 ISSN:2302-8149
e-ISSN:2540-9719

Analisis Fitur Musik dan Tren Popularitas Lagu di Spotify menggunakan K-Means dan CRISP-DM

Analysis of Music Features and Song Popularity Trends on Spotify Using K-Means and CRISP-DM

[1]Sari Marlia*, [2] Kiki Setiawan, [3]Christina Juliane
[1,2,3]Sistem Informasi Bisnis, Magister Sistem Informasi, STMIK LIKMI Bandung
Jalan Ir. H. Juanda No. 96, Lebakgede, Bandung, Indonesia
*e-mail: yooniversari@gmail.com

(***received*:** 13 December 2023, ***revised*:** 23 December 2024, ***accepted*:** 09 January 2024)

Abstrak

Spotify yang dikenal sebagai salah satu *platform streaming* musik terbaik, telah berperan penting dalam mengubah bagaimana pendengar mengakses, menikmati, dan berinteraksi dengan musik. Dengan adanya jutaan lagu dan data pengguna yang luas, *Spotify* memberikan peluang untuk memahami perilaku pendengar dan faktor-faktor yang berkontribusi terhadap keberhasilan dan popularitas sebuah lagu. Penelitian ini memiliki tujuan untuk mengkaji hubungan antara fitur-fitur musik dan popularitas lagu pada *platform* musik Spotify dengan melakukan analisis nilai SSE, nilai *euclidean distance*, dan nilai pusat klaster pada atribut *dataset loudness*, *danceability*, dan energi. Kerangka kerja yang digunakan dalam penelitian ini adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Algoritma klastering *K-Means* dan aplikasi olah *data mining Weka* digunakan untuk menguraikan fitur-fitur yang mempengaruhi kesuksesan dan popularitas lagu di *Spotify*. Hasil penelitian menunjukan bahwa kelompok/klaster 1, 2, dan 3 merupakan kelompok/klaster dengan lagu-lagu yang memiliki *loudness*, *danceability*, dan energi yang tinggi, sedang, dan rendah secara berurut. Lagu populer di Spotify saat ini semakin berfokus pada *loudness*, *danceability*, dan energi dengan tren yang menonjol, yaitu lagu-lagu dengan *loudness*, *danceability*, dan energi yang tinggi semakin populer, sementara lagu-lagu dengan *loudness*, *danceability*, dan energi yang rendah semakin kurang populer.

Kata kunci: CRISP-DM, K-Means, klastering, popularitas lagu, Spotify.

Abstract

Spotify, known as one of the best music streaming platforms, has played an important role in changing how listeners access, enjoy and interact with music. With millions of songs and extensive user data, Spotify provides an opportunity to understand listener behavior and the factors that contribute to a song's success and popularity. This research aims to examine the relationship between music features and the popularity of songs on the Spotify music platform by analyzing SSE values, Euclidean distance values, and cluster center values on the dataset attributes loudness, danceability, and energy. The framework used in this research is CRISP-DM (Cross-Industry Standard Process for Data Mining). The K-Means clustering algorithm and the Weka data mining application are used to decipher the features that influence the success and popularity of songs on Spotify. The research results show that groups/clusters 1, 2, and 3 are groups/clusters with songs that have high, medium, and low loudness, danceability, and energy respectively. Popular songs on Spotify are currently increasingly focused on loudness, danceability, and energy with a prominent trend, namely songs with high loudness, danceability, and energy are becoming more popular, while songs with low loudness, danceability, and energy are becoming less popular.

Keywords: *Clustering, CRISP-DM, K-Means, song's popularity, Spotify.*

 ISSN:2302-8149
e-ISSN:2540-9719

1 Pendahuluan

Musik telah hadir di dalam kehidupan manusia diawali dari jaman Yunani kuno hingga ke jaman modern [1]. Industri musik telah mengalami perkembangan yang signifikan dalam beberapa dekade terakhir terutama sejak era digital dan kemunculan *platform streaming* musik [2]. Sebagai salah satu pemimpin dalam industri *streaming* musik, Spotify telah memainkan peran sentral dalam mengubah cara pendengar mendapatkan, mengonsumsi, dan berinteraksi dengan musik. Dengan kehadiran jutaan lagu dan data pengguna yang besar, *Spotify* menyediakan kesempatan yang unik untuk memahami perilaku para pendengar dan faktor-faktor yang memengaruhi kesuksesan dan popularitas lagu.

Penelitian ini bertujuan untuk mengkaji hubungan antara fitur-fitur musik dan popularitas lagu dalam konteks *Spotify*. Dengan melakukan analisis fitur-fitur seperti *duration, danceability, energy, loudness, speechiness, accousticness, instrumentalness, liveness, valence, tempo,* dan *time signature*, maka elemen-elemen yang memengaruhi preferensi pendengar dan membantu memprediksi kesuksesan lagu dapat diidentifikasi. Tolak ukur yang digunakan untuk menentukan kesuksesan lagu pada studi ini adalah peringkat/*chart Billboard HOT* 100, *Billboard Global* 200 *excluding US*, dan *Billboard Global* 200 [3].

Dalam penelitian ini, metode analisis data/kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) dan algoritma klastering *K-Means* akan digunakan untuk menguraikan fitur-fitur yang mempengaruhi kesuksesan dan popularitas lagu di *Spotify*. Perangkat lunak yang digunakan untuk mengolah *dataset* adalah dengan menggunakan *Weka*. Hasil penelitian ini diharapkan akan memberikan pemahaman yang lebih dalam tentang bagaimana karakteristik musik berdampak pada kesuksesan dan popularitasnya di era digital saat ini.

Penelitian ini tidak hanya dapat memberikan wawasan mendalam terhadap industri musik, tetapi juga memiliki potensi untuk memberikan manfaat bagi para seniman, produsen musik, dan pemasar dalam pengembangan strategi yang lebih efektif dalam menciptakan musik yang populer. Dengan demikian, penelitian ini mengisi celah pengetahuan yang signifikan dalam memahami hubungan antara fitur musik dan popularitas lagu di era *Spotify*. Dengan demikian, artikel ini membuka jalan bagi pemahaman lebih mendalam tentang faktor-faktor yang memengaruhi popularitas lagu di *platform streaming* seperti *Spotify*, dan memberikan kontribusi penting dalam perkembangan ilmu musik dan analisis data.

2 Tinjauan Literatur

Beberapa studi sebelumnya telah mengeksplorasi faktor-faktor yang berkontribusi pada kesuksesan dan popularitas sebuah lagu dalam tangga lagu, termasuk melalui analisis fitur-fitur musiknya. Dengan pertumbuhan jumlah data yang terus meningkat dan kemajuan teknologi analitik, penelitian lebih lanjut diperlukan untuk menggali lebih dalam faktor-faktor yang mendasari kesuksesan lagu di *Spotify*.

Studi yang dilakukan oleh Musyarofah, dkk [4] menjelaskan bagaimana lagu pop paling populer dari tahun 2010 hingga 2019 diklasifikasikan menggunakan metode *K-Means* di *Spotify*. Hasil penelitian menunjukan bahwa tiga teratas (secara berurutan) persentase genre lagu dimiliki oleh genre pop sebesar 83,9%, genre EDM sebesar 3,05%, dan genre Boyband dan Hip Hop sebesar 2,54%. Sementara, persentase terendah dimiliki oleh genre permanent wave yaitu sebesar 0,678%.

Berbeda dengan studi yang dilakukan oleh Navisa, dkk [5], studi ini membandingkan algoritma klasifikasi mana yang menghasilkan performa terbaik berdasarkan proses *data mining* menggunakan CRISP-DM yaitu *Naive Bayes*, K-NN dan *Random Forest*. Studi tersebut memberikan informasi bahwa akurasi terbaik diperoleh pada algoritma *Naïve Bayes* dengan nilai sebesar 58,91% sedangkan algoritma yang memiliki kinerja terbaik adalah K-NN dan *Random Forest* dengan nilai 0,528.

Sedangkan studi yang dilakukan oleh Interiano, dkk [6] adalah menganalisis lebih dari 500.000 lagu yang dirilis di Inggris antara 1985 dan 2015 untuk memahami dinamika keberhasilan lagu, mencari korelasi antara fitur akustik dan keberhasilan, serta mengeksplorasi prediktabilitas keberhasilan lagu dengan menggunakan algoritma *random forests*. Studi terhadap 500.000 lagu yang dirilis dalam 30 tahun terakhir menemukan bahwa lagu yang sukses cenderung memiliki fitur musik yang lebih ceria, lebih bersemangat untuk berpesta, dan lebih mudah didengar. Studi ini juga berhasil memprediksi kesuksesan lagu dengan akurasi 74% hanya berdasarkan fitur akustik. Namun, penambahan variabel artis terkenal meningkatkan akurasi prediksi menjadi 86%, menunjukkan peran

 ISSN:2302-8149
e-ISSN:2540-9719

penting faktor non-musikal dalam kesuksesan. Studi ini menyimpulkan bahwa fitur musik memainkan peran penting dalam kesuksesan lagu, namun faktor non-musik seperti artis terkenal juga sangat berpengaruh.

Spotify merupakan salah satu dari sekian *platform streaming* musik yang terkenal dan diluncurkan sejak 7 Oktober 2008. *Spotify* hadir sejak teknologi internet semakin berkembang yang ditandai dengan *audio streaming* yang kian populer [7]. *Spotify* adalah layanan musik digital, *podcast*, dan video yang memberi akses ke jutaan lagu dan konten lain dari kreator di seluruh dunia dengan fungsi dasar seperti memutar musik tidak berbayar, tetapi pengguna juga bisa memilih untuk melakukan *upgrade* ke *Spotify* Premium. Dengan mengaktifkan premium, pengguna mendapatkan rekomendasi berdasarkan selera, membangun koleksi musik dan podcast dan lain-lain [8].

Data mining merupakan proses mencari anomali, pola, dan korelasi dalam *dataset* besar untuk memprediksi hasil yaitu dengan menggunakan berbagai teknik supaya informasi ini dapat digunakan untuk meningkatkan pendapatan, mengurangi biaya, meningkatkan hubungan dengan pelanggan, mengurangi risiko, dan lainnya [9]. Sedangkan menurut Agarwal [10] *data mining* adalah bidang perpotongan antara ilmu komputer dan statistik yang digunakan untuk menemukan pola dalam bank informasi yang memiliki tujuan utama adalah untuk mengekstrak informasi berguna dari berkas data dan membentuknya menjadi struktur yang dapat dipahami untuk penggunaan di masa depan.

Metode CRISP-DM merupakan metode yang cukup banyak diterapkan dalam data mining. Ada 6 tahapan di dalam metode ini, yaitu (1) *Business understanding*; (2) *Data understanding*; (3) *Data preparation*; (4) *Modelling*; (5) *Evaluation*; (6) *Deployment* [11].

Algoritma *K-Means* merupakan metode yang paling populer dan sederhana dalam klasterisasi data yang tujuannya adalah untuk membagi data menjadi kelompok (klaster) berdasarkan kesamaan fitur. Algoritma *K-Means* memerlukan nilai k yang harus ditentukan sebelum melakukan analisis klaster. Jika menggunakan nilai k yang berbeda, maka hasil pengelompokan juga akan berbeda. Beberapa penelitian terbaru telah menganalisis masalah inisialisasi yang berbeda, tetapi tidak mempertimbangkan situasi di mana algoritma hanya berkonvergensi ke lokal mininum yang buruk. Maksudnya adalah kadang-kadang algoritma *K-Means* bisa terjebak dalam solusi yang tidak optimal. Studi yang dilakukan Ikotun, dkk [12] oleh bertujuan untuk mengidentifikasi, mengambil, merangkum, dan menganalisis studi yang baru-baru ini diajukan yang terkait dengan peningkatan algoritma pengelompokan *K-Means* dengan teknik optimasi yang terinspirasi dari alam. Studi yang dilakukan oleh Tan, dkk [13] mengembangkan skema yang unifikasi untuk mempelajari nilai k dan memilih pusat awal untuk pengelompokan *K-Means* intinsik pada jenis yang sifatnya homogen.

Ada beberapa perangkat lunak yang dapat digunakan untuk mengolah *data mining*. Studi ini akan menggunakan Weka sebagai perangkat untuk mengolah data. Sama seperti perangkat pengolah *data mining* lainnya, Weka juga memiliki kelebihan dan kekurangan seperti yang tercantum pada Tabel 1, yaitu:

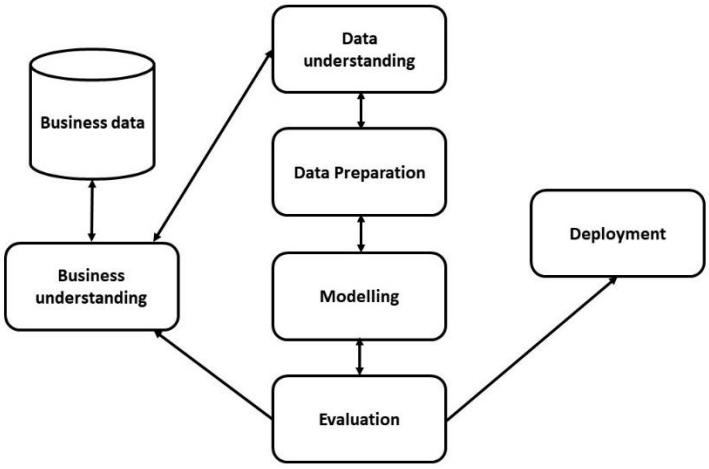
Tabel 1. Kelebihan dan Kekurangan Weka [14]

Kelebihan	Kekurangan
Antarmuka pengguna yang intuitif	Kurangnya dukungan komunitas yang aktif
Fleksibel	Memiliki keterbatasan dalam menangani volume data yang sangat besar
Mampu memroses berbagai format data	
Memiliki pendekatan algoritmik untuk menangani permasalahan klastering	
Memiliki fitur visualisasi yang dapat memberikan representasi data hasil proses data mining dalam bentuk gambar/grafik.	

3 Metode Penelitian

Penelitian ini dilakukan dengan menggunakan metode CRISP-DM yang memiliki alur proses seperti pada Gambar 1.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 1. Alur CRISP-DM [15]

Dataset diunduh dari *website Kaggle* dengan judul *Top Spotify Songs in 73 Countries (Daily Updated)* dengan tanggal *cut-off* data yaitu pada tanggal 31 Oktober 2023. Data mentah sejumlah 51.105 baris dengan 25 buah atribut, setelah dilakukan *data cleaning* jumlah baris menjadi 47.157 baris dengan jumlah atribut yang masih sama yaitu 25 buah. Aplikasi yang digunakan untuk mengolah data yaitu Weka dengan menerapkan metode *CRISP-DM* sebagai berikut:

3.1. Business Understanding

Pemahaman bisnis atau masalah dalam penelitian ini menunjuk kepada pemahaman tentang faktor-faktor apa saja yang mempengaruhi kesuksesan sebuah lagu di *Spotify*, dengan harapan dapat membantu industri musik dalam memproduksi lagu-lagu yang lebih disukai oleh pendengar dan berpotensi sukses di *chart* musik.

3.2. Data Understanding

Pada tahap pemahaman data di dalam penelitian ini mengacu pada dataset lagu-lagu teratas dari 73 negara seperti pada Tabel 2 berikut:

Tabel 2. Deskripsi atribut *dataset* [16]

Atribut	Indikator	Ukuran	Tipe data
spotify_id	Pengidentifikasi unik untuk lagu di basis data *Spotify*.	kategorikal	*string*
name	Judul lagu.	kategorikal	*string*
artists	Nama artis yang terkait dengan lagu.	kategorikal	*string*
daily_rank	Peringkat harian lagu dalam daftar *top* 50.	numerik	*integer*
daily_movement	Perubahan peringkat dibandingkan dengan hari sebelumnya.	kategorikal	*integer*
weekly_movement	Perubahan peringkat dibandingkan dengan minggu sebelumnya.	kategorikal	*integer*
country	Kode ISO negara dari *Playlist Top* 50. Jika *null*, maka *playlist* adalah *Global Top 50*.	kategorikal	*string*
snapshot_date	Tanggal ketika data dikumpulkan dari *API Spotify*.	kategorikal	*string*
popularity	Ukuran popularitas lagu saat ini di *Spotify*.	numerik	*integer*
is_explicit	Menunjukkan apakah lagu mengandung lirik eksplisit.	kategorikal	*boolean*
duration_ms	Durasi lagu dalam milidetik.	numerik	*integer*
album_name	Judul album tempat lagu tersebut berada.	kategorikal	*string*
album_release_date	Tanggal rilis album tempat lagu tersebut berada.	kategorikal	*string*
danceability	Ukuran seberapa cocok lagu untuk menari	numerik	*float*

 ISSN:2302-8149
e-ISSN:2540-9719

Atribut	Indikator	Ukuran	Tipe data
	berdasarkan berbagai elemen musik.		
energy	Ukuran intensitas dan tingkat aktivitas lagu.	numerik	*float*
key	Kunci lagu.	kategorikal	*integer*
loudness	Kekerasan keseluruhan lagu dalam desibel.	numerik	*float*
mode	Menunjukkan apakah lagu berada dalam kunci mayor atau minor.	kategorikal	*integer*
speechiness	Ukuran keberadaan kata-kata yang diucapkan dalam lagu.	numerik	*float*
acousticness	Ukuran kualitas akustik lagu.	numerik	*float*
instrumentalness	Ukuran kemungkinan lagu tidak mengandung vokal.	numerik	*float*
liveness	Ukuran keberadaan penonton langsung dalam rekaman.	numerik	*float*
valence	Ukuran positivitas musikal yang disampaikan oleh lagu.	numerik	*float*
tempo	Tempo lagu dalam beat per menit.	numerik	*float*
time_signature	Tanda waktu keseluruhan lagu.	kategorikal	*integer*

Koefisien korelasi *Pearson* akan diterapkan pada tahapan ini. Tujuannya adalah agar dapat melihat hubungan antara variabel-variabel yang ada di dalam pemahaman data. Koefisien korelasi *Pearson* dapat digunakan untuk mengukur kekuatan hubungan linier antara dua variabel. Rumus koefisien korelasi Pearson adalah seperti persamaan (1) berikut:

$$r_{XY} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \quad (1)$$

Keterangan:
r_{XY} = koefisien korelasi; n = banyak data; x = variabel pertama; y = variabel kedua

Ukuran hubungan antara variabel ditunjukkan oleh sebuah angka yang dikenal sebagai koefisien korelasi. Rentang nilai dari koefisien korelasi adalah antara -1 hingga +1. Koefisien korelasi dengan nilai positif menandakan adanya hubungan yang searah antara satu variabel dengan variabel lainnya. Di sisi lain, koefisien korelasi dengan nilai negatif menandakan adanya hubungan yang berlawanan arah antara satu variabel dengan variabel lainnya. Nilai koefisien korelasi yang mendekati -1 atau 1 mengindikasikan adanya hubungan yang kuat. Sebaliknya, jika nilai koefisien korelasi mendekati 0, ini menandakan adanya hubungan yang lemah [17].

3.3. Data Preparation

Pada tahap persiapan data dilakukan aktivitas menyusun data mentah dari *dataset* untuk memilah, membersihkan, menyusun, dan mengintegrasikan data [5]. Dalam penelitian ini dilakukan pembersihan data. Tahap pembersihan data ini perlu dilakukan. Dalam artikel [18] disebutkan bahwa pembersihan data, bertujuan untuk memperbaiki kualitas data dengan menemukan dan menghapus kesalahan dan ketidaksesuaian yang ada.

3.4. Modelling

Penelitian ini menggunakan model algoritma klasterisasi *K-Means*. Berikut adalah langkah-langkah menggunakan algoritma *K-Means* untuk proses klastering:

3.4.1. Tahap inisialisasi yaitu menentukan jumlah klaster (K) berdasarkan kondisi data yang ada. Metode penentuan jumlah klaster yang hendak digunakan adalah metode *Elbow*. Dalam menentukan jumlah klaster yang optimal perlu menghitung nilai *Sum of Square Error* (SSE) dengan menggunakan rumus dalam persamaan (2) berikut [19]:

$$SSE = \Sigma_{K=1}^{K} \Sigma s S_k \|X_i - C_k\|^2 \quad (2)$$

Keterangan:
K = jumlah klaster; X_i = atribut dari data ke-i (i=1, 2, 3, 4, …, …, n);

 ISSN:2302-8149
e-ISSN:2540-9719

C_k = atribut titik pusat klaster ke-i (i=1, 2, 3, 4, …, n)

3.4.2. Tahap inisialisasi *centroid*/pusat klaster yaitu dengan memilih sejumlah data sebagai pusat klaster secara random. Nilai pusat klaster dapat ditentukan melalui persamaan (3) berikut:

$$\bar{x}_{\ell j} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{..i} \quad (3)$$

Keterangan:
$x_{\ell j}$ = nilai rata-rata variabel ℓ (ℓ=1,2,…,*p*) pada klaster ke-*j* (*j*=1,2,…,*k*);
n_j = jumlah objek pada klaster ke-*j;*
$x_{..i}$ = nilai objek ke-*i* (i=1,2,…,n_j) pada variabel ℓ dan klaster ke-*j*

3.4.3. Proses klastering yaitu dengan penghitungan jarak antar data dengan tiap pusat klaster. Rumus penghitungan jarak salah satunya akan menggunakan rumus *Euclidean Distance* pada persamaan (4) sebagai berikut [19]:

$$d(x,y) = |x-y|\sqrt{\Sigma_{i=1}^{n}(x_i - y_i)^2} \quad (4)$$

Keterangan:
i = indeks atribut; n = jumlah data; x_i = atribut dari data ke-i (i=1, 2, 3, 4, …, …, n);
y_i = atribut dari pusat klaster ke-i (i=1, 2, 3, 4, …, …, n)

3.5. Evaluation

Pada tahap ini, menurut Damayanti, dkk [20] dilakukan 4 tahapan, yaitu: (1) melakukan evaluasi model yang dipakai dalam tahap pemodelan guna mendapatkan kualitas serta efektivitas sebelum digunakan; (2) menetapkan model yang dapat memenuhi tujuan pada tahapan awal; (3) menetapkan apakah ada persoalan penting dari bisnis atau penelitan yang belum atau tidak terpecahkan dengan baik; (4) pengambilan keputusan yang berkaitan dengan penggunaan hasil *data mining*. Metode *Elbow* dapat digunakan dalam mengevaluasi metode klastering *K-Means*. Metode *Elbow* mencari nilai k yang menghasilkan SSE terendah, yang menunjukkan bahwa klaster-klaster yang terbentuk cukup kompak dan minim *overlapping*.

3.6. Deployment

Deployment pada CRISP-DM merupakan tahap terakhir dalam proses *data mining*. Tahap ini bertujuan untuk mengimplementasikan model *data mining* yang telah dikembangkan ke dalam sistem operasional. Tujuan dari *deployment* adalah agar model tersebut dapat digunakan secara efektif dan efisien oleh pengguna. Tahapan ini dapat berupa pembuatan aplikasi atau laporan [21].

4 Hasil dan Pembahasan

Tahapan *business understanding* merupakan tahapan/langkah pertama dalam mengidentifikasi masalah. Ada beberapa faktor yang memengaruhi kesuksesan lagu di Spotify, yaitu: 1) Karakteristik lagu: genre musik, tempo, struktur lagu, melodi, vokal, lirik, dan kualitas produksi; 2) Karakteristik pendengar: demografi (usia, jenis kelamin, lokasi), preferensi musik, dan kebiasaan mendengarkan musik; 3) Strategi pemasaran dan promosi: *platform streaming* musik, media sosial, *influencer*, dan konser; 4) Faktor eksternal: tren musik terkini, peristiwa budaya, dan kondisi ekonomi.

Tahapan *data understanding* merupakan tahapan/langkah kedua dalam mengidentifikasi masalah. Pada tahapan ini diketahui menggunakan *dataset* yang diunduh dari *website Kaggle* yang berisikan sejumlah 51.105 baris dengan 25 buah atribut dan tanggal *cut-off* data yaitu pada 31 Oktober 2023. Pada tahapan ini juga dilakukan juga penghitungan koefisien korelasi *Pearson* untuk mencari hubungan antara popularitas lagu terhadap *loudness, danceability*, dan juga energi. Setelah dilakukan penghitungan, maka didapatkan seperti yang tercantum pada Tabel 3 sebagai berikut:

Tabel 3. Koefisien Korelasi Pearson Popularitas Lagu

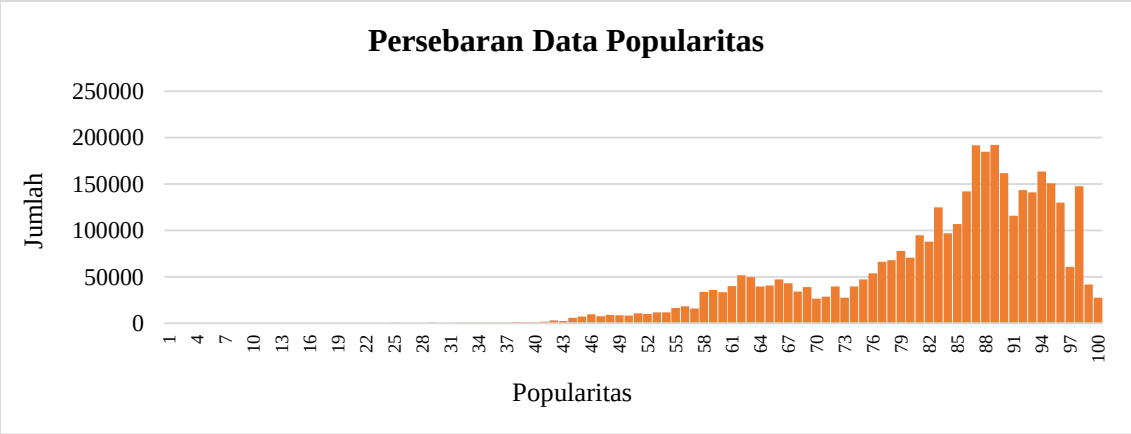
	Klaster	r_{XY}
1	Popularitas terhadap *loudness*	0,76

 ISSN:2302-8149
e-ISSN:2540-9719

Klaster		r_{XY}
2	Popularitas terhadap *danceability*	0,993
3	Popularitas terhadap energi	0,96

Nilai koefisien korelasi Pearson (r_{XY}) berdasarkan hasil Tabel 3 di atas menunjukan bahwa popularitas lagu terhadap *loudness* memiliki hubungan positif yang kuat. Sementara popularitas lagu terhadap *danceability* dan energi masing-masing memiliki hubungan positif yang sangat kuat. Hal ini dapat berarti bahwa 1) Lagu dengan musik yang keras dan energik (*loudness* tinggi) cenderung lebih populer dibandingkan lagu yang tenang dan lembut (*loudness* rendah); 2) Lagu dengan ritme yang *upbeat* dan mudah untuk ditarikan (*danceability* tinggi) cenderung lebih populer dibandingkan lagu dengan ritme yang lambat dan monoton (*danceability* rendah); 3) Lagu dengan musik yang penuh semangat dan dinamis (energi tinggi) cenderung lebih populer dibandingkan lagu yang datar dan kurang bergairah (energi rendah).

Tahapan *data preparation* merupakan tahapan/langkah ketiga dalam mengidentifikasi masalah. Pertama-tama dilakukan beberapa metode pembersihan data dasar seperti menghapus entri duplikat atau baris kosong sebelum memulai analisis utama. Selanjutnya, data dideskripsikan berdasarkan popularitas seperti pada Gambar 2 sebagai berikut:



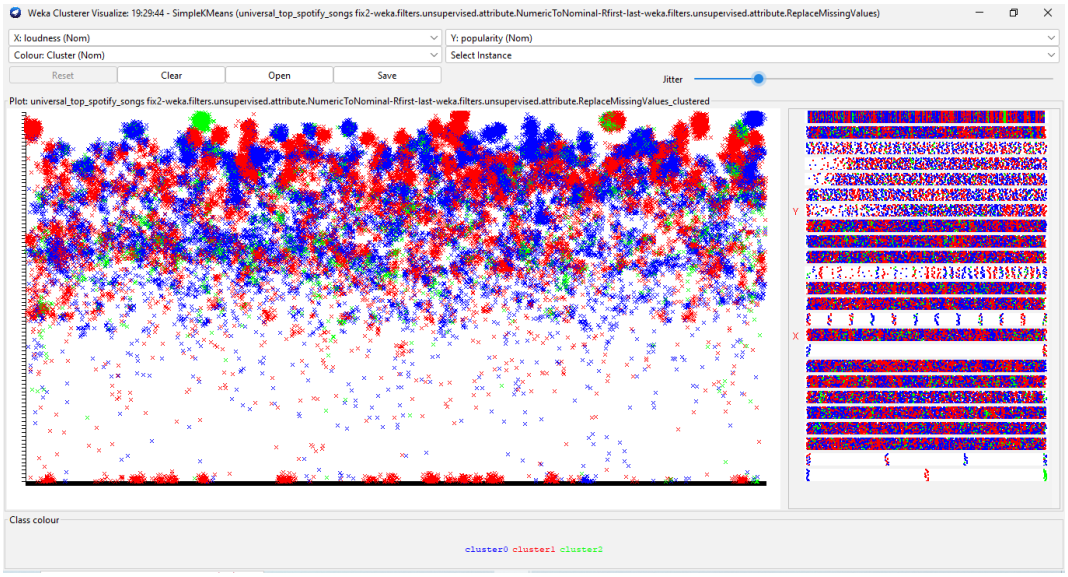
Gambar 2. Penyebaran Data Popularitas.

Berdasarkan hasil olah data, lagu-lagu yang dirilis pada tahun 2023 memiliki tingkat kepopuleran yang tinggi di *Spotify*. Hal ini dapat disebabkan oleh beberapa faktor, seperti promosi yang lebih gencar, ketersediaan lagu yang lebih luas, dan ketertarikan pendengar terhadap hal-hal baru, hal ini bisa dilihat pada Gambar 3:

Gambar 3. Jumlah Streaming Lagu terhadap Tahun Perilisan Album di Spotify.

 ISSN:2302-8149
e-ISSN:2540-9719

Selanjutnya, pada tahapan *modelling* yang merupakan tahapan keempat dari kerangka kerja CRISP-DM, kriteria jenis data diintegrasikan berdasarkan kolom *file* csv ke dalam satu *file*. Penelitian ini menggunakan metode *K-Means* dengan tujuan klastering data ke dalam beberapa kriteria/karakteristik sesuai dengan popularitas, yang selanjutnya akan dilakukan analisa dengan data *loudness*, *danceability*, dan energi. Pada proses ini diwujudkan untuk meminimalisir variasi antar data yang ada di dalam suatu klaster dan mengembangkan variasi data yang ada di klaster lainnya. Popularitas *Global Top 10* ditunjukan seperti pada Gambar 4 berikut:



Gambar 4. Persebaran Data Popularitas terhadap *Loudness*.

Gambar 4 menunjukan hasil analisis klusterisasi *K-Means* terhadap data lagu-lagu dalam aplikasi *Spotify*. Analisis ini menggunakan dua atribut, yaitu *loudness* dan *popularity*. Terdapat tiga kelompok/klaster lagu yang dapat diidentifikasi, yaitu:

a. Kelompok 1 (merah): lagu-lagu yang memiliki *loudness* yang tinggi dan popularitas yang tinggi.
b. Kelompok 2 (biru): Lagu-lagu yang memiliki *loudness* yang sedang dan popularitas yang sedang.
c. Kelompok 3 (hijau): Lagu-lagu yang memiliki *loudness* yang rendah dan popularitas yang rendah.

Tabel 4. Analisis Klasterisasi Popularitas terhadap *Loudness*.

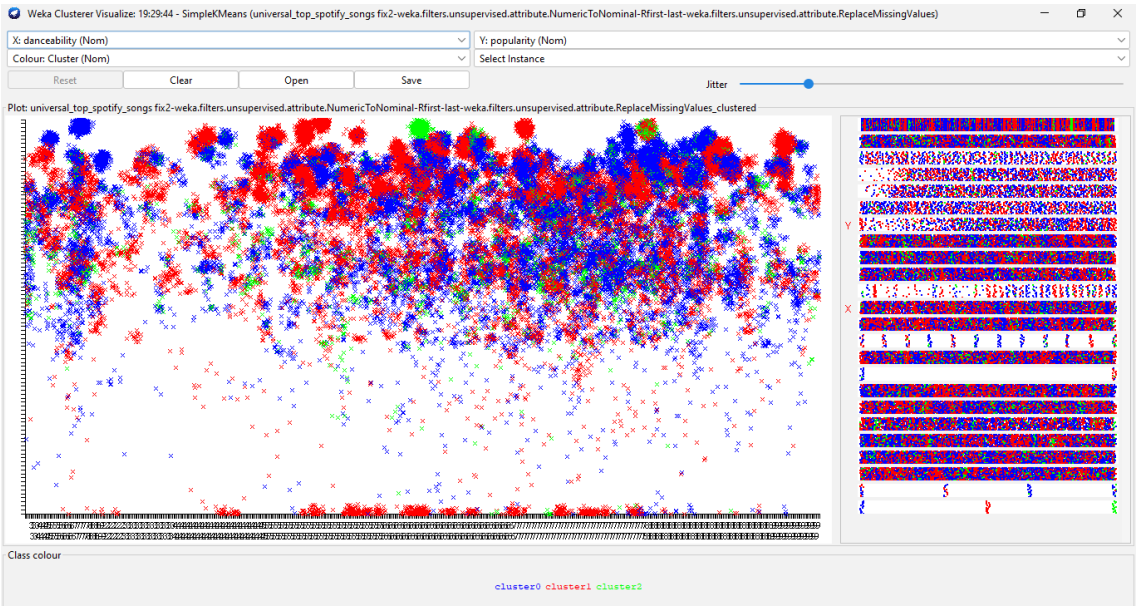
Klaster	Nilai SSE	Nilai Euclidean distance	Nilai pusat klaster
1	1,09	0,87	(12,5, 60)
2	0,86	0,69	(50,40)
3	1,22	0,96	(25,20)

Pada Tabel 4, nilai SSE adalah jumlah kuadrat jarak antara titik-titik dalam suatu klaster dengan pusat klaster tersebut. Semakin kecil nilai SSE, maka semakin baik titik-titik dalam klaster tersebut terpusat pada pusat klaster. Sedangkan, nilai *euclidean distance* adalah jarak antara dua titik dalam ruang *Euclidean*. Semakin kecil nilai *euclidean distance*, maka semakin dekat dua titik tersebut. Berdasarkan nilai SSE dan *euclidean distance*, dapat disimpulkan bahwa klaster 1 dan klaster 2 memiliki titik-titik yang lebih terpusat daripada klaster 3. Hal ini menunjukkan bahwa klaster 1 dan klaster 2 memiliki hubungan popularitas terhadap *loudness* yang lebih jelas daripada klaster 3.

Berdasarkan analisis klusterisasi K-Means terhadap data lagu-lagu dalam aplikasi Spotify, dapat disimpulkan bahwa musik populer di Spotify saat ini semakin bervariasi dalam hal *loudness*. Lagu-lagu dengan *loudness* yang tinggi seperti pop, *rock*, dan elektronik kemudian lagu-lagu dengan *loudness* sedang seperti *hip-hop*, *r&b*, dan *folk* semakin populer, sedangkan lagu-lagu dengan *loudness* yang rendah seperti *jazz*, *blues*, klasik semakin kurang populer. Tren ini dapat disebabkan oleh beberapa faktor, seperti perubahan selera masyarakat, perkembangan teknologi musik, serta

 ISSN:2302-8149
e-ISSN:2540-9719

persaingan antar artis dan genre musik. Tren-tren tersebut dapat berubah seiring perkembangan jaman. Namun, tren-tren tersebut memberikan gambaran umum mengenai musik populer di *Spotify* saat ini.



Gambar 5. Persebaran Data Popularitas Terhadap *Danceability*.

Gambar 5 merupakan hasil analisis klasterisasi K-Means terhadap data lagu-lagu dalam aplikasi Spotify yang menggunakan dua atribut yaitu, popularitas dan *danceability*. Hal ini menunjukkan bahwa terdapat tiga kelompok/klaster lagu yang dapat diidentifikasi, yaitu:

a. Kelompok 1 (merah): Lagu-lagu yang memiliki popularitas yang tinggi dan *danceability* yang tinggi.
b. Kelompok 2 (biru): Lagu-lagu yang memiliki popularitas yang sedang dan *danceability* yang sedang.
c. Kelompok 3 (hijau): Lagu-lagu yang memiliki popularitas yang rendah dan *danceability* yang rendah.

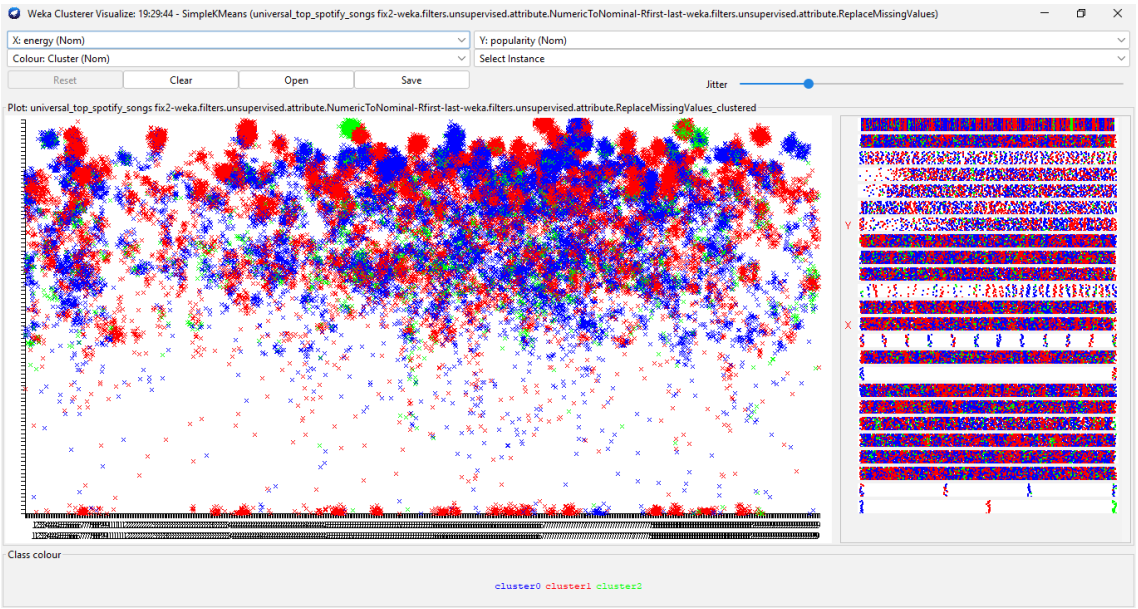
Terdapat 3 klaster pada hubungan popularitas terhadap *danceability* seperti yang ditunjukan pada Tabel 5. Berdasarkan nilai SSE dan *euclidean distance*, dapat disimpulkan bahwa ketiga klaster memiliki titik-titik yang terpusat dengan baik. Hal ini menunjukkan bahwa ketiga klaster memiliki hubungan popularitas terhadap *danceability* yang cukup jelas.

Tabel 5. Analisis Klasterisasi Popularitas terhadap *Danceability*.

Klaster	Nilai SSE	Nilai Euclidean distance	Nilai pusat klaster
1	0,99	0,99	(40,70)
2	0,95	0,97	(60,50)
3	1,03	1,01	(20,30)

Berdasarkan analisis klasterisasi K-Means terhadap data lagu-lagu dalam aplikasi Spotify yang menggunakan dua atribut yaitu, popularitas dan *danceability*, dapat disimpulkan bahwa musik populer di Spotify saat ini semakin berfokus pada *danceability*. Lagu-lagu dengan *danceability* yang tinggi semakin populer, sedangkan lagu-lagu dengan *danceability* yang rendah semakin kurang populer. Tren ini dapat disebabkan oleh beberapa faktor, seperti perubahan selera masyarakat, perkembangan teknologi musik, serta persaingan antar artis.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 6. Persebaran Data Popularitas Terhadap Energi.

Gambar 6 merupakan hasil analisis klasterisasi *K-Means* terhadap data lagu-lagu dalam aplikasi Spotify yang menggunakan dua atribut, yaitu popularitas dan energi. Hal ini menunjukkan bahwa terdapat tiga kelompok/klaster lagu yang dapat diidentifikasi, yaitu:

a. Kelompok 1 (merah): Lagu-lagu yang memiliki popularitas yang tinggi dan energi yang tinggi.
b. Kelompok 2 (biru): Lagu-lagu yang memiliki popularitas yang sedang dan energi yang sedang.
c. Kelompok 3 (hijau): Lagu-lagu yang memiliki popularitas yang rendah dan energi yang rendah.

Terdapat 3 klaster pada hubungan popularitas terhadap energi. Berdasarkan nilai SSE dan *euclidean distance* pada Tabel 6, dapat disimpulkan bahwa ketiga klaster memiliki titik-titik yang terpusat dengan sangat baik. Hal ini menunjukkan bahwa ketiga klaster memiliki hubungan popularitas terhadap energi yang sangat jelas.

Tabel 6. Analisis Klasterisasi Popularitas terhadap Energi.

Klaster	Nilai SSE	Nilai Euclidean distance	Nilai pusat klaster
1	0,99	0,99	(70,60)
2	1,02	1,01	(50,40)
3	1,05	1,03	(30,20)

Berdasarkan analisis klasterisasi *K-Means* terhadap data lagu-lagu dalam aplikasi *Spotify*, dapat disimpulkan bahwa musik populer di *Spotify* saat ini semakin berfokus pada energi. Lagu-lagu dengan energi yang tinggi semakin populer, sedangkan lagu-lagu dengan energi yang rendah semakin kurang populer. Tren ini dapat disebabkan oleh beberapa faktor, seperti 1) perubahan selera masyarakat yang semakin menyukai musik yang lebih dinamis, menyenangkan, dan membangkitkan semangat; 2) perkembangan teknologi musik yang memungkinkan produser musik untuk menciptakan lagu-lagu dengan energi yang lebih tinggi; 3) persaingan antar genre musik yang mendorong para musisi untuk menciptakan lagu-lagu yang lebih unik dan menarik perhatian pendengar.

Berdasarkan hasil pengolahan data, berikut adalah hasil analisis data untuk lagu yang berada dalam peringkat Top 10 di aplikasi *Spotify* yang diurutkan berdasarkan artis yang paling populer seperti pada Tabel 7.

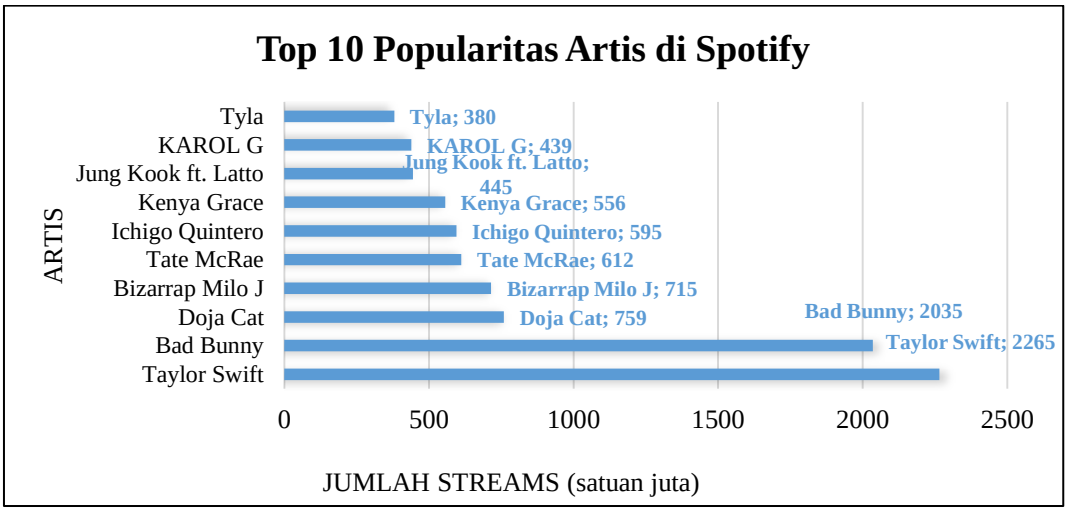
Tabel 7. Persentase Popularitas Artis dalam Peringkat Top 10 Spotify.

Artis	Popularitas
Taylor Swift	26%
Bad Bunny	23%

 ISSN:2302-8149
e-ISSN:2540-9719

Artis	Popularitas
Doja Cat	9%
Bizarrap Milo J	8%
Tate McRae	7%
Ichigo Quintero	7%
Kenya Grace	6%
Jung Kook ft. Latto	5%
KAROL G	5%
Tyla	4%

Berdasarkan Tabel 7, *Taylor Swift* menempati urutan pertama sebagai penyanyi paling populer dengan tingkat popularitas sebesar 26% dan *Bad Bunny* berada pada urutan kedua sebagai penyanyi kedua populer dengan tingkat popularitas sebesar 23%. Selisih tingkat persentase keduanya sebesar 3%. Hal ini menunjukan bahwa kedua penyanyi tersebut memiliki tingkat popularitas yang paling tinggi saat ini. Hal tersebut dapat disebabkan karena banyaknya masyarakat, baik itu fans ataupun *casual listeners* yang menyukai dan sering memutar lagu-lagu karya kedua penyanyi tersebut di *Spotify*. Melalui gambar 7 berikut dapat disimpulkan bahwa jumlah angka *streaming*/pemutaran lagu di *Spotify* terbanyak dimiliki oleh *Taylor Swift* dan *Bad Bunny*. Berdasarkan data peringkat mingguan pada *website Billboard*, *Taylor Swift* menempati peringkat pertama di dalam *Chart Billboard HOT 100 Top* 10 yang diambil pada tanggal 4 November 2023 sementara artis-artis lainnya tersebar di chart *Billboard* HOT 100, *Billboard* 200, dan *Billboard* 200 *Global* [22].



Gambar 7. Peringkat Top 10 Popularitas Artis di *Spotify*.

Tahapan kelima dalam kerangka kerja *CRISP-DM* yaitu evaluasi, maka diperoleh hasil seperti pada Tabel 8, Tabel 9, dan Tabel 10 sebagai berikut:

Tabel 8. Evaluasi Kualitas Klastering K-Means berdasarkan Nilai SSE Popularitas Lagu terhadap *Loudness*.

Klaster	Kompaktitas Klaster	Visualisasi Klaster
1	Memiliki kekompakan yang sedang karena nilai SSE-nya 1,09 berada di antara klaster 2 dan 3	Memiliki bentuk yang berada di antara klaster 2 dan 3
2	Klaster yang paling kompak karena memiliki nilai SSE terkecil, yaitu 0,86	Memiliki bentuk yang paling kompak dan terpusat
3	Klaster yang paling tidak kompak karena memiliki nilai SSE terbesar, yaitu 1,22	Memiliki bentuk yang paling menyebar

Tabel 9. Evaluasi Kualitas Klastering K-Means berdasarkan Nilai SSE Popularitas Lagu terhadap *Danceability*.

 ISSN:2302-8149
e-ISSN:2540-9719

Klaster	Kompaktitas Klaster	Visualisasi Klaster
1	Klaster yang paling tidak kompak karena memiliki nilai SSE terbesar, yaitu 1,02	Memiliki bentuk yang paling menyebar
2	Klaster yang paling kompak karena memiliki nilai SSE terkecil, yaitu 0,95	Memiliki bentuk yang paling kompak dan terpusat
3	Klaster yang memiliki kekompakan yang sedang karena nilai SSE-nya berada di antara klaster 2 dan 1	Memiliki bentuk yang berada di antara klaster 1 dan 2

Tabel 9. Evaluasi Kualitas Klastering K-Means berdasarkan Nilai SSE Popularitas Lagu terhadap Energi.

Klaster	Kompaktitas Klaster	Visualisasi Klaster
1	Klaster yang paling kompak karena memiliki SSE terkecil, yaitu 0,99	Memiliki bentuk yang paling kompak dan terpusat
2	Klaster yang memiliki kekompakan sedang karena nilai SSE-nya berada di antara klaster 1 dan 3	Memiliki bentuk yang paling menyebar
3	Klaster yang paling tidak kompak karena nilainya SSE nya paling besar yaitu 1,05	Memiliki bentuk yang berada di antara klaster 1 dan 2

Kemudian untuk tahapan terakhir yaitu *deployment* di mana dilakukan pembuatan laporan, laporan dapat dibuat apabila evaluasi model klastering telah selesai dilakukan [20].

5 Kesimpulan

Penelitian dalam bentuk klasterisasi menggunakan *dataset* lagu terpopuler yang diambil dari *website Kaggle* yang dianalisis menggunakan metode *K-Means* adalah untuk mengetahui kesuksesan lagu dalam bentuk lagu yang berada dalam 10 besar pada *Spotify*. Lagu populer di *Spotify* saat ini semakin berfokus pada *loudness*, *danceability*, dan energi dengan tren yang menonjol, yaitu lagu-lagu dengan *loudness*, *danceability*, dan energi yang tinggi semakin populer, sementara lagu-lagu dengan *loudness*, *danceability*, dan energi yang rendah semakin kurang populer. Tren-tren ini dapat disebabkan oleh beberapa faktor, seperti perubahan selera masyarakat, perkembangan teknologi musik, serta persaingan antar artis dan genre musik. Penelitian ini masih memiliki beberapa kekurangan. Maka dari itu, saran untuk penelitian ke depannya. Pertama, memperluas cakupan data yang digunakan. Data yang digunakan dalam penelitian ini hanya berasal dari aplikasi *Spotify*. Untuk mendapatkan gambaran yang lebih lengkap mengenai tren musik populer, perlu dilakukan analisis terhadap data dari berbagai sumber, seperti layanan streaming musik lainnya seperti *Apple Music, Youtube Music, Youtube, Pandora*, dll. Kedua, memperluas jangkauan atribut. Selain *loudness, danceability*, dan energi, terdapat beberapa atribut lain yang dapat digunakan untuk menganalisis tren musik populer, seperti tempo, mode, dan *genre*. Dengan memperluas jangkauan atribut yang digunakan, penelitian ini dapat memberikan gambaran yang lebih lengkap mengenai tren musik populer. Ketiga, untuk mendapatkan pemahaman yang lebih mendalam, perlu dilakukan analisis lebih lanjut terhadap data yang ada. Analisis lebih lanjut dapat dilakukan dengan menggunakan metode statistik yang lebih kompleks, seperti analisis regresi atau analisis faktor. Keempat, dapat menggunakan algoritma klastering lain seperti *K-Means* ++ atau *Hierarchical Clustering*, mengubah jumlah klaster menjadi 2 atau 4, dan mengubah nilai pusat klaster awal secara manual.

Referensi

[1] H. Martopo, "Sejarah Musik Sebagai Sumber Pengetahuan Ilmiah Untuk Belajar Teori, Komposisi, Dan Praktik Musik," *Harmonia: Journal of Arts Research and Education*, vol. 13, no. 2, 2013.

[2] I. Ruddin, H. Santoso, and R. E. Indrajit, "Digitalisasi Musik Industri: Bagaimana Teknologi Informasi Mempengaruhi Industri Musik di Indonesia," *Jurnal Pendidikan Sains dan Komputer*, vol. 2, no. 01, 2022, doi: 10.47709/jpsk.v2i01.1395.

[3] Billboard, "Billboard Charts." Accessed: Nov. 15, 2023. [Online]. Available: https://www.billboard.com/charts/

 ISSN:2302-8149
e-ISSN:2540-9719

[4] U. L. Musyarofah, S. N. Alima, and D. S. Y. Kartika, "KLASIFIKASI TOP 50 SPOTIFY TAHUN 2010-2019 MENGGUNAKAN METODE K-MEANS CLUSTERING," *Prosiding Seminar Nasional Teknologi dan Sistem Informasi*, vol. 2, no. 1, 2022, doi: 10.33005/sitasi.v2i1.300.
[5] S. Navisa, Luqman Hakim, and Aulia Nabilah, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM," *Jurnal Sistem Cerdas*, vol. 4, no. 2, 2021, doi: 10.37396/jsc.v4i2.162.
[6] M. Interiano, K. Kazemi, L. Wang, J. Yang, Z. Yu, and N. L. Komarova, "Musical trends and predictability of success in contemporary songs in and out of the top charts," *R Soc Open Sci*, vol. 5, no. 5, 2018, doi: 10.1098/rsos.171274.
[7] S. Y. M. Netti and I. Irwansyah, "Spotify: Aplikasi Music Streaming untuk Generasi Milenial," *Jurnal Komunikasi*, vol. 10, no. 1, 2018, doi: 10.24912/jk.v10i1.1102.
[8] Spotify, "Apa itu Spotify?," support.spotify.com. Accessed: Nov. 15, 2023. [Online]. Available: https://support.spotify.com/id-id/article/what-is-spotify/
[9] SAS Institute, "Data Mining: what it is & why it matters," SAS Insights: Analytics and Data Science Insights.
[10] S. Agarwal, "Data mining: Data mining concepts and techniques," in *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*, 2014. doi: 10.1109/ICMIRA.2013.45.
[11] P. Chapman *et al.*, "CRISP-DM -Cross-Industry Standard Process for Data Mining- 1.0 Step-by-step data mining guide.," *CRISP-DM Consortium*, 2000.
[12] A. M. Ikotun, M. S. Almutari, and A. E. Ezugwu, "K-means-based nature-inspired metaheuristic algorithms for automatic data clustering problems: Recent advances and future directions," *Applied Sciences (Switzerland)*, vol. 11, no. 23. 2021. doi: 10.3390/app112311246.
[13] C. Tan, H. Zhao, and H. Ding, "Statistical initialization of intrinsic K-means clustering on homogeneous manifolds," *Applied Intelligence*, vol. 53, no. 5, 2023, doi: 10.1007/s10489-022-03698-8.
[14] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, 2019, doi: 10.34148/teknika.v8i1.95.
[15] L. Medeiros, "The CRISP-DM methodology," Medium.
[16] A. Asaniczka, "Top Spotify Songs in 73 Countries (Daily Updated)," Kaggle.
[17] A. Rahmawati and E. Setyowati, "K-Means Cluster Analysis for District or City Clustering in Bengkulu Province based on The Number of Base Transceiver Stations and The Strength of Cell Phone Signal," *CESS (Journal of Computer Engineering, System and Science)*, vol. 8, no. 1, 2023, doi: 10.24114/cess.v8i1.40913.
[18] F. Ridzuan and W. M. N. Wan Zainon, "A review on data cleansing methods for big data," in *Procedia Computer Science*, 2019. doi: 10.1016/j.procs.2019.11.177.
[19] M. S. Pangestu and M. A. Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means," *Sainteks*, vol. 19, no. 2, 2022, doi: 10.30595/sainteks.v19i2.14495.
[20] S. E. Damayanti and S. K. Kuswayati, "Analisis Dan Implementasi Framework CRISP-DM (Cross Industry Standard Process For Data Mining) Untuk Clustering Perguruan Tinggi Swasta," *ejournal sttbandung*, 2018.
[21] A. Pambudi, "PENERAPAN CRISP-DM MENGGUNAKAN MLR K-FOLD PADA DATA SAHAM PT. TELKOM INDONESIA (PERSERO) TBK (TLKM) (STUDI KASUS: BURSA EFEK INDONESIA TAHUN 2015-2022)," *Jurnal Data Mining dan Sistem Informasi*, vol. 4, no. 1, p. 1, Mar. 2023, doi: 10.33365/jdmsi.v4i1.2462.
[22] Billboard, "Billboard Hot 100™," billboard.