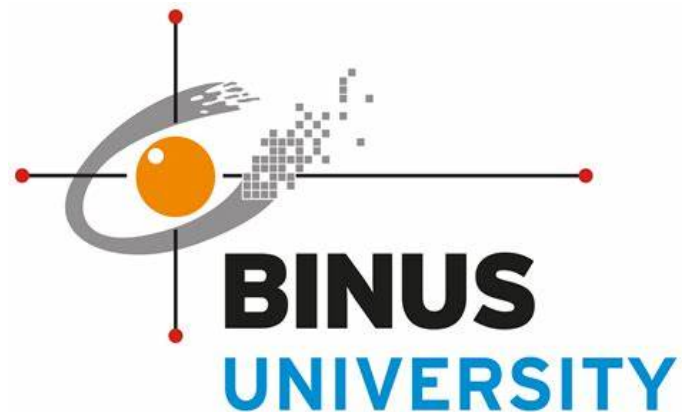


## **JUDUL MAKALAH**

**Disusun guna memenuhi tugas mata kuliah**

Dosen Pengampu:

**FEPRI PUTRA PANGHURIAN, S.Kom, M.T.I.**



### **Kelompok 1**

Albertus Christian	2702255703
Archi Setio	2702255962
Dominikus Sebastian Ramli	2702329664
Ignatius Abraham Aristio Kusnadi	2702243590
Vincent Virgo	2702250381
Wilbert Bernardi	2702255786

### **KELAS LD01**

**PROGRAM STUDI COMPUTER SCIENCE**

**BINA NUSANTARA UNIVERSITY**

**TAHUN 2024/2025**

## DAFTAR ISI

	Hlm
<b>COVER</b> .....	1
<b>DAFTAR ISI</b> .....	2
<b>BAB I: PENDAHULUAN</b>	
1.1 Latar Belakang .....	3
1.2 Rumusan Masalah .....	4
1.4 Tujuan Penulisan .....	4
<b>BAB II: Metodologi</b>	
2.1 Materi Pertama.....	
2.1.1 Sub-Materi Pertama.....	
2.1.2 Sub-Materi Kedua.....	
2.2 Materi Kedua.....	
2.3 Materi Ketiga.....	
<b>BAB III: Hasil dan Pembahasan</b>	
3.1 Pengambilan Dataset.....	
3.2 Data Cleaning.....	
3.3 Distribusi Data dan Boxplot.....	
3.4 Visualisasi Popularitas Lagu dan Artis.....	
3.5 Clustering Lagu dengan K-Means.....	
3.6 Analisis Popularitas setiap Cluster.....	
<b>BAB IV: Kesimpulan dan Saran</b> .....	
<b>DAFTAR PUSTAKA</b> .....	
<b>DAFTAR TABEL (Jika Ada)</b> .....	iv
<b>DAFTAR GAMBAR (Jika Ada)</b> .....	v
<b>DAFTAR LAMPIRAN (Jika Ada)</b> .....	vi

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Dalam era digital saat ini, platform streaming musik telah menjadi gerbang utama bagi jutaan orang untuk menemukan, mengonsumsi, dan berinteraksi dengan musik. Spotify, Apple Music, dan Shazam adalah tiga pemain kunci dalam ekosistem musik digital ini, masing-masing dengan jutaan pengguna aktif dan koleksi lagu yang sangat luas. Ini menjadikan gabungan data dari platform-platform tersebut sebagai sumber yang sangat kaya dan representatif untuk melakukan analisis mendalam mengenai tren musik, preferensi pendengar, dan dinamika industri musik secara keseluruhan.

Dataset yang kami pilih sebagai fokus utama dalam makalah ini adalah kompilasi data yang mencakup metrik dari ketiga platform terkemuka tersebut. Informasi yang tersedia meliputi atribut audio seperti tempo, nada, energi, dan danceability, serta data deskriptif seperti popularitas lagu, genre (jika tersedia atau dapat disimpulkan), dan tanggal rilis. Selain itu, data ini juga menyediakan metrik penting terkait keterlibatan lagu di berbagai playlist dan chart di Spotify, Apple Music, dan Deezer, serta data dari Shazam mengenai chart. Dengan demikian, dataset ini memungkinkan kami untuk melakukan analisis multidimensional yang komprehensif, menjelajahi hubungan antara berbagai faktor dan mendapatkan wawasan yang mendalam tentang ekosistem musik digital yang kompleks.

Popularitas dan relevansi gabungan platform ini dalam budaya populer saat ini menjamin bahwa temuan dari analisis dataset ini akan memiliki signifikansi praktis dan teoretis yang tinggi. Hasil penelitian ini dapat memberikan wawasan berharga bagi berbagai pemangku kepentingan, termasuk musisi, produser, label rekaman, agregator, dan pengembang aplikasi musik. Lebih lanjut, analisis dataset ini juga dapat berkontribusi pada pemahaman yang lebih luas tentang bagaimana teknologi digital telah mengubah cara orang menemukan, mengonsumsi, dan berinteraksi dengan musik di skala global.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka, ada beberapa rumusan masalah yang ditentukan yaitu, sebagai berikut :

- a. Bagaimana lagu-lagu di berbagai *streaming platform* dapat dikelompokkan (di-cluster) menjadi segmen-segmen yang berbeda berdasarkan kombinasi atribut-atribut audionya?
- b. Bagaimana atribut audio (misalnya, BPM, energi, *danceability*) berkorelasi dengan popularitas lagu di berbagai *streaming platform*?
- c. Bagaimana keterkaitan antara clustering tipe lagu dengan banyaknya stream di *streaming platform*?
- d. Apakah terdapat perbedaan signifikan dalam rata-rata jumlah streams (sebagai indikator popularitas) antar segmen musik (cluster) yang telah diidentifikasi?

## 1.3 Tujuan Penulisan

Tujuan penulisan berisi pernyataan-pernyataan penting yang berisi jawaban dari rumusan masalah. Tujuan penulisan dituliskan dengan poin-poin sebagai berikut:

- a. Menganalisis korelasi antara atribut audio (seperti BPM, energi, dan danceability) dengan tingkat popularitas lagu di platform Spotify.
- b. Mengidentifikasi bagaimana jumlah kolaborasi antar artis dan jaringan keterkaitan mereka berkontribusi pada jangkauan dan popularitas lagu di Spotify.
- c. Mengelompokkan (melakukan clustering) lagu-lagu di Spotify ke dalam segmen-segmen musik yang berbeda berdasarkan kombinasi atribut-atribut audionya.
- d. Menganalisis dan membandingkan rata-rata jumlah streams (sebagai indikator popularitas) di antara segmen-segmen musik (cluster) yang berbeda.

## **BAB II**

### **Metodologi**

Penelitian ini menggunakan pendekatan kuantitatif berbasis data yang diperoleh dari Spotify, dengan tujuan mengelompokkan lagu populer berdasarkan karakteristik audio menggunakan metode K-Means Clustering. Proses ini dilaksanakan melalui beberapa tahapan penting mulai dari pengumpulan dan eksplorasi data hingga analisis lanjutan dan visualisasi hasil.

#### **2.1 Dataset dan Tools**

##### **2.1.1 Pengumpulan Data**

Dataset yang digunakan berjudul *Popular Spotify Songs*, berupa file .csv yang mencatat berbagai informasi penting tentang lagu-lagu populer di berbagai platform

streaming musik seperti Spotify, Apple Music, dan Deezer. Data ini mencakup lebih dari 20 atribut seperti:

- **track\_name** = Judul Lagu
- **artist(s)\_name** = Nama penyanyi atau grup
- **released\_year, released\_month, released\_day**: Informasi tanggal rilis
- **streams**: Jumlah pemutaran lagu
- **bpm**: Tempo dalam Beats Per Minute
- **danceability\_%, energy\_%, valence\_%, acousticness\_%, instrumentalness\_%, liveness\_%, speechiness\_%**: Fitur audio dalam bentuk persentase

Dataset dimuat menggunakan Python melalui library **pandas** dengan encoding **'latin1'** untuk menangani karakter khusus pada nama artis dan lagu.

### 2.1.2 Tools dan Lingkungan Analisis

Proses analisis dilakukan menggunakan *Google Colab* sebagai lingkungan pemrograman berbasis cloud. Adapun tools/libraries yang digunakan meliputi:

- **pandas**: Untuk manipulasi data tabular
- **matplotlib** dan **seaborn**: untuk visualisasi
- **sklearn**: untuk implementasi K-Means dan PCA
- **numpy**: Untuk operasi numerik
- **StandardScaler**: Untuk normalisasi fitur numerik

## 2.2 Pra-pemrosesan dan Eksplorasi Data

### 2.2.1 Pembersihan Data

Langkah awal adalah mengecek jumlah data kosong dan duplikat menggunakan **df.isnull().sum()** dan **df.duplicated().sum()**. Kolom yang tidak relevan seperti **'key'**

dihapus untuk menyederhanakan analisis. Nilai dalam kolom **streams**. diubah menjadi numerik dengan **pd.to\_numeric**.

## 2.22 Eksplorasi Data (EDA)

Eksplorasi data dilakukan dengan visualisasi *box plot* untuk memahami distribusi dan potensi outlier dalam fitur numerik. Fitur-fitur seperti:

- bpm, danceability\_%, energy\_%, valence\_%
- acousticness\_%, instrumentalness\_%, liveness\_%, speechiness\_%

dianalisis secara individual. Selain itu, grafik batang horizontal (**barh**) digunakan untuk membandingkan lagu-lagu teratas di Spotify dan Apple Music berdasarkan jumlah playlist.

Visualisasi garis (line chart) juga digunakan untuk menampilkan total stream tahunan dari 5 artis dengan jumlah stream tertinggi, sehingga tren musik mereka dapat diamati secara kronologis.

## 2.3 K-Means Clustering untuk Segmentasi Lagu

### 2.3.1 Seleksi Fitur dan Normalisasi

Tujuh fitur audio utama dipilih sebagai variabel input:

- valence\_%, danceability\_%, energy\_%
- acousticness\_%, instrumentalness\_%
- liveness\_%, speechiness\_%

Fitur-fitur ini dinormalisasi menggunakan **StandardScaler** agar semua berada dalam skala yang seragam, mencegah fitur dengan skala besar mendominasi hasil clustering.

### 2.3.2 Penentuan Jumlah Klaster (K)

Jumlah klaster optimal ditentukan menggunakan *Elbow Method*. Dengan mengukur *Within-Cluster Sum of Squares* (WCSS) dari K=1 hingga K=10, titik siku pada grafik menunjukkan jumlah cluster terbaik. Hasil menunjukkan nilai optimal K=6.

### 2.3.3 Implementasi K-Means

Dengan K=6, algoritma K-Means diterapkan pada data terstandarisasi untuk mengelompokkan lagu-lagu berdasarkan kemiripan fitur. Hasil pengelompokan disimpan dalam kolom baru **Music\_Cluster**.

### 2.3.4 Visualisasi Hasil Clustering

Visualisasi dilakukan menggunakan dua metode:

- **PCA (Principal Component Analysis):** Mengurangi dimensi data ke 2 komponen utama untuk divisualisasikan dalam scatter plot berwarna berdasarkan cluster.
- **Pair Plot:** Visualisasi hubungan antar fitur numerik dengan warna berdasarkan cluster untuk melihat pemisahan antar grup.

## 2.4 Interpretasi Klaster

Setelah klaster terbentuk, dilakukan analisis lebih lanjut untuk menafsirkan karakteristik masing-masing klaster:



- Rata-rata nilai setiap fitur dihitung untuk tiap klaster.
- Rata-rata jumlah stream juga dihitung untuk menilai popularitas relatif antar klaster.
- Nama deskriptif diberikan pada setiap klaster berdasarkan karakteristik dominan, seperti:
  - Cluster 0: Upbeat & Joyful Dance Hits
  - Cluster 1: High-Energy Live Performances
  - Cluster 2: Intense & Brooding Rhythmic Tracks
  - Cluster 3: Rhythmic & Speech-Driven Grooves
  - Cluster 4: Atmospheric & Melancholic Instrumentals
  - Cluster 5: Mellow & Reflective Acoustic Pieces

Terakhir, untuk setiap klaster ditampilkan 5 lagu teratas berdasarkan jumlah stream sebagai contoh representatif.

## BAB III

### Hasil dan Pembahasan

#### 3.1 Pengambilan Dataset

Dataset yang digunakan merupakan kumpulan lagu populer Spotify dari tahun-tahun terakhir, dengan fitur-fitur seperti nama lagu, nama artis, jumlah artis, tanggal rilis, serta metrik performa seperti jumlah streams, keberadaan di playlist Spotify, Apple Music, dan platform lainnya.

	track_name	artist(s)_name	artist_count	released_year	released_month	released_day	in_spotify_playlists	in_spotify_charts	streams	in_apple_playlists	...	bpm	key	mode	danceability_%	valence_%	energy_%	acousticness_%	instrumentalness_%	liveness_%	speechiness_%	Music_cluster
0	Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	2	2023	7	14	553	147	141381703	43	...	125	B	Major	80	89	83	31	0	8	4	4
1	LALA	Myke Towers	1	2023	3	23	1474	48	133716286	48	...	92	C#	Major	71	61	74	7	0	10	4	4
2	vampire	Olivia Rodrigo	1	2023	6	30	1397	113	140003974	94	...	138	F	Major	51	32	53	17	0	31	6	2
3	Cruel Summer	Taylor Swift	1	2019	8	23	7858	100	800840817	116	...	170	A	Major	55	58	72	11	0	11	15	2
4	WHERE SHE GOES	Bad Bunny	1	2023	5	18	3133	50	303236322	84	...	144	A	Minor	65	23	80	14	63	11	6	4

#### 3.2 Data Cleaning

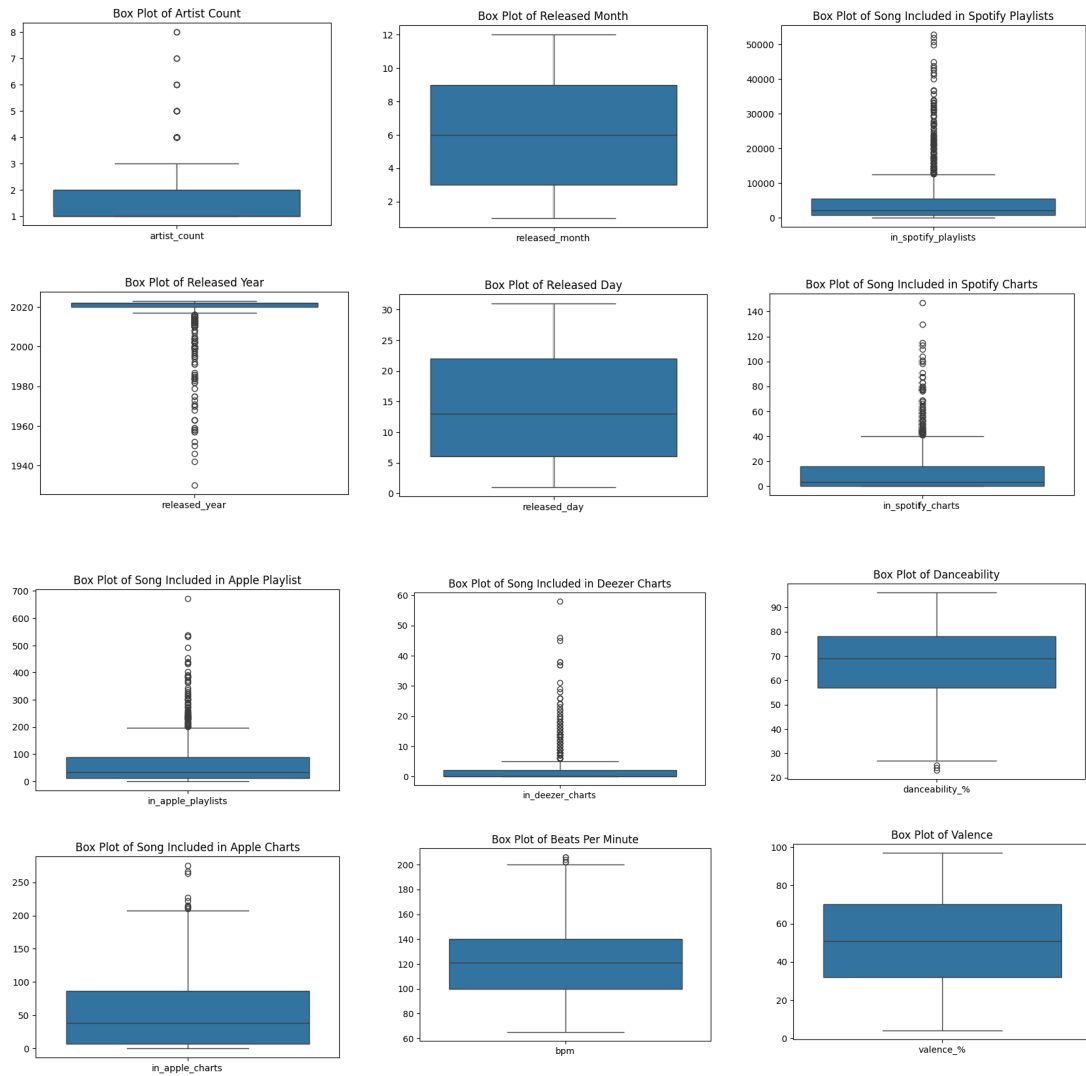
Pembersihan dataset dilakukan dengan mengecek nilai kosong dan duplikat. Beberapa kolom yang tidak relevan, seperti kolom key, dihapus agar analisis lebih fokus. Kolom streams juga dikonversi menjadi tipe numerik agar dapat diproses dengan benar.

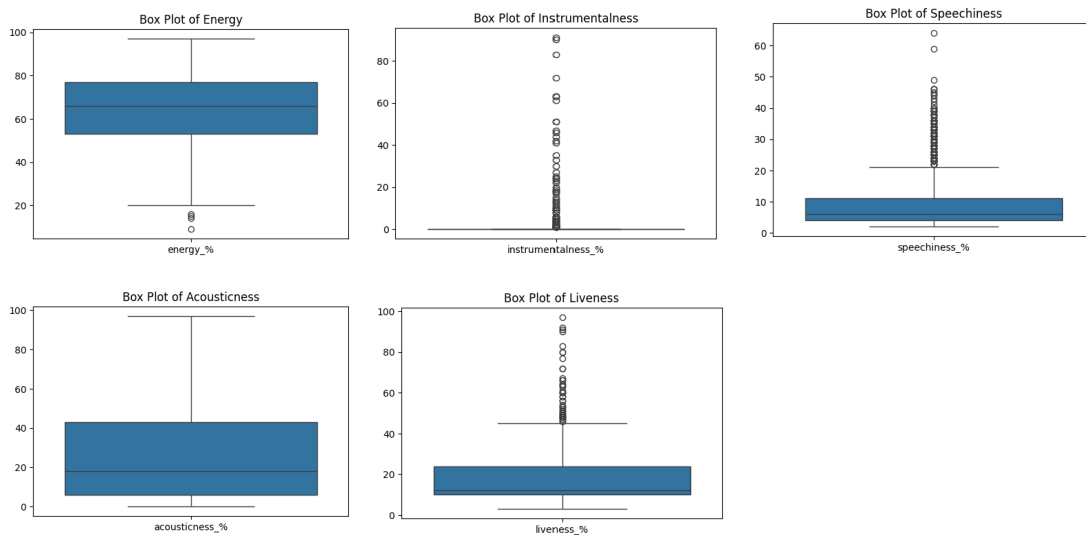
	track_name	artist(s)_name	artist_count	released_year	released_month	released_day	in_spotify_playlists	in_spotify_charts	streams	in_apple_playlists	...	bpm	key	mode	danceability_%	valence_%	energy_%	acousticness_%	instrumentalness_%	liveness_%	speechiness_%	Music_cluster
0	Seven (feat. Latto) (Explicit Ver.)	Latto, Jung Kook	2	2023	7	14	553	147	141381703.0	43	...	125	Major		80	89	83	31	0	8	4	0
1	LALA	Myke Towers	1	2023	3	23	1474	48	133716286.0	48	...	92	Major		71	61	74	7	0	10	4	0
2	vampire	Olivia Rodrigo	1	2023	6	30	1397	113	140003974.0	94	...	138	Major		51	32	53	17	0	31	6	2
3	Cruel Summer	Taylor Swift	1	2019	8	23	7858	100	800840817.0	116	...	170	Major		55	58	72	11	0	11	15	2
4	WHERE SHE GOES	Bad Bunny	1	2023	5	18	3133	50	303236322.0	84	...	144	Minor		65	23	80	14	63	11	6	4

#### 3.3 Distribusi Data dan Boxplot

Dilakukan visualisasi distribusi data menggunakan boxplot untuk fitur-fitur numerik utama dalam dataset. Boxplot digunakan untuk mengidentifikasi sebaran data sekaligus mendeteksi adanya outlier pada variabel seperti jumlah artis per lagu, tahun, bulan, dan hari perilisan lagu, serta atribut popularitas di platform Spotify, Apple Music, dan Deezer. Selain itu, fitur musik seperti beats per minute, danceability,

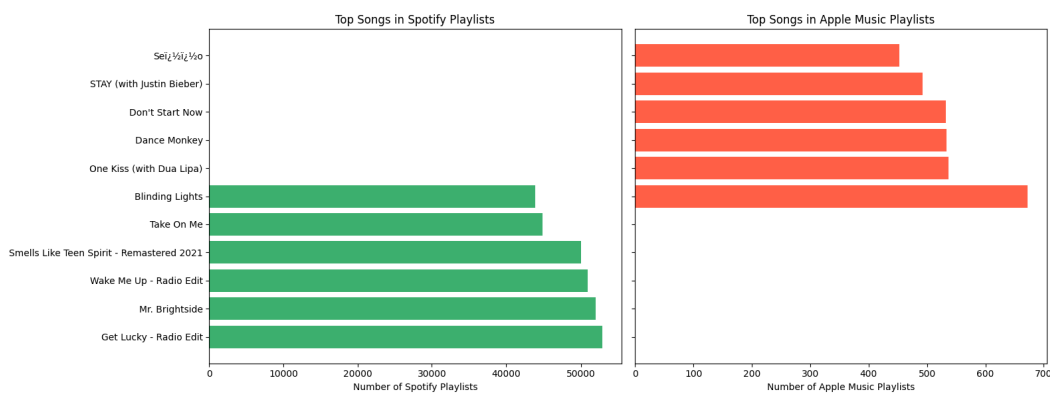
energy, valence, acousticness, instrumentality, liveness, dan speechiness juga divisualisasikan

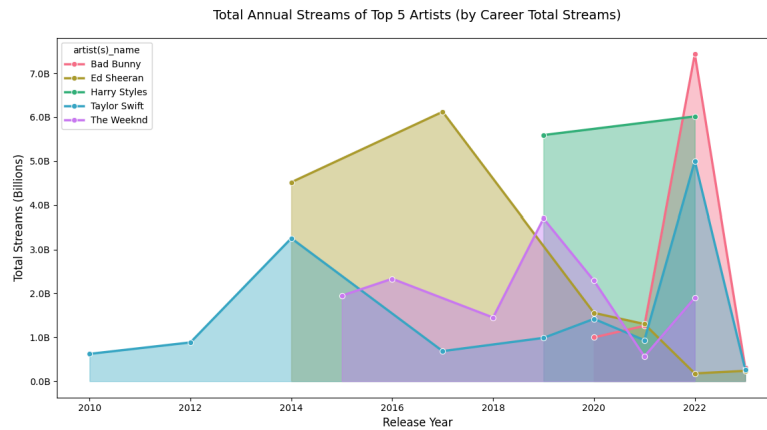




### 3.3 Visualisasi Popularitas Lagu dan Artis

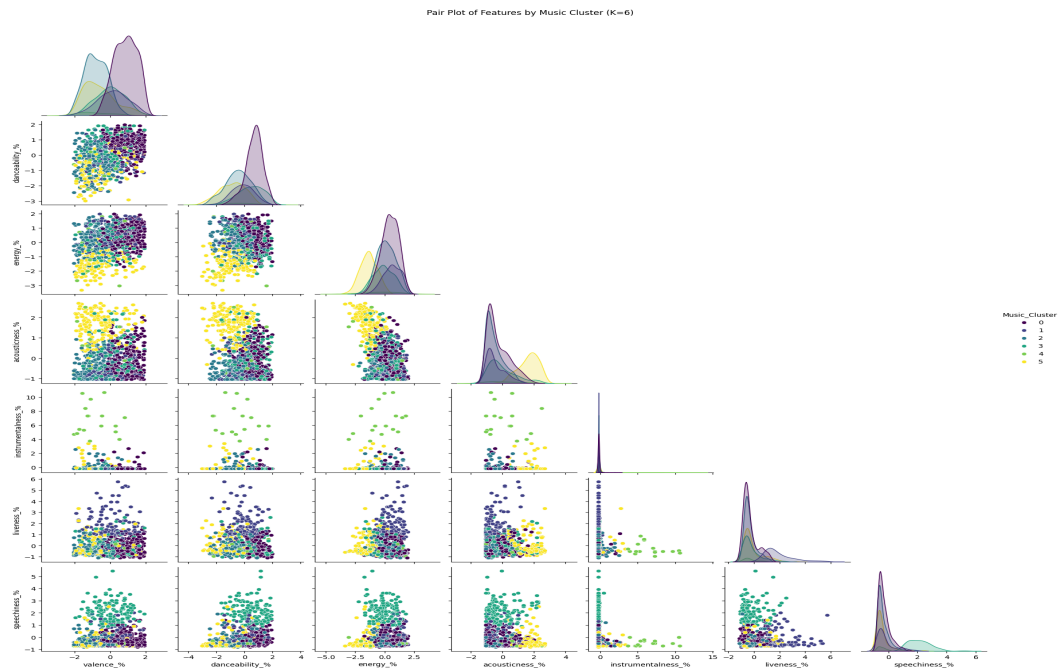
Dilakukan visualisasi lagu-lagu terpopuler berdasarkan jumlah playlist di Spotify dan Apple Music. Visualisasi ini menampilkan enam lagu dengan jumlah kemunculan terbanyak di playlist masing-masing platform. Selain itu, dilakukan analisis tren tahunan total jumlah streaming dari lima artis dengan total streaming tertinggi. Hasil visualisasi menunjukkan adanya peningkatan signifikan pada tahun-tahun tertentu, yang kemungkinan terkait dengan perilisan album atau musik baru.





### 3.4 Clustering Lagu dengan K-Means

Untuk mengelompokkan lagu berdasarkan karakteristik musiknya, dilakukan clustering menggunakan algoritma K-Means dengan fitur-fitur seperti danceability, valence, energy, dan lain-lain. Data fitur terlebih dahulu dinormalisasi agar skala antar fitur seimbang. Clustering ini membantu mengidentifikasi pola-pola dalam data musik dan memudahkan analisis karakteristik lagu berdasarkan kelompoknya.



### 3.5 Analisis Popularitas setiap Cluster

Untuk menilai tingkat popularitas dari masing-masing kluster musik, dilakukan analisis rata-rata jumlah streams per kluster. Hasilnya menunjukkan bahwa kluster "Mellow & Reflective Acoustic Pieces" memiliki rata-rata streams tertinggi, disusul oleh "Intense & Brooding Rhythmic Tracks" dan "Upbeat & Joyful Dance Hits". Hal ini mengindikasikan bahwa musik dengan nuansa tenang dan emosional ternyata memiliki daya tarik yang besar di kalangan pendengar.

Average 'streams' per Music Cluster (Whole Numbers):

```
Music_Cluster
5      581635370
2      539048559
0      533662613
1      495382415
3      382790853
4      344317106
```

Name: streams, dtype: int64

Average streams with Cluster Names (Whole Numbers):

Music_Cluster	Cluster_Name	streams
0	5 Mellow & Reflective Acoustic Pieces	581635370
1	2 Intense & Brooding Rhythmic Tracks	539048559
2	0 Upbeat & Joyful Dance Hits	533662613
3	1 High-Energy Live Performances	495382415
4	3 Rhythmic & Speech-Driven Grooves	382790853
5	4 Atmospheric & Melancholic Instrumentals	344317106

Selanjutnya, ditampilkan lima lagu teratas dalam setiap kluster berdasarkan jumlah streams-nya. Visualisasi ini memperlihatkan lagu-lagu populer yang menjadi representasi dominan di masing-masing kluster, seperti "Shape of You", "Blinding Lights", dan "Someone You Loved". Daftar ini membantu memahami karakteristik musik dalam setiap kluster.

```

--- Top 5 Songs in Each Cluster (by Streams) ---

--- Upbeat & Joyful Dance Hits (Cluster 0) ---
0          track_name      streams
1          Shape of You  3562543890
2          Dance Monkey  2864791672
3 Sunflower - Spider-Man: Into the Spider-Verse  2808096550
4          Believer      2594040133
5          Closer        2591224264

--- High-Energy Live Performances (Cluster 1) ---
0          track_name      streams
1          One Dance     2713922350
2          As It Was     2513188493
3          Watermelon Sugar  2322580122
4 MONTERO (Call Me By Your Name)  1735441776
5          In The End    1624165576

--- Intense & Brooding Rhythmic Tracks (Cluster 2) ---
0          track_name      streams
1          Blinding Lights  3703895074
2 STAY (with Justin Bieber)  2665343922
3          Perfect        2559529074
4          Sweater Weather  2282771485
5          Something Just Like This  2204080728

--- Rhythmic & Speech-Driven Grooves (Cluster 3) ---
0          track_name      streams
1          Starboy        2565529693
2          Love Yourself  2123309722
3 Lose Yourself - Soundtrack Version  1829992958
4          No Role Modelz  1791000570
5          LA CANCIIIZ    1435127549

--- Atmospheric & Melancholic Instrumentals (Cluster 4) ---
0          track_name      streams
1          Apocalypse     841749534
2          Freaks         824420218
3 Master of Puppets (Remastered)  704171068
4          After Dark     646886885
5          Alien Blues    370068639

--- Mellow & Reflective Acoustic Pieces (Cluster 5) ---
0          track_name      streams
1          Someone You Loved  2887241814
2 Say You Won't Let Go  2420461338
3 lovely - Bonus Track  2355719893
4          Lucid Dreams    2288695111
5          Photograph      2236667932
-----

```

## **BAB IV**

### **Kesimpulan dan Saran**

#### **3.1 Kesimpulan**

Melalui analisis clustering menggunakan algoritma K-Means, lagu-lagu populer berhasil dikelompokkan ke dalam enam klaster berdasarkan atribut audio seperti energy, valence, danceability, dan lainnya. Masing-masing klaster menunjukkan karakteristik musikal yang unik, seperti lagu-lagu akustik yang mellow, atau lagu-lagu dengan ritme intens dan enerjik. Analisis rata-rata jumlah streams menunjukkan bahwa klaster “Mellow & Reflective Acoustic Pieces” menjadi yang paling populer di antara lainnya. Temuan ini membuktikan bahwa metode unsupervised learning seperti K-Means dapat dimanfaatkan untuk memahami segmentasi musik secara efektif, serta memberikan wawasan tentang preferensi pendengar dalam skala besar.

#### **3.2 Saran**

- Penambahan data pendukung seperti genre atau lirik lagu bisa membantu membuat klaster yang lebih akurat dan bermakna.
- Evaluasi model clustering bisa ditambahkan agar kualitas hasil klaster lebih terukur.
- Mencoba menggunakan metode lain seperti Hierarchical Clustering untuk melihat apakah ada struktur klaster lain yang bisa ditemukan.

Analisis ini dapat dikembangkan lebih lanjut untuk membantu sistem rekomendasi lagu atau pembuatan playlist otomatis berdasarkan klaster yang terbentuk.