



(Research/Review) Article

Klasterisasi Lagu Populer dan Eksplorasi Subgenre Spotify 2024 dengan K-Medoids

Alfia Nurlaili Tahiyat^{1*}, Bima Maulana², Ade Eka Saputra³, Lusiana Efrizoni⁴, Rah-maddeni⁵

¹ Universitas Sains dan Teknologi Indonesia, e-mail : alfiatahiyat@gmail.com

² Universitas Sains dan Teknologi Indonesia, e-mail : bimsbima99@gmail.com

³ Universitas Sains dan Teknologi Indonesia; e-mail : adecka998@gmail.com

⁴ Universitas Sains dan Teknologi Indonesia; e-mail : lusiana@stmik-amik-riau.ac.id

⁵ Universitas Sains dan Teknologi Indonesia; e-mail : rahmaddeni@sar.ac.id

* Corresponding Author : Alfia Nurlaili Tahiyat

Abstract: Spotify's genre classification system remains too broad, often grouping songs with distinct characteristics into the same category. For example, Pop Ballads and Dance Pop are frequently classified under "Pop" despite significant differences in tempo, emotion, and production style. This leads to inaccurate song recommendations. This study applies the K-Medoids algorithm to enhance song classification based on Spotify Playlist Count, Spotify Playlist Reach, and Spotify Popularity. The CRISP-DM methodology guides business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Clustering results without popularity ranking reveal three main groups: songs with low playlist count but high reach (dominated by light hip-hop), songs with high playlist count and reach (dominated by contemporary R&B), and songs with low popularity (dominated by dance). After ranking by popularity, clusters became more defined, with alternative pop dominating the high-reach cluster, contemporary R&B in the popular cluster, and dance pop in the less popular cluster. Evaluation using a Silhouette Score of 0.5014 indicates good cluster quality. Additionally, this study successfully identified the 15 most popular songs on Spotify in 2024. These findings can help Spotify refine its recommendation system by incorporating subgenre-based classification, ensuring more accurate search results aligned with user preferences and evolving music trends.

Keywords: K-Medoids, Spotify, genre clustering, subgenre.

Abstrak: Spotify memiliki sistem klasifikasi genre yang masih terlalu luas, menyebabkan lagu dengan karakteristik berbeda dikelompokkan dalam kategori yang sama. Misalnya, Pop Ballads dan Dance Pop sering digabungkan dalam kategori "Pop", meskipun keduanya memiliki perbedaan signifikan dalam tempo, emosi, dan gaya produksi. Hal ini berdampak pada rekomendasi lagu yang kurang akurat. Penelitian ini menerapkan algoritma K-Medoids untuk meningkatkan klasifikasi lagu berdasarkan Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity. Metodologi CRISP-DM digunakan dalam tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Hasil klasterisasi tanpa pengurutan popularitas menunjukkan tiga kelompok utama: lagu dengan jumlah playlist rendah tetapi jangkauan luas (dominan hip-hop ringan), lagu dengan playlist dan jangkauan tinggi (dominan contemporary R&B), serta lagu dengan popularitas rendah (dominan dance). Setelah dilakukan pengurutan berdasarkan popularitas, klaster semakin terdefinisi dengan genre dominan alternative pop pada klaster lagu dengan reach tinggi, contemporary R&B untuk lagu populer, serta dance pop untuk lagu kurang populer. Evaluasi menggunakan Silhouette Score sebesar 0.5014 mengindikasikan kualitas klaster yang baik. Selain itu, penelitian ini berhasil mengidentifikasi 15 lagu terpopuler di Spotify 2024. Hasil penelitian ini dapat digunakan oleh Spotify untuk meningkatkan sistem rekomendasi berbasis subgenre, sehingga hasil pencarian lagu lebih akurat sesuai preferensi pengguna dan tren musik digital yang berkembang.

Kata kunci: K-Medoids, Spotify, pengelompokan genre, subgenre.

Received: April 05, 2025

Revised: April 10, 2025

Accepted: April 28, 2025

Published: April 30, 2025

Curr. Ver.: April 30, 2025



Copyright: © 2025 by the authors.
Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Pendahuluan

Spotify adalah salah satu layanan streaming musik terbesar di dunia dengan lebih dari 600 juta pengguna aktif bulanan pada tahun 2024, di mana 236 juta di antaranya merupakan pelanggan berbayar. Platform ini menawarkan akses ke lebih dari 100 juta lagu dengan berbagai genre. Namun, sistem klasifikasi genre yang digunakan masih terlalu luas, menyebabkan lagu dengan karakteristik berbeda sering kali dikelompokkan dalam kategori yang sama. Misalnya, Pop Ballads dan Dance Pop sering digabungkan dalam kategori “Pop,” meskipun keduanya memiliki perbedaan signifikan dalam tempo, emosi, dan gaya produksi. Akibatnya, pengguna mungkin menerima rekomendasi lagu yang kurang akurat, yang berpotensi menurunkan pengalaman mendengarkan mereka.

Meningkatkan akurasi rekomendasi musik menjadi penting, terutama karena preferensi pengguna semakin beragam dan industri musik digital berkembang pesat. Salah satu pendekatan yang dapat digunakan adalah dengan mengelompokkan lagu berdasarkan metrik objektif seperti Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity. Metrik ini mencerminkan penerimaan dan eksposur lagu di dalam ekosistem Spotify, yang dapat menjadi dasar dalam membangun sistem rekomendasi yang lebih presisi. Dengan menerapkan metode klusterisasi yang lebih akurat, sistem rekomendasi dapat memberikan saran yang lebih relevan kepada pengguna, serta memberikan wawasan yang lebih dalam tentang tren konsumsi musik digital bagi industri musik.

Berbagai penelitian telah mengeksplorasi metode pengelompokan dalam analisis data musik. Suastrri (2020) menunjukkan bahwa K-Means Clustering dapat meningkatkan akurasi rekomendasi lagu berbasis data Spotify. Namun, penelitian oleh Fitradhi dkk. (2023) menemukan bahwa metode ini memiliki kelemahan dalam menangani outlier, yang berdampak pada stabilitas kluster. Sementara itu, Kim dkk. (2021) menggunakan pendekatan berbasis representasi numerik musik untuk menyempurnakan pemetaan genre, tetapi metode tersebut masih menghadapi kesulitan dalam membedakan subgenre yang mirip. Anoohya dkk. (2023) menemukan bahwa K-Medoids lebih stabil dan akurat dibandingkan K-Means dalam sistem rekomendasi musik karena kemampuannya dalam menangani data yang tidak merata. Zaidah dkk. (2021) mengevaluasi K-Means, K-Medoids, dan Agglomerative Clustering untuk klasifikasi genre, dan menyimpulkan bahwa K-Medoids lebih efektif dalam menangani pola popularitas yang kompleks.

Dalam penelitian terkait klusterisasi, Ramadhani dkk. (2024) membandingkan K-Means dan K-Medoids untuk analisis data kemiskinan dan menemukan bahwa meskipun K-Means memiliki nilai Silhouette Coefficient yang lebih tinggi, K-Medoids lebih unggul dalam menangani data dengan distribusi tidak merata. Hal ini menunjukkan bahwa algoritma K-Medoids lebih stabil dalam kondisi data yang kompleks. Sementara itu, penelitian oleh Ardiansyah dkk. (2022) menekankan bahwa metode klusterisasi seperti K-Means dan Decision Tree dapat meningkatkan efektivitas prediksi berbasis data numerik, yang relevan dalam konteks pengelompokan lagu berdasarkan popularitas dan jangkauan.

Metode K-Medoids dipilih dalam penelitian ini karena kemampuannya dalam membentuk kluster yang lebih stabil serta menangani distribusi data musik yang tidak merata. Berbeda dengan metode lain yang bergantung pada rata-rata pusat kluster seperti K-Means, K-Medoids menggunakan titik data yang mewakili medoid, sehingga lebih robust terhadap outlier. Penelitian sebelumnya menunjukkan bahwa modifikasi K-Medoids dapat meningkatkan skalabilitas dan efisiensi untuk kumpulan data besar, membuatnya cocok untuk beragam data musik. Tidak seperti Agglomerative Clustering, K-Medoids menawarkan struktur kluster yang fleksibel, yang ideal untuk analisis genre musik yang lebih spesifik.

2. Kajian Pustaka atau Penelitian Terkait

Teori Klasterisasi dalam Sistem Rekomendasi Musik

Klasterisasi merupakan teknik dalam pembelajaran tak terawasi yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristiknya. Dalam konteks sistem rekomendasi musik, klasterisasi memungkinkan pengelompokan lagu berdasarkan berbagai parameter, seperti popularitas, jangkauan audiens, serta frekuensi dimasukkannya lagu ke dalam playlist. Metode ini digunakan untuk membangun segmentasi pengguna maupun konten, yang dapat meningkatkan akurasi sistem rekomendasi[8].

Beberapa algoritma populer untuk klasterisasi termasuk K-Means, K-Medoids, Agglomerative Clustering, dan DBSCAN. Di antara metode tersebut, K-Medoids dikenal memiliki keunggulan dalam mengatasi outlier dan distribusi data yang tidak merata, karena menggunakan medoid (titik aktual dalam data) sebagai pusat klaster[5].

K-Medoids sebagai Algoritma Klasterisasi Stabil

K-Medoids menjadi alternatif utama dari K-Means dalam berbagai penelitian karena ketahanannya terhadap noise dan outlier (Anoohya et al., 2023) [4]. Berbeda dengan K-Means yang menghitung rata-rata sebagai pusat klaster, K-Medoids memilih data aktual yang paling representatif dari klaster tersebut. Hal ini menjadikannya lebih stabil dalam kondisi data yang kompleks, termasuk data musik digital yang sangat beragam dari segi popularitas dan eksposur.

Zaidah dkk. (2021) membandingkan tiga algoritma (K-Means, K-Medoids, dan Agglomerative Clustering) dalam klasifikasi genre Spotify dan menemukan bahwa K-Medoids menghasilkan klaster yang lebih akurat, terutama dalam menangani distribusi popularitas yang kompleks. Sementara itu, penelitian oleh Gandhi dan Srivastava (2021) juga menunjukkan bahwa modifikasi K-Medoids dapat meningkatkan efisiensi pada dataset besar [10].

Klasterisasi Data Musik Spotify

Penelitian terhadap data musik Spotify telah berkembang pesat dalam beberapa tahun terakhir. Suastru (2023) menggunakan K-Means untuk mengelompokkan lagu pop berdasarkan data Spotify dan menunjukkan peningkatan akurasi rekomendasi [1]. Namun, Fitradhi et al. (2023) menemukan kelemahan K-Means dalam menangani outlier dan stabilitas klaster [2].

Kim et al. (2021) menggunakan representasi numerik musik untuk menyempurnakan pemetaan genre, namun pendekatan ini kurang efektif dalam membedakan subgenre yang memiliki kesamaan karakteristik emosional dan produksi [3]. Mufidah dkk. (2017) mengusulkan Fuzzy C-Means untuk pengelompokan lagu berdasarkan emosi, tetapi hasilnya menunjukkan batas antar-klaster yang masih kabur [11].

Sidora dan Harani (2023) menerapkan KNN dan Algoritma Genetika untuk sistem rekomendasi musik Spotify. Namun, penelitian tersebut lebih berfokus pada kecocokan preferensi pengguna, bukan eksplorasi struktur data lagu secara menyeluruh seperti dalam klasterisasi berdasarkan metrik popularitas [13].

Kesenjangan Penelitian dan Posisi Kontribusi

Mayoritas penelitian sebelumnya cenderung fokus pada pengelompokan lagu berdasarkan fitur musikal (acousticness, energy, danceability) atau metadata pengguna, namun masih minim eksplorasi terhadap metrik distribusi lagu seperti playlist count, playlist reach, dan popularitas secara terintegrasi. Selain itu, belum banyak studi yang menyoroti bagaimana hasil

klasterisasi berdasarkan metrik popularitas dapat digunakan untuk mengekstraksi pola sub-genre yang muncul secara alami dalam struktur data.

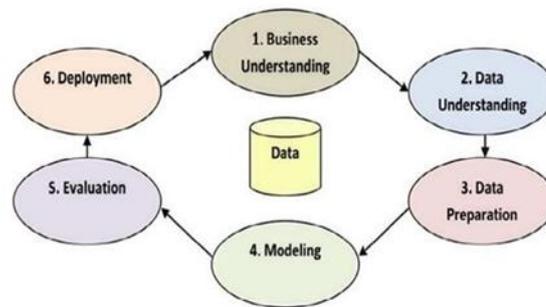
Penelitian ini menutup kesenjangan tersebut dengan:

- a. Menerapkan algoritma K-Medoids untuk mengelompokkan lagu-lagu populer berdasarkan parameter objektif distribusi di Spotify;
- b. Mengintegrasikan pendekatan eksploratif terhadap hasil klasterisasi untuk menemukan subgenre yang dominan;
- c. Memberikan kontribusi praktis dalam membangun sistem rekomendasi berbasis sub-genre yang lebih akurat;
- d. Mengidentifikasi 15 lagu terpopuler berdasarkan klasterisasi data, bukan hanya popularitas absolut.

Dengan pendekatan ini, penelitian ini tidak hanya memperluas pemahaman mengenai pola distribusi popularitas lagu, tetapi juga menawarkan implikasi praktis bagi pengembangan sistem rekomendasi musik dan strategi pemasaran berbasis data.

3. Metode yang Diusulkan

Penelitian ini menggunakan pendekatan Cross-Industry Standard Process for Data Mining (CRISP-DM) sebagai kerangka kerja dalam pengolahan data dan membangun model klasterisasi lagu populer di Spotify. CRISP-DM terdiri dari enam tahapan utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Metode ini dipilih karena memberikan struktur yang jelas dalam proses analisis data, mulai dari pemahaman masalah hingga implementasi hasil yang dapat diterapkan pada sistem rekomendasi musik.



Gambar 1. Metode Penelitian Cross-Industry Standard Process for Data Mining (CRISP-DM)

Business Understanding

Tahap ini berkonsentrasi pada pemahaman terhadap sistem rekomendasi musik di Spotify dan permasalahan dalam klasifikasi genre lagu yang masih terlalu umum. Sistem genre yang ada sering kali menggabungkan lagu dengan karakteristik berbeda dalam kategori yang sama, seperti Pop Ballad dan Dance Pop, sehingga rekomendasi menjadi kurang akurat. Hal ini menyebabkan pengguna mendapatkan lagu yang tidak sesuai dengan preferensi mereka, yang berdampak pada menurunnya pengalaman mendengarkan musik.

Untuk mengatasi masalah ini, penelitian ini bertujuan untuk mengembangkan pengelompokan genre lagu yang lebih spesifik menggunakan algoritma K-Medoids, dengan mempertimbangkan pola popularitas lagu di Spotify. Dengan mengelompokkan lagu

berdasarkan parameter Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity, hasil klusterisasi diharapkan dapat meningkatkan akurasi rekomendasi dan memberikan gambaran lebih lanjut tentang tren musik yang berkembang di Spotify tahun 2024.

Data Understanding

Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle dengan judul Most Streamed Spotify Songs 2024. Dataset ini dipilih karena menyediakan informasi terbaru mengenai lagu-lagu yang paling sering diputarkan di Spotify pada tahun 2024, sehingga relevan untuk menganalisis tren musik digital dan pola distribusi popularitas lagu. Dataset awal terdiri dari 4.600 lagu, dengan pemilihan data didasarkan pada kelengkapan atribut serta relevansinya terhadap tujuan penelitian.

Parameter utama yang digunakan dalam penelitian ini meliputi:

Track dan Artist, digunakan untuk mencari informasi genre dan subgenre lagu dari Spotify, yang kemudian dilabeli secara manual untuk menentukan pengelompokannya.

Spotify Playlist Count, menunjukkan jumlah playlist yang memasukkan lagu tersebut, yang mencerminkan seberapa sering lagu dimasukkan dalam daftar putar oleh pengguna.

Spotify Playlist Reach, mengukur cakupan audiens dari playlist yang memuat lagu, sebagai indikator jangkauan lagu di berbagai segmen pendengar.

Spotify Popularity, menunjukkan tingkat popularitas lagu berdasarkan interaksi pengguna, seperti jumlah pemutaran dan penyimpanan dalam playlist.

Ketiga parameter numerik (Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity) menjadi dasar dalam proses klusterisasi karena mencerminkan keterlibatan pengguna dan visibilitas lagu dalam platform. Dengan pendekatan ini, penelitian bertujuan untuk mengidentifikasi pola distribusi popularitas lagu serta meningkatkan akurasi rekomendasi berbasis subgenre.

Data Preparation

Sebelum masuk ke proses klusterisasi, tahap persiapan data dilakukan untuk memastikan kualitas data setelah memahami karakteristik data. Langkah pertama dalam tahap ini adalah memilih kolom yang relevan, yaitu Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity. Kolom-kolom ini merupakan dasar pembentukan kluster.

Selanjutnya, dilakukan pembersihan data dengan menghapus entri yang duplikat dan memperbaiki nilai yang hilang. Metode StandardScaler dinormalisasi untuk memastikan bahwa setiap parameter memiliki rentang nilai yang sama. Ini memastikan bahwa perbedaan skala antar-fitur tidak menyebabkan bias dalam pembentukan kluster. Distribusi Spotify Playlist Count dan Spotify Playlist Reach menunjukkan kecenderungan skewed right, yang menandakan sebagian besar lagu memiliki eksposur rendah, sedangkan Spotify Popularity lebih mendekati distribusi normal.

Proses persiapan data ini sangat penting karena data yang tidak bersih atau tidak terstandarisasi dapat menyebabkan model klusterisasi menghasilkan kelompok yang tidak optimal. Dengan data yang telah dinormalisasi, proses klusterisasi dapat berjalan dengan lebih akurat dan stabil.

Modeling

Setelah data dipersiapkan, tahap pemodelan dilakukan dengan menerapkan algoritma K-Medoids untuk mengelompokkan lagu berdasarkan popularitas dan distribusinya dalam playlist. Algoritma ini dipilih karena kemampuannya dalam menangani outlier dan menghasilkan kluster yang lebih stabil dibandingkan metode lain seperti K-Means.

Penentuan jumlah kluster optimal dilakukan menggunakan Silhouette Score, yang menghasilkan nilai optimal $k = 3$. Model pertama dibuat tanpa pengurutan popularitas, sementara model kedua dibuat setelah mengurutkan lagu berdasarkan Spotify Popularity untuk mengidentifikasi pola yang lebih spesifik.

Hasil klusterisasi menunjukkan tiga kelompok utama. Pada klusterisasi tanpa pengurutan popularitas, Kluster 0 berisi lagu dengan jumlah playlist rendah tetapi jangkauan luas, Kluster 1 mencakup lagu dengan jumlah playlist dan jangkauan tinggi, dan Kluster 2 berisi lagu dengan popularitas rendah. Setelah dilakukan pengurutan berdasarkan popularitas, Kluster 0 didominasi oleh alternative pop, Kluster 1 tetap mencerminkan lagu populer dengan dominasi contemporary R&B, dan Kluster 2 lebih banyak berisi dance pop.

Dalam model ini, setiap medoid dipilih untuk mewakili kluster tertentu. Medoid adalah titik data dalam kluster yang paling mendekati pusat kluster tersebut, berbeda dengan K-Means yang menggunakan rata-rata titik data sebagai pusatnya. Model ini memberi kemampuan untuk mengelompokkan lagu dengan lebih akurat sesuai dengan pola distribusi popularitasnya.

Evaluation

Evaluasi dilakukan dengan menganalisis hasil klusterisasi berdasarkan karakteristik musik dan mengukur kualitas kluster menggunakan Silhouette Score. Hasil skor sebesar 0.5014 menunjukkan bahwa klusterisasi memiliki pemisahan yang cukup jelas, dengan scatter plot berbasis Principal Component Analysis (PCA) yang menggambarkan distribusi kluster secara visual. Klusterisasi sebelum dan sesudah pengurutan popularitas menunjukkan bahwa pengelompokan berdasarkan subgenre semakin jelas setelah pengurutan dilakukan. Dengan demikian, metode K-Medoids terbukti efektif dalam mengelompokkan lagu berdasarkan popularitas dan distribusinya di Spotify.

Deployment

Hasil klusterisasi ini dapat dimanfaatkan oleh berbagai pemangku kepentingan. Pengembang sistem rekomendasi dapat meningkatkan akurasi pemilihan lagu dengan mempertimbangkan pola popularitas dan distribusinya dalam playlist. Industri musik dapat menggunakan wawasan dari hasil klusterisasi untuk strategi pemasaran berbasis data, seperti penyusunan daftar putar yang lebih sesuai dengan tren genre. Selain itu, akademisi dan peneliti dapat menjadikan penelitian ini sebagai referensi dalam pengembangan lebih lanjut terkait klusterisasi data musik dan penerapan algoritma dalam sistem rekomendasi.

Penelitian ini juga memberikan peluang bagi Spotify untuk meningkatkan sistem pencarian berbasis subgenre. Dengan hasil klusterisasi ini, sistem pencarian lagu dapat menjadi lebih akurat, memungkinkan pengguna untuk menemukan lagu yang lebih sesuai dengan preferensi mereka berdasarkan subgenre yang lebih spesifik.

4. Hasil dan Pembahasan

Pada bagian ini, penulis perlu menjelaskan perangkat keras dan perangkat lunak yang digunakan, sumber dataset, analisis data awal, hasil, serta analisis atau pembahasan hasil. Penyajian hasil menggunakan gambar, grafik, dan tabel sangat dianjurkan. Selain itu, rumus atau alat ukur evaluasi juga harus disertakan di bagian ini. Harus ada diskusi atau analisis; tidak hanya menuliskan ulang hasil dalam bentuk kalimat, tetapi juga menjelaskan ket-erkaitannya dengan hipotesis awal. Selain itu, bagian ini harus membahas dan menguraikan temuan-temuan penting dari penelitian.

Hasil pengelompokan menggunakan algoritma K-Medoids menunjukkan bahwa lagu-lagu dalam dataset dapat dikelompokkan menjadi tiga kluster berdasarkan Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity. Klasterisasi dilakukan dalam dua tahap, yaitu tanpa pengurutan popularitas dan setelah dilakukan pengurutan popularitas, guna memahami bagaimana distribusi popularitas lagu memengaruhi pola klasterisasi.

Business Understanding

Spotify merupakan salah satu platform streaming musik terbesar di dunia dengan lebih dari 600 juta pengguna aktif bulanan pada tahun 2024. Sistem klasifikasi genre yang digunakan masih terlalu luas, menyebabkan beberapa lagu dengan karakteristik berbeda tetap dikelompokkan dalam kategori yang sama. Hal ini berdampak pada ketidaktepatan sistem rekomendasi lagu, sehingga pengguna tidak mendapatkan saran lagu yang sesuai dengan preferensinya. Oleh karena itu, penelitian ini bertujuan untuk meningkatkan klasifikasi genre berdasarkan popularitas lagu dengan menerapkan algoritma K-Medoids. Metode ini dipilih karena kemampuannya dalam menangani outlier dan menghasilkan kluster yang lebih stabil dibandingkan metode lain seperti K-Means.

Dengan mempertimbangkan tiga parameter utama, Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity, penelitian ini diharapkan dapat menghasilkan pengelompokan lagu yang lebih akurat, meningkatkan pengalaman pengguna, serta memberikan wawasan yang lebih mendalam mengenai tren konsumsi musik digital dan pola distribusi popularitas lagu di Spotify.

Data Understanding

Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle dengan judul Most Streamed Spotify Songs 2024. Dataset ini terdiri dari 4.600 lagu, yang dipilih berdasarkan kelengkapan atribut dan relevansinya terhadap penelitian ini. Data ini dianalisis untuk mengidentifikasi nilai yang hilang, duplikasi, dan inkonsistensi sebelum masuk ke proses klasterisasi.

Informasi Dataset:

```
<class 'pandas.core.frame.DataFrame'> RangeIndex: 4600 entries, 0 to 4599 Data columns (total 4 columns):
```

Tabel 1. Informasi Dataset (Hasil Penelitian, 2025).

#	Column	Non-Null Count	Dtype
0	Track	4600 non-null	object
1	Artist	4595 non-null	object
2	Spotify Playlist Reach	4530 non-null	object object
3	Spotify Playlist Count	4528 non-null	float64
4	Spotify Popularity	3796 non-null	

Tabel 1 menyajikan informasi umum mengenai dataset yang digunakan dalam penelitian ini.

Berikut adalah rincian dari setiap kolom dalam dataset:

- a. Track – Berisi judul lagu yang tersedia dalam dataset. Kolom ini memiliki 4.600 nilai non-null, yang berarti tidak ada nilai yang hilang dalam entri judul lagu.
- b. Artist – Berisi nama artis yang menyanyikan lagu tersebut. Kolom ini memiliki 4.595 nilai non-null, menunjukkan bahwa terdapat 5 entri yang kosong, yang kemungkinan merupakan kesalahan dalam pengumpulan data.
- c. Spotify Playlist Count – Menunjukkan jumlah playlist yang memasukkan lagu tersebut. Kolom ini memiliki 4.530 nilai non-null, sehingga terdapat 70 nilai yang hilang, yang mungkin terjadi karena lagu tersebut tidak banyak dimasukkan dalam playlist.
- d. Spotify Playlist Reach – Mengukur jangkauan audiens dari playlist yang memuat lagu tersebut. Dengan 4.528 nilai non-null, terdapat 72 nilai yang hilang, yang mungkin disebabkan oleh keterbatasan data dari playlist dengan jangkauan kecil.
- e. Spotify Popularity – Mengindikasikan tingkat popularitas lagu berdasarkan interaksi pengguna Spotify, seperti jumlah pemutaran dan penyimpanan dalam playlist. Kolom ini memiliki 3.796 nilai non-null, sehingga terdapat 804 nilai yang hilang, yang kemungkinan besar berasal dari lagu-lagu yang kurang populer atau tidak memiliki data interaksi yang cukup.

Dalam penelitian ini, Track dan Artist digunakan untuk mencari informasi genre dan subgenre lagu dari Spotify, yang kemudian dilabeli secara manual untuk menentukan kategori yang lebih spesifik. Sementara itu, tiga parameter numerik (Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity) menjadi dasar dalam proses klusterisasi karena mencerminkan keterlibatan pengguna dan visibilitas lagu dalam platform.

Dari hasil eksplorasi data ini, ditemukan bahwa beberapa atribut memiliki nilai yang hilang, terutama pada Spotify Popularity. Oleh karena itu, dilakukan pembersihan data untuk menghapus entri yang tidak lengkap sebelum masuk ke tahap analisis lebih lanjut. Dengan data yang telah dibersihkan dan dinormalisasi, proses klusterisasi dapat berjalan dengan lebih akurat dan stabil.

Data Preparation

Penanganan Missing Values

Dalam proses pembersihan data, langkah pertama yang dilakukan adalah mengidentifikasi jumlah missing values di setiap kolom untuk mengetahui sejauh mana kekurangan data dapat memengaruhi analisis. Hal ini dilakukan dengan metode `.isnull().sum()`, yang memungkinkan kita melihat jumlah nilai yang hilang di setiap atribut, terutama pada Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity. Setelah itu, dilakukan penghapusan missing values menggunakan `.dropna()`, seperti yang ditunjukkan pada Gambar 2, sehingga dataset yang awalnya memiliki beberapa entri dengan nilai kosong kini berkurang menjadi 3.795 baris dan 5 kolom.

Meskipun metode ini efektif dalam memastikan hanya data yang lengkap yang digunakan, ada alternatif lain untuk menangani missing values tanpa menghapus baris, seperti menggunakan imputasi dengan nilai statistik. Teknik ini menggantikan nilai kosong dengan mean jika distribusi data normal, atau median jika terdapat outliers yang berpotensi mengganggu hasil analisis. Alternatif lain yang juga dapat diterapkan adalah interpolasi, yang mengisi nilai kosong dengan pendekatan linear berdasarkan pola data yang ada. Namun, dalam penelitian ini, metode `.dropna()` dipilih untuk memastikan hanya data yang sepenuhnya bersih yang digunakan dalam proses analisis klusterisasi menggunakan algoritma K-Medoids.

```

data = data.dropna()
print(f"Data setelah menghapus missing values: {data.shape}")

track_column = data[['Track', 'Artist']]

scaler = StandardScaler()
data_scaled = scaler.fit_transform(data[['Spotify Playlist Count',
                                         'Spotify Playlist Reach',
                                         'Spotify Popularity']])

print("Data berhasil dinormalisasi.\n")
    
```

Data setelah menghapus missing values: (3795, 5)
 Data berhasil dinormalisasi.

Gambar 2. Menghapus Missing Value (Hasil Penelitian, 2025)

Normalisasi Data

Normalisasi data menggunakan StandardScaler bertujuan untuk menyelaraskan skala antarvariabel dalam dataset guna menghindari dominasi satu fitur terhadap fitur lainnya dalam proses klusterisasi. Pada penelitian ini, normalisasi diterapkan pada tiga atribut numerik utama, yaitu Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity. Metode StandardScaler bekerja dengan mentransformasikan data sehingga memiliki distribusi dengan rata-rata nol dan standar deviasi satu.

Normalisasi menjadi langkah krusial dalam penerapan algoritma K-Medoids karena algoritma ini mengandalkan metrik jarak, seperti Euclidean Distance, dalam menentukan pusat medoid dan mengelompokkan data. Tanpa normalisasi, perbedaan skala antarvariabel dapat menyebabkan bias dalam perhitungan jarak, di mana fitur dengan nilai yang lebih besar akan lebih berpengaruh dibandingkan fitur lainnya. Dengan normalisasi, setiap variabel berkontribusi secara seimbang, sehingga klaster yang terbentuk lebih representatif dan sesuai dengan karakteristik data sebenarnya.

```

# Menampilkan hasil normalisasi
print("\n=== Hasil Normalisasi Data ===")
print(normalized_df.head())
    
```

=== Hasil Normalisasi Data ===

	Track	Artist	Playlist Count (Normalized)
0	MILLION DOLLAR BABY	Tommy Richman	-0.529306
1	Not Like Us	Kendrick Lamar	-0.564788
2	i like the way you kiss me	Artemas	-0.207404
3	Flowers	Miley Cyrus	2.729730
4	Houdini	Eminem	-0.849544

	Playlist Reach (Normalized)	Popularity (Normalized)
0	5.496221	1.761932
1	4.786041	1.761932
2	5.978907	1.761932
3	3.560378	1.328959
4	4.040638	1.514519

Gambar 3. Normalisasi Data (Hasil Penelitian, 2025)

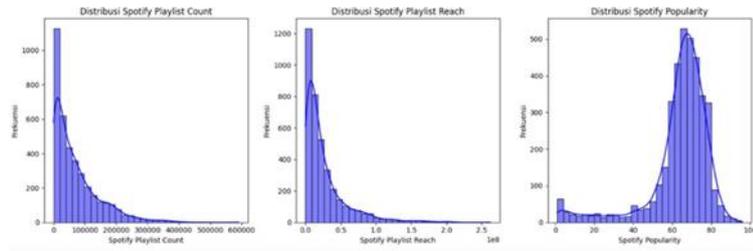
Distribusi Data

Gambar 4 menunjukkan histogram distribusi dari tiga atribut numerik utama dalam dataset yaitu Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity. Histogram pada Spotify Playlist Count dan Spotify Playlist Reach menunjukkan distribusi yang sangat skewed ke kanan, yang mengindikasikan bahwa sebagian besar lagu hanya muncul di sedikit playlist dan memiliki jangkauan yang rendah, sementara hanya sebagian kecil lagu yang memiliki eksposur luas. Sebaliknya, atribut Spotify Popularity cenderung mengikuti distribusi normal, dengan mayoritas lagu berada dalam rentang popularitas menengah antara 40 hingga 80.

Analisis distribusi ini, sebagaimana ditampilkan pada Gambar 4, memiliki implikasi signifikan dalam proses klusterisasi lagu berdasarkan popularitas dan jangkauan playlist. Distribusi yang tidak merata dapat menyebabkan bias dalam penentuan pusat kluster jika tidak

ditangani dengan baik, misalnya melalui normalisasi data. Selain itu, pemahaman terhadap pola distribusi membantu dalam memilih parameter yang lebih optimal dalam algoritma K-Medoids untuk memastikan bahwa kluster yang terbentuk benar-benar mencerminkan segmentasi musik yang sesuai dengan perilaku pengguna di Spotify.

Gambar 4. Distribusi Data (Hasil Penelitian, 2025)



Modeling

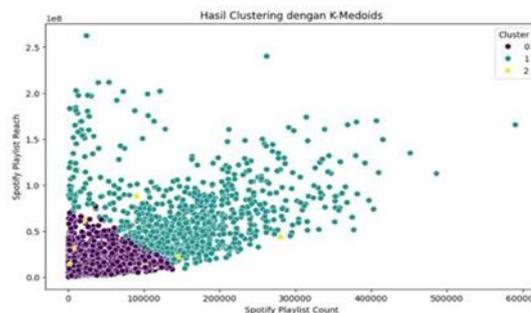
Scatterplot

Sebelum menerapkan algoritma K-Medoid, dilakukan analisis eksploratori menggunakan scatterplot untuk memahami distribusi lagu berdasarkan jumlah playlist (Spotify Playlist Count) dan jangkauan (Spotify Playlist Reach).

Gambar 5 menunjukkan hasil klusterisasi dengan scatterplot, di mana setiap titik merepresentasikan satu lagu, dengan warna yang berbeda menunjukkan kluster yang terbentuk. Dari scatterplot ini, dapat diamati pola distribusi sebagai berikut:

- a. Cluster 0 (Ungu): Lagu-lagu yang termasuk dalam kluster ini memiliki jumlah playlist rendah,
- b. lagu niche yang memiliki banyak pemutaran dalam lingkup tertentu tetapi tidak tersebar luas.

Secara keseluruhan, hasil menunjukkan bahwa distribusi lagu di Spotify tidak merata, dengan sebagian besar lagu memiliki jumlah playlist dan jangkauan yang rendah, sementara hanya sebagian kecil yang memiliki eksposur sangat tinggi. Pola ini kemudian divalidasi lebih lanjut melalui proses klusterisasi menggunakan algoritma K-Medoids.



Gambar 5. Scatterplot (Hasil Penelitian, 2025)

b. Identifikasi Pola Popularitas Lagu

Hasil klusterisasi memberikan wawasan mengenai bagaimana lagu-lagu dengan karakteristik tertentu, seperti jumlah playlist dan jangkauan, dapat dikelompokkan dengan baik. Ini membantu dalam memahami tren musik yang berkembang di Spotify dan memberikan informasi yang lebih mendalam tentang pola distribusi lagu dalam ekosistem digital.

c. Dukungan untuk Strategi Musik dan Pemasaran

Hasil klusterisasi ini dapat digunakan untuk menentukan kategori lagu yang memiliki potensi untuk ditingkatkan popularitasnya serta lagu-lagu yang sudah mendominasi dalam distribusi playlist. Informasi ini dapat dimanfaatkan oleh industri musik dalam menyusun strategi promosi yang lebih efektif, baik dalam bentuk kampanye pemasaran digital maupun dalam rekomendasi berbasis preferensi pengguna. Secara keseluruhan, pemodelan K-Medoids efektif dalam mengelompokkan lagu berdasarkan popularitas dan distribusinya, memberikan hasil klusterisasi yang jelas dan dapat digunakan sebagai dasar analisis tren musik lebih lanjut.

Deployment

Implementasi Hasil

Berdasarkan hasil analisis, 15 lagu terpopuler di Spotify 2024 telah berhasil diidentifikasi. Lagu-lagu ini memiliki tingkat popularitas yang tinggi berdasarkan jumlah playlist dan reach yang dicapai. Keberhasilan pengurutan ini menunjukkan bahwa metode yang digunakan mampu menyusun daftar lagu populer dengan akurat dan dapat menjadi dasar dalam pengembangan sistem rekomendasi yang lebih presisi.

Selain itu, proses klusterisasi juga berjalan dengan baik, mengelompokkan lagu-lagu berdasarkan pola tertentu yang muncul dalam data. Klusterisasi ini memungkinkan adanya pemetaan subgenre yang lebih spesifik, yang dapat menjadi pertimbangan bagi Spotify dalam meningkatkan pengalaman pengguna. Dengan adanya subgenre yang lebih terperinci, pengguna dapat menikmati rekomendasi lagu yang lebih sesuai dengan preferensi mereka, sehingga pengalaman mendengarkan menjadi lebih personal dan relevan.

15 Lagu Terpopuler:

	Track	Artist
36	A Bar Song (Tipsy)	Shaboozey
21	Espresso	Sabrina Carpenter
0	MILLION DOLLAR BABY	Tommy Richman
12	LUNCH	Billie Eilish
2	i like the way you kiss me	Artemas
1	Not Like Us	Kendrick Lamar
40	BIRDS OF A FEATHER	Billie Eilish
7	Gata Only	FloyyMenor
10	I Had Some Help (feat. Morgan Wallen)	Post Malone
734	Tell Ur Girlfriend	Lay Bankz
920	End of Beginning	Djo
78	CHIHITO	Billie Eilish
16	Fortnight (feat. Post Malone)	Taylor Swift
13	Like That	Future
452	Cruel Summer	Taylor Swift

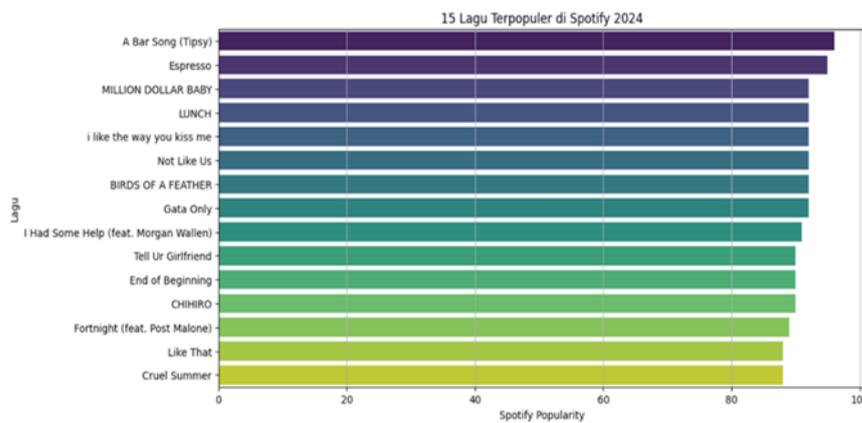
	Spotify Playlist Count	Spotify Playlist Reach	Spotify Popularity
36	25766	180237463	96.0
21	24425	262343414	95.0
0	30716	196631588	92.0
12	13800	197280692	92.0
2	54331	211607669	92.0
1	28113	174597137	92.0
40	10871	202626837	92.0
7	40094	211236940	92.0
10	16219	168766158	91.0
734	10335	116287670	90.0
920	64517	191969474	90.0
78	11028	144854768	90.0
16	12784	177932568	89.0
13	43025	108168634	88.0
452	121479	201964859	88.0

Gambar 13. Lagu Terpopuler Spotify 2024 (Hasil Penelitian, 2025)

Grafik

Untuk mendukung hasil analisis, grafik batang digunakan sebagai alat bantu visualisasi data dalam menampilkan pengurutan lagu terpopuler di Spotify 2024. Grafik ini mempermudah pemahaman terhadap perbedaan tingkat popularitas masing-masing lagu dan memberikan gambaran yang lebih jelas mengenai dominasi beberapa lagu tertentu.

Dari grafik, terlihat bagaimana distribusi popularitas lagu-lagu yang masuk dalam daftar 15 besar. Variasi warna dalam grafik membantu membedakan setiap lagu, sehingga mempermudah interpretasi data. Dengan visualisasi ini, pihak industri musik dapat lebih mudah menganalisis tren yang sedang berkembang dan menggunakan informasi tersebut untuk strategi pemasaran atau rekomendasi lagu yang lebih efektif.



Gambar 14. Grafik Batang Lagu Terpopuler Spotify 2024 (Hasil Penelitian, 2025)

5. Kesimpulan

Penelitian ini berhasil mengelompokkan lagu di Spotify berdasarkan popularitas dan distribusinya dalam playlist menggunakan algoritma K-Medoids. Hasil klusterisasi mengidentifikasi tiga kelompok utama yaitu lagu dengan jangkauan luas namun sedikit playlist, lagu dengan popularitas tinggi dan distribusi luas, serta lagu dengan tingkat popularitas rendah. Evaluasi menggunakan Silhouette Score sebesar 0.5014 menunjukkan pemisahan klaster yang baik, membuktikan efektivitas K-Medoids dalam analisis popularitas musik digital.

Temuan ini memiliki implikasi praktis bagi industri musik dan platform streaming, terutama dalam merancang strategi pemasaran yang lebih terarah. Klusterisasi dapat membantu dalam mengidentifikasi lagu dengan potensi peningkatan popularitas serta mendukung pengembangan sistem rekomendasi berbasis subgenre untuk meningkatkan relevansi bagi pengguna. Selain itu, penelitian ini juga mengidentifikasi 15 lagu terpopuler di Spotify tahun 2024 berdasarkan jangkauan dan distribusinya.

Untuk penelitian selanjutnya, disarankan penggunaan dataset yang lebih besar dan lebih beragam guna meningkatkan generalisasi hasil. Kombinasi algoritma lain, seperti DBSCAN atau Hierarchical Clustering, juga dapat dieksplorasi untuk memahami pola distribusi musik secara lebih mendalam. Selain itu, integrasi data lain, seperti durasi mendengarkan atau interaksi pengguna, dapat memberikan wawasan lebih luas terhadap tren konsumsi musik di platform streaming

Daftar Pustaka

- [1]. Anoohya, B., Nikitha, J., Jyothi, N., Anish, I., & Bhargavi, G. (2023). Recommendation system for song data using K-Means and K-Medoids clustering algorithm. www.ijnrd.org
- [2]. Ardiansyah, B., Daulay, I., & Hutagaol, R. (n.d.). K-Means and Decision Tree algorithm for prediction of postgraduate students admission in University of Indonesia. SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat. <https://journal.irpi.or.id/index.php/sentimas>

- [3]. Firmansyah, M. I., Rohman, R. S., Marsusanti, E., Bina, U., Informatika, S., & Disetujui, D. D. (2023). Penerapan algoritma kluster K-Means untuk fitur atribut pada layanan streaming musik Spotify. *Journal Computer Science*, 2(2).
- [4]. Fitriadi, N. R., Hidayat, M. F., Saputro, T. W., Alifian, M. G., & Sari, A. P. (2023). Laporan final project machine learning rekomendasi musik Spotify menggunakan metode K-Means.
- [5]. Gandhi, G., & Srivastava, R. (n.d.). Analysis and implementation of modified K-Medoids algorithm to increase scalability and efficiency for large dataset. <http://www.ijret.org>
- [6]. Irawan, E., Siregar, S. P., Damanik, I. S., Saragih, I. S., Studi Sistem Informasi, P., & Tunas Bangsa, S. A. (n.d.). Implementasi K-Medoids untuk pengelompokan sebaran maha-siswa baru (Eka Irawan). *Jurnal Riset Sistem Informasi dan Teknik Informatika (JURASIK)*, 5, 275–281. <https://tunasbangsa.ac.id/ejournal/index.php/jurasik>
- [7]. Junaidi, K., Aurelie, A. C., Israwana, S., Putra, A., & Ryu, A. (2023). Implementasi MapReduce pada dataset Spotify top music untuk mengetahui artist yang paling banyak didengar dalam kurun waktu 10 tahun. *Jurnal Teknologi*, 2(2). <https://jurnal.net-plg.com/>
- [8]. Kim, S., Park, J., Seong, K., Cho, N., Min, J., & Hong, H. (2021). Music-circles: Can music be represented with numbers? <http://arxiv.org/abs/2102.13350>
- [9]. Mantik, J., & Privandhani, N. A. (2022). (2022) 1542-1550 Accredited.
- [10]. Marlia, S., et al. (n.d.). Analisis fitur musik dan tren popularitas lagu di Spotify menggunakan K-Means dan CRISP-DM. *SISTEMASI: Jurnal Sistem Informasi*. <http://sistemasi.ftik.unisi.ac.id>
- [11]. Mufidah, M., Amirah, A., Widodo, A. W., & Dewi, C. (2017). Pengelompokan lagu ber-dasarkan emosi menggunakan algoritma Fuzzy C-Means. <http://j-ptiik.ub.ac.id>
- [12]. Nisa, C., & Yustanti, W. (2021). Studi perbandingan algoritma kluster dalam penge-lompokan persediaan produk (Studi kasus: Subdirektorat Perencanaan Sarana Prasarana dan Logistik PTN X). *Journal of Emerging Information Systems and Business Intelligence (JEISBI)*, 2(3), 14–20. <https://ejournal.unesa.ac.id/index.php/JEISBI/article/view/41103>
- [13]. Pangestu, C., Shaufiah, S., & Wijaya, R. (2024). X Spotify cares clustering analysis using K-Means and K-Medoids. *Jurnal Media Informatika Budidarma*, 8(1), 497. <https://doi.org/10.30865/mib.v8i1.7279>
- [14]. Ramadhani, J., Anugraha, Y. S., Fauzan, A., & Efrizoni, L. (2024). Jilang Ramadhani 1, Yoga Safitra Anugraha 2. <http://ojsamik.amikmitragama.ac.id>
- [15]. Rohman, N., & Wibowo, A. (2024). Clustering of popular Spotify songs in 2023 using K-Means method and silhouette coefficient. *Jurnal Pilar Nusa Mandiri*, 20(1), 18–24. <https://doi.org/10.33480/pilar.v20i1.4937>
- [16]. Sari, B. N. (2016). Identification of tuberculosis patient characteristics using K-Means clustering. *Scientific Journal of Informatics*, 3(2), 129–138. <https://doi.org/10.15294/sji.v3i2.7909>
- [17]. Sidora, L. I., & Harani, N. H. (2023). Sistem rekomendasi musik Spotify menggunakan KNN dan algoritma genetika.
- [18]. Suastrri, L. (n.d.). 8_Clustering pop songs based on Spotify data using K-Means.
- [19]. Werner, A. (2020). Organizing music, organizing gender: Algorithmic culture and Spotify recommendations. *Popular Communication*, 18(1), 78–90. <https://doi.org/10.1080/15405702.2020.1715980>
- [20]. Zaidah, A. R., Septiarani, C. I., Nisa, S., Yusuf, A., & Wahyudi, N. (2021). Komparasi algoritma K-Means, K-Medoid, Agglomerative clustering terhadap genre Spotify. *Jurnal Ilmiah Ilmu Komputer*, 7(1), 49–54. <https://ejournal.fikom-unasman.ac.id>