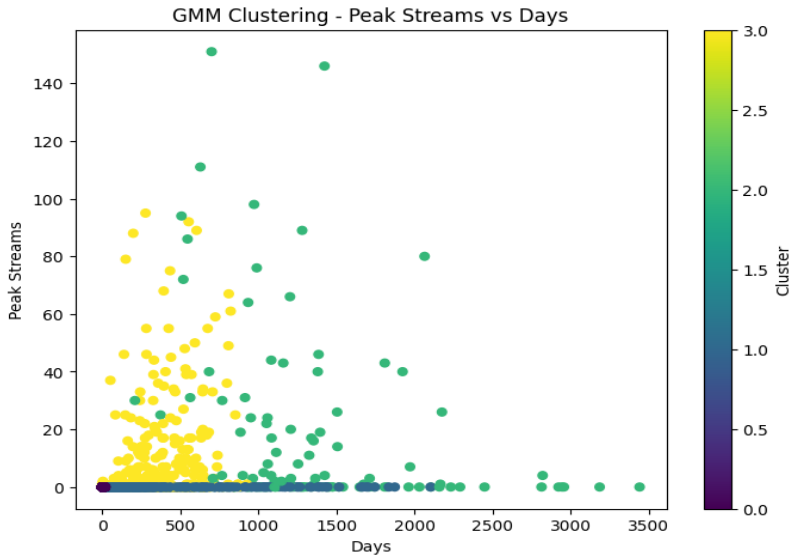
Gambar 5 S*catter Plot Total Stream vs Days*

Gambar 5 menunjukkan *scatter plot* antara Total Streams dan Days berdasarkan klaster GMM. Terlihat bahwa sebagian besar lagu memiliki Total Streams rendah dan durasi bertahan yang pendek, sedangkan hanya sedikit lagu yang mencapai jutaan streams dan bertahan lama di chart. Klaster dengan warna cerah menggambarkan lagu yang lebih populer dan stabil, sementara klaster berwarna gelap menunjukkan lagu dengan popularitas rendah.

Gambar 6 menampilkan hubungan antara Peak Streams dan Days dalam klaster GMM. Mayoritas lagu memiliki nilai *Peak Streams* rendah, dengan hanya sedikit yang mencapai puncak *streams* tinggi. Klaster dengan warna terang umumnya menunjukkan lagu yang memiliki puncak performa lebih baik dan bertahan lebih lama di chart, sedangkan klaster gelap berisi lagu yang kurang populer. Visualisasi ini memperlihatkan bahwa klaster GMM memiliki tumpang tindih antar kelompok.

Scatter plot hasil *clustering* menggunakan algoritma *Gaussian Mixture Model* (GMM) yang menampilkan hubungan antara Peak Streams dan Days menggambarkan hubungan antara jumlah puncak pemutaran lagu dan lamanya lagu tersebut berada di tangga lagu Spotify. Berdasarkan visualisasi, terlihat bahwa lagu-lagu dengan Peak Streams tinggi dan Days panjang cenderung berada dalam satu *cluster* yang mewakili lagu-lagu dengan tingkat popularitas tinggi, sementara lagu-lagu dengan Peak Streams rendah dan Days pendek dikelompokkan dalam cluster dengan popularitas rendah. Batas-batas antar cluster tampak tumpang tindih, yang menunjukkan bahwa metode GMM menghasilkan clustering yang lebih halus daripada K-Means. Hal ini menunjukkan bahwa GMM mampu menangkap keberadaan lagu-lagu dengan karakteristik campuran, seperti lagu-lagu yang sempat populer tetapi tidak bertahan lama, sehingga menghasilkan pemetaan variasi tingkat popularitas lagu di Spotify yang lebih realistis.

P-ISSN:1693-3877 / E-ISSN: 2407-1544
DOI:10.30873



Gambar 6 *Scatter Plot Peak Streams vs Days*

E. Perbandingan Hasil K-Means dan GMM

Perbandingan antara algoritma K-Means dan *Gaussian Mixture Model* (GMM) dilakukan untuk menilai efektivitas masing-masing metode dalam mengelompokkan data lagu Spotify. Kedua algoritma menggunakan jumlah optimal empat klaster, tetapi memiliki pendekatan yang berbeda dalam pembentukan klaster.

Algoritma K-Means menggunakan pendekatan berbasis jarak Euclidean, di mana setiap item data dikelompokkan ke dalam klaster dengan jarak terdekat ke pusat klaster (*centroid*). Pendekatan ini menghasilkan batas klaster yang jelas dan kelompok yang relatif terpisah. Metode ini sederhana, cepat, dan efisien untuk kumpulan data besar, tetapi kurang fleksibel dalam menangani data yang tumpang tindih.

Sebaliknya, algoritma *Gaussian Mixture Model* (GMM) menggunakan pendekatan probabilistik yang mengasumsikan bahwa data berasal dari kombinasi beberapa distribusi Gaussian. Setiap item data memiliki probabilitas tertentu untuk masuk ke dalam setiap klaster. Oleh karena itu, GMM mampu membentuk batas klaster yang lebih halus dan lebih cocok untuk data yang terdistribusi tidak seragam.

Berdasarkan hasil evaluasi menggunakan *Silhouette Score,* diperoleh nilai 0,731 untuk K-Means dan 0,256 untuk GMM. Nilai-nilai ini menunjukkan bahwa K-Means menghasilkan klaster yang lebih terpisah dan homogen dibandingkan GMM. Sementara itu, nilai 0,256 untuk GMM menunjukkan bahwa batas antar klaster masih tumpang tindih, menunjukkan karakteristik yang serupa antar kelompok lagu.

Tabel 3. Hasil Klaster

Klaster	K-Means (Data)	GMM (Data)
0	4652	2580
1	409	2086
2	25	122
3	122	420

Berdasarkan Tabel 3, terlihat bahwa distribusi anggota klaster yang dihasilkan K-Means dan GMM berbeda cukup jelas. Pada K-Means, klaster 0 mendominasi jumlah anggota (4652 data), sedangkan klaster lain jauh lebih kecil (misalnya klaster 2 hanya 25 data). Pada GMM, sebaran data juga tidak

P-ISSN:1693-3877 / E-ISSN: 2407-1544
DOI:10.30873

merata, namun proporsinya berbeda: klaster 0 berisi 2580 data dan klaster 1 berisi 2086 data, sementara klaster 2 dan 3 relatif lebih kecil. Perbedaan distribusi ini menunjukkan bahwa mekanisme "hard assignment" pada K-Means cenderung menarik sebagian besar data ke klaster dominan, sedangkan "soft assignment" pada GMM membuat pembagian data lebih fleksibel, walaupun tetap membentuk klaster mayor dan minor.

Dengan demikian, algoritma K-Means lebih unggul dalam mengelompokkan data Spotify secara eksplisit dan efisien, sementara GMM menghasilkan hasil yang lebih halus tetapi kurang jelas pemisahannya. Meskipun demikian, GMM tetap berguna ketika pemodelan probabilistik diperlukan atau ketika data menunjukkan variasi non-linier yang kompleks.

IV. SIMPULAN

Penelitian ini berhasil melakukan perbandingan kinerja antara algoritma K-Means dan Gaussian Mixture Model (GMM) dalam analisis popularitas lagu spotify Indonesia. Hasil evaluasi internal menunjukan bahwa algoritma K-Means memiliki kinerja yang lebih baik dalam menghasilkan klaster yang jelas dan terpisah, dengan nilai Silhouette Score sebesar 0,731 dibandingkan dengan GMM yang memperoleh nilai 0,256. Melalui proses klasterisasi tersebut, diperoleh empat kelompok utama popularitas lagu, yaitu Lagu Bintang/Viral, Lagu Hits Kuat, Lagu Evergreen, dan Lagu Rendah Popularitas. Temuan ini menunjukkan bahwa algoritma K-Means mampu memberikan segmentasi yang lebih tegas dan mudah diinterpretasikan, sehingga dapat dimanfaatkan untuk mendukung pengambilan keputusan strategis dalam industri musik, khususnya terkait waktu perilisan, perancangan strategi promosi, serta analisis tren pendengar di platform digital.

Berdasarkan hasil penelitian, disarankan agar pelaku industri musik termasuk label rekaman, musisi independen, maupun analisis data mempertimbangkan penerapan algoritma *K-Means* dalam menganalisis tren popularitas lagu pada *platform streaming*. Penelitian selanjutnya disarankan untuk mengeksplorasi variabel musikal tambahan seperti *danceability, energy*, dan *valence*, serta mengintegrasikannya dengan data demografis dan perilaku pengguna untuk memperoleh pemahaman yang lebih komprehensif mengenai karakteristik audiens. Selain itu, perbandingan dengan metode klasterisasi lain yang mampu menangani data non-linear juga dapat menjadi arah penelitian lanjutan guna meningkatkan akurasi dan interpretabilitas hasil klasterisasi.

DAFTAR PUSTAKA

[1] Agrawal, T., & Vankadara, S. K., 2023, Predicting Music Popularity: A Machine Learning Approach Using Spotify Data, International Journal of Intelligent Systems and Applications in Engineering, Vol. 11, No. 6, 724–733.

[2] Auliasari, K., & Kertaningtyas, M., 2022, Analisis cluster atribut audio pada lagu terpopuler aplikasi TikTok, Jurnal Sains dan Informatika, Vol. 8, No. 2, 140–149.

[3] Damayanti, S. E., Fajriana, M. I., Meilani, D., & Fatimah, S. S., 2024, Menjelajahi Wawasan Industri Musik: Klasterisasi Lagu Terpopuler di Spotify 2024 Dengan Metode K-Means Clustering, Prosiding Seminar Nasional CORISINDO 2024, Universitas Teknologi Bandung. https://corisindo.utb.ac.id/proceedings/2024/188

[4] Fitradhi, N. R., Hidayat, M. F., Saputro, T. W., Alifian, M. G., & Sari, A. P., 2023, Rekomendasi musik Spotify menggunakan metode K-Means, Prosiding Seminar Nasional Informatika Bela Negara (SANTIKA), Vol. 3, No. 3, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

[5] Jain, A. K., & Mishra, S., 2020, A Comprehensive Study on Gaussian Mixture Models for Unsupervised Learning, IEEE Access, Vol. 8, 135890–135905.

[6] Maulana, A., & Puspitaniangrum, D., 2023, Perbandingan Algoritma K-Means dan Gaussian Mixture Model (GMM) untuk Klasterisasi Pola Cuaca Ekstrem, Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), Vol. 7, No. 4, 812–819.

P-ISSN:1693-3877 / E-ISSN: 2407-1544
DOI:10.30873

[7] Netti, S. Y. M., & Irwanyah, I., 2018, Spotify: Aplikasi musik streaming untuk generasi milenial, Jurnal Komunikasi, Vol. 10, No. 1, 1–16.
[8] Patel, R., & Singh, V. K., 2021, K-Means vs. GMM: A Comparative Analysis for Clustering Complex Datasets, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 12, No. 5, 112–119.
[9] Spotify., 2024, Apa itu Spotify?, Spotify Help Center. https://support.spotify.com/id-id/article/what-is-spotify/
[10] Sujono, A., Rahmawati, D., & Putra, A., 2022, Analisis tren musik populer di Indonesia melalui data streaming Spotify menggunakan pendekatan data mining, Jurnal Teknologi Informasi dan Ilmu Komputer, Vol. 9, No. 4, 789–798.
[11] Susanto, A., & Wijaya, D., 2019, Analisis Performa Algoritma K-Means untuk Klasterisasi Data Skala Besar, Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), Vol. 6, No. 3, 289–296.
[12] Syafitri, W., Andreswari, R., & Mulia, H., 2024, Clustering Pop Songs Based On Spotify Data Using K-Means And K-Medoids Algorithm, Jurnal Media Informatika Budidarma, Vol. 8, No. 1, 213–221.
[13] Waghmare, S., & Bhalchandra, A., n.d., A Comparative Study of K-Means and GMM Clustering Algorithms on IRIS Dataset, International Journal of Computer Applications, Vol. 174, No. 46, 46–51.