

Analisis Klaster Lagu Indonesia pada Platform Spotify Menggunakan Algoritma K-Means (Studi Kasus: Spotify)

Cluster Analysis of Indonesian Songs on the Spotify Platform Using the K-Means Algorithm (Case Study: Spotify)

Rizki Nur Azim¹, Sudin Saepudin²

^{1,2}Program Studi Sistem Informasi, Universitas Nusa Putra

E-mail: rizki.azim_si22@nusaputra.ac.id, sudin.saepudin@nusaputra.ac.id

Abstrak

Perkembangan platform streaming seperti Spotify telah merevolusi konsumsi musik di Indonesia, menghasilkan data performa lagu yang kompleks dari Spotify Charts. Penelitian ini bertujuan mengelompokkan lagu Indonesia berdasarkan popularitas menggunakan algoritma K-Means Clustering. Variabel yang digunakan—*peak*, *previous*, *streak*, dan *streams*—dipilih karena secara langsung mencerminkan dinamika performa lagu di tangga lagu, meliputi posisi terbaik, pergerakan peringkat, ketahanan, dan volume pemutaran, sehingga lebih representatif terhadap resonansi pasar dibandingkan fitur audio. Pendekatan kuantitatif deskriptif diterapkan pada data Spotify Charts wilayah Indonesia periode Januari 2020 hingga Desember 2024, dengan sampel purposive sebanyak 5.316 data setelah pembersihan. Penentuan jumlah kluster optimal menggunakan Elbow Method menghasilkan $k=3$, dengan evaluasi kualitas kluster melalui Silhouette Score sebesar 0,4968. Hasil menunjukkan tiga kluster: Cluster 0 (populer menengah, rata-rata streams 3,8 juta), Cluster 1 (kurang populer, rata-rata streams 569 ribu), dan Cluster 2 (sangat populer, rata-rata streams 54 juta). Penelitian menyimpulkan bahwa K-Means efektif mengidentifikasi pola popularitas lagu secara objektif sebagai dasar strategi promosi berbasis data.

Kata kunci: Algoritma Klustering, Data Mining, K-Means, Musik Indonesia, Spotify Charts

Abstract

The growth of streaming platforms like Spotify has revolutionized music consumption in Indonesia, generating complex song performance data from Spotify Charts. This study aims to cluster Indonesian songs by popularity using the K-Means Clustering algorithm. The selected variables—peak, previous, streak, and streams—were chosen because they directly reflect song performance dynamics on the charts, encompassing best position, rank movement, longevity, and play volume, making them more representative of market resonance than audio features alone. A quantitative descriptive approach was applied to Indonesia Spotify Charts data spanning January 2020 to December 2024, with a purposive sample of 5,316 records post-cleaning. The optimal number of clusters was determined via the Elbow Method at $k=3$, with cluster quality evaluated using a Silhouette Score of 0.4968. Findings reveal three clusters: Cluster 0 (medium popularity, 3.8M average streams), Cluster 1 (low popularity, 569K average streams), and Cluster 2 (high popularity, 54M average streams). The study concludes that K-Means effectively identifies objective popularity patterns to support data-driven promotion strategies.

Keywords: Clustering Algorithm, Data Mining, Indonesian Music, K-Means, Spotify Chart

1. PENDAHULUAN

Perkembangan teknologi informasi yang masif telah merevolusi ekosistem industri musik global, mengubah paradigma konsumsi dari kepemilikan fisik menjadi aksesibilitas berbasis *streaming*. Platform musik digital seperti Spotify telah menjadi katalis utama dalam transformasi ini, memungkinkan distribusi konten yang cepat, fleksibel, dan melampaui batasan geografis bagi jutaan pengguna di seluruh dunia. Di Indonesia, adopsi layanan *streaming* menunjukkan tren pertumbuhan yang signifikan, seiring dengan peningkatan penetrasi internet dan pergeseran perilaku audiens yang kini lebih mengutamakan kenyamanan akses *real-time*. Dinamika ini tidak hanya menciptakan ruang baru bagi musisi lokal untuk mempromosikan karya, tetapi juga menempatkan platform *streaming* sebagai pusat gravitasi ekonomi kreatif digital nasional [1].

Dalam konteks domestik, Spotify telah menjadi salah satu media utama bagi audiens Indonesia untuk mengeksplorasi preferensi musikal mereka, dengan pangsa pasar yang terus bersaing ketat dengan layanan lain seperti YouTube Music [2]. Fenomena ini tercermin dari meningkatnya jumlah lagu lokal yang masuk dalam tangga lagu digital, yang mengindikasikan bahwa industri musik nasional kini semakin kompetitif dan responsif terhadap algoritma platform. Namun, di balik kemudahan akses tersebut, terdapat kompleksitas data perilaku pengguna yang dihasilkan setiap harinya—termasuk jumlah pemutaran, durasi *streak*, serta posisi lagu dalam tangga lagu—yang sering kali belum dimanfaatkan secara optimal oleh para pelaku industri untuk memahami pola popularitas secara komprehensif [1].

Penelitian mengenai perilaku musik digital sebenarnya telah banyak dilakukan, terutama yang berfokus pada ekstraksi fitur audio seperti tempo, *energy*, dan *danceability* untuk keperluan sistem rekomendasi atau klasifikasi genre. Beberapa studi terdahulu, seperti yang dilakukan oleh peneliti di bidang sistem informasi, cenderung menggunakan teknik *data mining* untuk memetakan karakteristik musikal lagu guna meningkatkan presisi *playlist* personal. Meskipun pendekatan ini memberikan pemahaman mendalam tentang komposisi teknis musik, studi-studi tersebut sering kali mengabaikan metrik performa tangga lagu yang secara langsung mencerminkan resonansi pasar terhadap suatu karya di dunia nyata [1].

Terdapat dikotomi yang mencolok dalam literatur antara penelitian yang berorientasi pada konten audio dan penelitian yang berorientasi pada performa pasar. Sementara studi berbasis fitur audio (seperti tempo dan *loudness*) telah mencapai tingkat maturitas yang baik, penggunaan data *Spotify Charts*—seperti *peak*, *previous*, *streak*, dan *streams*—sebagai variabel utama dalam analisis kluster masih sangat terbatas, khususnya pada cakupan lagu-lagu Indonesia [3]. Keterbatasan ini menjadi celah penelitian (*research gap*) yang krusial, mengingat karakteristik populer tidak selalu berbanding lurus dengan fitur musikal yang dominan. Akibatnya, pelaku industri sering kali kesulitan dalam membedah pola popularitas yang lebih bersifat dinamis dan kontekstual, seperti lagu yang sedang viral dibandingkan dengan lagu yang memiliki basis pendengar setia dalam jangka panjang [3].

Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan menerapkan algoritma *K-Means Clustering* untuk mengelompokkan lagu-lagu Indonesia di Spotify berdasarkan variabel performa popularitas: *peak*, *previous*,

streak, dan *streams* [3]. Melalui pendekatan analitik ini, penelitian ini berupaya menjawab bagaimana karakteristik performa popularitas dapat dipetakan secara objektif ke dalam klaster-klaster yang bermakna, seperti kategori lagu yang sedang naik daun, lagu populer, atau lagu yang mengalami penurunan performa. Kebaruan penelitian ini terletak pada pemanfaatan metrik performa tangga lagu sebagai variabel determinan dalam *clustering*, yang memberikan kontribusi teoritis bagi pengembangan analisis data musik digital serta kontribusi praktis bagi musisi dan label rekaman dalam merumuskan strategi promosi dan distribusi yang berbasis data (*data-driven decision making*) [3].

1. METODOLOGI PENELITIAN

Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan teknik data mining K-Means Clustering untuk menganalisis pola performa popularitas lagu Indonesia dari data Spotify Charts, termasuk variabel *peak*, *previous*, *streak*, dan *streams*. Pendekatan ini memungkinkan penggambaran objektif dan pengelompokan data numerik secara sistematis, sesuai dengan definisi penelitian kuantitatif yang menekankan pengukuran dan analisis statistik. Sugiyono menjelaskan bahwa metode kuantitatif deskriptif efektif untuk mengidentifikasi pola data sekunder seperti charts musik, sementara K-Means mendukung penemuan struktur tersembunyi [4], [5].

Instrumen dan Teknik Analisis Data

Instrumen utama mencakup Spotify Charts sebagai sumber data primer, dataset terstruktur dalam Microsoft Excel, serta perangkat lunak Python dengan library Pandas, NumPy, Scikit-learn, JupyterLab, dan Matplotlib untuk pengolahan dan visualisasi. Teknik analisis data meliputi pra-pemrosesan (pembersihan duplikat, *missing values*, normalisasi Min-Max), penentuan klaster optimal via Elbow Method, klasterisasi K-Means berbasis Euclidean distance, evaluasi Silhouette Coefficient, serta visualisasi melalui scatter plot, boxplot, dan heatmap. Sudaryono dan Emzir menekankan bahwa instrumen non-responden seperti software analisis data mining memastikan validitas dalam penelitian kuantitatif berbasis data sekunder, dengan normalisasi untuk kesetaraan skala variabel. Creswell mendukung penggunaan mixed tools untuk analisis klaster guna meningkatkan reliabilitas hasil [6], [7].

Populasi dan Sampel

Populasi penelitian terdiri dari seluruh lagu Indonesia yang masuk Spotify Charts wilayah Indonesia (regional, harian/mingguan) dalam periode tertentu, didefinisikan sebagai lagu oleh musisi asal Indonesia yang dipublikasikan resmi di Spotify. Sampel diambil secara purposive dari lagu-lagu terpopuler berdasarkan data performa (*peak*, *previous*, *streak*, *streams*), dengan seleksi ketat untuk menghindari duplikat dan memastikan representativitas pola popularitas. Teknik pengumpulan melibatkan akses situs Spotify Charts, ekstraksi data ke spreadsheet, dan verifikasi konsistensi, sebagaimana direkomendasikan Sugiyono untuk sampling non-probability dalam studi deskriptif data mining [4], [8], [9].

Prosedur Penelitian

Prosedur dimulai dengan studi literatur untuk landasan teori K-Means dan clustering musik, diikuti pengumpulan data dari Spotify Charts melalui tahapan akses situs, pemilihan kategori Indonesia, ekstraksi variabel, pembersihan, dan penyusunan dataset. Selanjutnya, pra-pemrosesan data (normalisasi $X' = \frac{X - X_{min}}{X_{max} - X_{min}}$ penentuan k optimal (Elbow Method pada WCSS), klusterisasi erisasi iteratif K-Means $J = \sum_{i=1}^k \sum_{x \in C_i} \|x - c_i\|^2$, evaluasi (Silhouette Coefficient), visualisasi (scatter plot, boxplot, heatmap), analisis karakteristik kluster, dan interpretasi untuk mengidentifikasi kelompok popularitas. Alur ini mengikuti prinsip sistematis Creswell untuk desain kuantitatif, memastikan konvergensi dan interpretasi logis, sementara Emzir menyoroti pentingnya prosedur iteratif untuk validasi kluster.

3. HASIL DAN PEMBAHASAN

Gambaran Umum Data Penelitian

Tabel 1 Data Spotify Chart

No	track	artist	track name	peak	prev	streak	streams
1	For Revenge	Serana		1428	1	1,180,531	721,336,096
2	Feby Putri	Runtuh		1645	1	772,988	698,855,057
3	Anggi Marito	Tak Segampang Itu		1193	1	1,587,709	660,834,582
4	Hindia	Rumah Ke Rumah		2325	6	850,039	655,570,413
5	Nadin Amizah	Bertaut		2144	1	708,303	655,235,467
6	Raim Laode	Komang		1155	1	1,780,047	624,330,037
7	Last Child	Duka		1736	9	752,610	617,076,288
8	Tulus	Monokrom		3406	11	732,440	598,954,003
9	Virgoun	Selamat (Selamat Tinggal)		2453	6	661,240	597,592,808
10	Lyodra	Pesan Terakhir		1727	3	917,057	583,341,409
11	Juicy Luicy	Lantas		1895	2	819,431	581,171,297
12	Pamungkas	Monolog		2192	1	1,639,729	581,125,413
13	Nadhif Basalamah	penjaga hati		921	1	1,469,908	526,525,364
14	Vierra	Seandainya		1870	12	666,772	503,206,803
15	Hindia	Evaluasi		1601	1	574,234	501,626,078
16	Andra & The Backbone	Sempurna		3177	13	548,619	484,381,005
17	Tulus	Hati-Hati di Jalan		1501	1	1,461,480	479,830,213
18	.Feast	Nina		594	1	1,159,363	474,446,402
19	Batas Senja	Nanti Kita Seperti Ini		1073	3	821,350	467,315,291
20	Suara Kayu	Kembali Pulang		1157	8	669,116	447,824,296
.....
5320	Ghaisa Kenang	Maaf		1	189	1,001	1,001

Pemrosesan Data

1. Data Cleaning

Pada tahap data cleaning dilakukan proses pembersihan data untuk memastikan kualitas dataset yang digunakan. Proses ini meliputi penghapusan data yang memiliki nilai kosong (missing value), data tidak valid, serta data duplikat [10].

Berdasarkan hasil preprocessing, jumlah data berkurang dari 5320 menjadi 5316 data, sehingga terdapat 4 data yang dihapus. Data yang dihapus tersebut

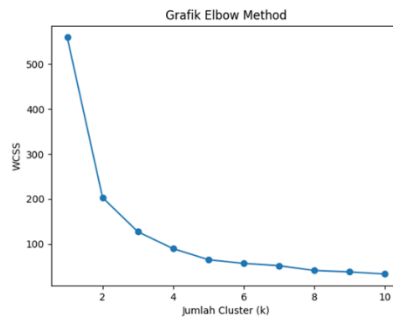
tidak memenuhi kriteria kelayakan karena mengandung nilai yang tidak lengkap atau tidak dapat dikonversi ke dalam bentuk numerik. Dengan demikian, dataset yang digunakan pada tahap analisis merupakan data yang telah bersih dan siap diolah.

2. Normalisasi Data

Tahap selanjutnya adalah normalisasi data yang bertujuan untuk menyamakan skala antar variabel. Variabel yang digunakan dalam penelitian memiliki rentang nilai yang berbeda, terutama pada variabel *streams* yang memiliki nilai jauh lebih besar dibandingkan variabel lainnya.

Normalisasi dilakukan menggunakan metode Min-Max Normalization yang mengubah nilai data ke dalam rentang 0 hingga 1. Proses ini bertujuan agar setiap variabel memiliki kontribusi yang seimbang dalam proses clustering, mengingat algoritma K-Means menggunakan perhitungan jarak dalam menentukan kemiripan data [11].

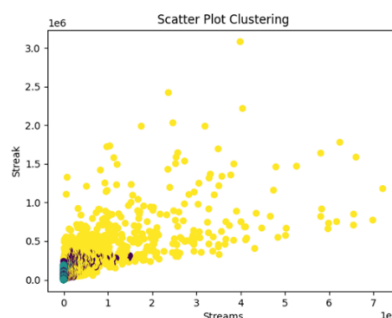
Penentuan Jumlah Kluster (Elbow Method)



Gambar 1 Elbow Method

Grafik Elbow menggambarkan penurunan signifikan WCSS terhadap jumlah cluster (k) hingga titik siku pada $k=3$, di mana penurunan mulai melandai, sehingga jumlah cluster optimal adalah 3. Nilai WCSS yang semakin kecil menandakan variasi data dalam cluster semakin baik.

Hasil Clustering K-Means



Gambar 2 Scatter Plot Clustering

Scatter plot digunakan untuk memvisualisasikan hasil clustering berdasarkan atribut *streams* dan *streak*. Setiap titik pada grafik merepresentasikan satu data lagu, sedangkan warna menunjukkan cluster tempat data tersebut berada.

Dari grafik tersebut terlihat bahwa data terbagi menjadi beberapa kelompok yang memiliki pola distribusi tertentu. Hal ini menunjukkan bahwa algoritma K-Means berhasil mengelompokkan data berdasarkan kemiripan karakteristik

Evaluasi Clustering

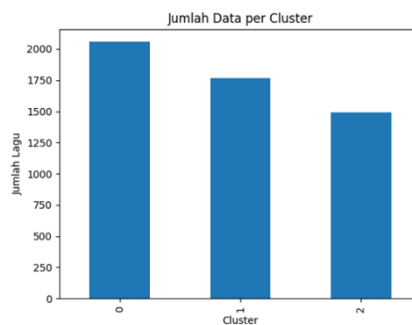
Evaluasi hasil clustering dilakukan menggunakan metode Silhouette Score. Metode ini digunakan untuk mengukur seberapa baik suatu data ditempatkan dalam cluster yang sesuai.

Nilai Silhouette Score yang diperoleh adalah sebesar: **0,4968**

Nilai Silhouette Score berada pada rentang -1 hingga 1. Nilai mendekati 1 menunjukkan bahwa data dalam cluster sangat kompak dan terpisah dengan baik dari cluster lain. Nilai mendekati 0 menunjukkan bahwa cluster saling tumpang tindih, sedangkan nilai negatif menunjukkan adanya kesalahan pengelompokan [12].

Berdasarkan nilai yang diperoleh, dapat disimpulkan bahwa kualitas clustering berada pada kategori cukup baik.

Jumlah Data Pada Setiap Cluster

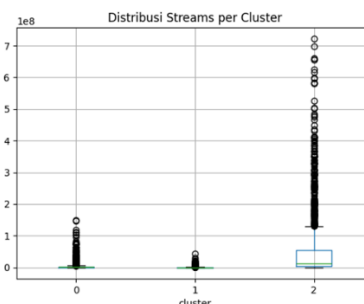


Gambar 3 Jumlah Data per Cluster

Grafik ini menunjukkan distribusi jumlah data pada masing-masing cluster, data yang di hasilkan dari grafik di atas, **Cluster 0 = 2057**, **Cluster 1 = 1768**, dan **Cluster 2 = 1491** Analisis terhadap grafik ini penting untuk mengetahui keseimbangan distribusi data.

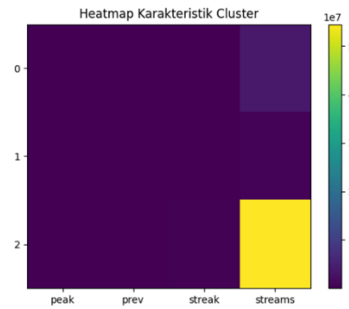
Jika jumlah data pada setiap cluster relatif seimbang, maka model clustering dapat dikatakan stabil. Namun jika terdapat perbedaan yang signifikan, maka hal tersebut menunjukkan adanya dominasi karakteristik tertentu dalam dataset.

Analisis Karakteristik Cluster



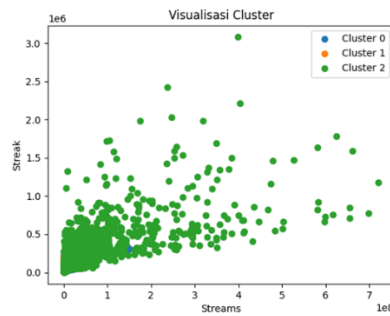
Gambar 4 Distribusi Streams per Cluster

Boxplot digunakan untuk melihat distribusi nilai *streams* pada setiap cluster. Melalui boxplot, dapat diketahui nilai median, rentang interkuartil, serta keberadaan outlier. Dari grafik tersebut terlihat adanya perbedaan distribusi antara masing-masing cluster. Hal ini menunjukkan bahwa setiap cluster memiliki karakteristik popularitas yang berbeda



Gambar 5 Heatmap Karakteristik Cluster

Heatmap digunakan untuk menampilkan nilai rata-rata dari setiap atribut pada masing-masing cluster dalam bentuk visual. Melalui heatmap, dapat terlihat dengan jelas perbedaan pola antar cluster, sehingga mempermudah proses interpretasi.



Gambar 6 Visualisasi Cluster

Visualisasi ini memberikan gambaran yang lebih jelas mengenai pemisahan antar cluster dengan menggunakan warna yang berbeda. Dari grafik tersebut terlihat bahwa masing-masing cluster memiliki pola penyebaran yang berbeda, yang menunjukkan bahwa proses clustering telah berhasil dilakukan dengan baik

Interpretasi Cluster

1. Cluster 0 (Lagu Populer Menengah)

Tabel 2 Cluster 0

track name	track artist	streams	streak
Menjaga Hati	Yovie & Nuno	149177942	312765
Kelingan Mantan	NDX A.K.A.	146645890	283209
Sampai Akhir Waktu	Yovie & Nuno	110645538	264740
Abadi - Indo Version	Dendi Nata	103685334	331669
Aku, Kamu Dan Samudra	Rebellion Rose	96632296	277378

Cluster 0 menunjukkan rata-rata streams 3,8 juta dan streak 60 ribu, mencerminkan popularitas menengah yang stabil dengan ketahanan lama di tangga lagu. Nilai peak 66 dan previous 84 mengindikasikan performa konsisten di peringkat menengah, bukan puncak viral, sehingga merepresentasikan lagu stabil tanpa lonjakan ekstrem.

2. Cluster 1 (Lagu Kurang Populer)

Tabel 3 Cluster 1

track name	track artist	streams	streak
Aku Ikhlas	Aftershine	30739882	199971
Sayap Pelindungmu	TheOvertunes	19557029	77205
Tresno liyane	Northsle	18322124	233365
Seamin Tak Seiman	Mahen	15117951	196833
Waking Up Together With You	Ardhito Pramono	12982253	205247

Cluster 1 memiliki rata-rata streams terendah (569 ribu) dan streak 40 ribu, menunjukkan ketahanan rendah di tangga lagu. Meski peak mencapai 10 (posisi tinggi sesaat), previous 160 mengindikasikan penurunan tajam dan ketidakstabilan, sehingga merepresentasikan lagu dengan popularitas fluktuatif atau rendah secara keseluruhan.

3. Cluster 2 (Lagu Sangat Populer / Viral Kuat)

Tabel 4 Cluster 2

track name	track artist	streams	streak
Serana	For Revenge	721336096	1180531
Runtuh	Feby Putri	698855057	772988
Tak Segampang Itu	Anggi Marito	660834582	1587709
Rumah Ke Rumah	Hindia	655570413	850039
Bertaut	Nadin Amizah	655235467	708303

Cluster 2 memiliki rata-rata streams tertinggi (54 juta) dan streak 251 ribu, mencerminkan popularitas sangat tinggi dengan ketahanan lama di tangga lagu. Nilai previous 19 menunjukkan posisi konsisten di puncak, meski peak bervariasi, sehingga merepresentasikan lagu viral, stabil, dan berdaya tarik kuat.

Hasil dan Pembahasan

Berdasarkan seluruh tahapan penelitian yang telah dilakukan, mulai dari pengumpulan data, preprocessing, normalisasi menggunakan metode Min-Max, penentuan jumlah klaster dengan Elbow Method, hingga proses klasterisasi menggunakan algoritma K-Means, diperoleh hasil pengelompokan data lagu Indonesia ke dalam tiga klaster utama.

Hasil visualisasi menunjukkan bahwa data telah terbagi ke dalam kelompok yang cukup jelas. Meskipun masih terdapat beberapa titik yang berdekatan antar klaster, secara umum pola pemisahan data sudah terlihat dengan baik. Hal ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan data berdasarkan kemiripan karakteristik yang dimiliki.

Berdasarkan analisis karakteristik kluster, diperoleh tiga kategori utama, yaitu kluster lagu terpopuler menengah, kluster lagu yang kurang populer, dan kluster lagu yang sangat populer / Viral kuat. Perbedaan antar kluster ditentukan berdasarkan variabel *peak*, *previous*, *streak*, dan *streams* yang digunakan dalam penelitian.

Hasil evaluasi menggunakan Silhouette Coefficient menghasilkan nilai sebesar 0,4968 yang menunjukkan bahwa kualitas kluster berada pada kategori cukup baik. Nilai ini mengindikasikan bahwa data dalam satu kluster memiliki tingkat kemiripan yang cukup tinggi serta terdapat pemisahan yang cukup jelas antar kluster.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa metode K-Means dapat digunakan untuk mengidentifikasi pola popularitas lagu Indonesia secara objektif. Hasil pengelompokan ini memberikan gambaran mengenai dinamika popularitas lagu di platform Spotify serta dapat dimanfaatkan sebagai dasar dalam analisis tren musik digital.

KESIMPULAN

Penelitian ini berhasil mengelompokkan lagu Indonesia di Spotify Charts ke dalam tiga kluster utama menggunakan algoritma K-Means, dengan temuan bahwa Cluster 0 merepresentasikan lagu populer menengah (rata-rata streams 3,8 juta, streak 60 ribu), Cluster 1 menunjukkan lagu kurang populer atau fluktuatif (streams 569 ribu, streak 40 ribu), dan Cluster 2 mencerminkan lagu sangat populer atau viral kuat (streams 54 juta, streak 251 ribu). Jumlah kluster optimal ditentukan melalui Elbow Method pada $k=3$, didukung Silhouette Score 0,4968 yang mengindikasikan kualitas clustering cukup baik, sehingga metode ini efektif mengungkap pola performa popularitas berdasarkan variabel *peak*, *previous*, *streak*, dan *streams* dari 5316 data bersih.

Meskipun demikian, keterbatasan penelitian meliputi ketergantungan pada data sekunder Spotify Charts yang rentan terhadap bias algoritma platform dan kurangnya variabel audio seperti tempo atau danceability, yang dapat memperkaya interpretasi. Selain itu, periode data terbatas berpotensi tidak menangkap tren musiman jangka panjang. Untuk penelitian mendatang, disarankan mengintegrasikan fitur audio Spotify API, membandingkan dengan platform lain seperti YouTube Music, serta menerapkan algoritma clustering lain seperti DBSCAN untuk validasi lebih lanjut. Secara praktis, hasil ini memberi implikasi bagi musisi dan label rekaman dalam merancang strategi promosi data-driven, seperti memprioritaskan lagu potensial viral di Cluster 2, sementara lagu stabil di Cluster 0 dapat difokuskan pada penguatan basis penggemar setia.

DAFTAR PUSTAKA

- [1] P. Aplikasi *et al.*, "Pemanfaatan aplikasi musik spotify sebagai media hiburan di kalangan generasi z di kota pekanbaru," no. 6955, 2024.
- [2] dkk. Penulis, "Segmentasi Pengguna Spotify Berdasarkan Preferensi Musik Menggunakan K-Means," *Journal of Natural and Applied Sciences (JNATIA)*, 2024, doi: 10.24843/jnatia.

- [3] M. I. Firmansyah, R. S. Rohman, and E. Marsusanti, "Penerapan Algoritma Klustering K-Means Untuk Fitur Atribut Pada Layanan Streaming Musik Spotify," vol. 2, no. 2, pp. 58–66, 2023.
- [4] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan RnD*, 27th ed. Bandung: Alfabeta, 2019.
- [5] S. Arikunto, *Metodologi Penelitian Kuantitatif*. 2010.
- [6] J. W. Cresswell, *Research design: Qualitative, quantitative, and mixed methods approaches*, 5th ed. SAGE Publications, 2021.
- [7] A. J. Onwuegbuzie and N. L. Leech, "Enhancing the Interpretation of Significant Findings: The Role of Mixed Methods Research," *The Qualitative Report*, 2015, doi: 10.46743/2160-3715/2004.1913.
- [8] Sudaryono, *Metodologi Penelitian*, 1st ed. Depok: Rajawali Pers, 2018.
- [9] Emzir, *Metodologi Penelitian Pendidikan*, 1st ed. Jakarta: Rajawali Pers, 2020.
- [10] I. Ghozali, *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25 Edisi 9*, 9th ed. Semarang: Badan Penerbit UNDIP, 2018.
- [11] S. Santoso, *Menguasai Statistik SPSS 25*. Jakarta: PT Elex Media Komputindo, 2018.
- [12] D. K. A. Astutik, A. Indrasetianingsih, and F. Fitriani, "Penerapan Text Mining pada Analisis Sentimen Pengguna Twitter Layanan Transportasi Online Menggunakan Metode Density Based Spatial Clustering of Applications with Noise (DBSCAN) dan K-Means," *Statistika*, vol. 15, no. 1, pp. 184–194, 2022.