

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan dalam teknologi informasi dan komunikasi telah mengubah cara kita mengakses serta menikmati musik. Salah satu platform yang paling berpengaruh dalam hal ini adalah *spotify*, yang telah menjadi salah satu layanan *streaming* musik paling populer di dunia. Dengan lebih dari 500 juta pengguna aktif, *spotify* memberikan akses ke jutaan lagu dari berbagai genre, memungkinkan pengguna untuk menemukan dan mendengarkan musik yang sesuai dengan preferensi mereka. Namun menurut (Nuriska et al., 2023), dengan banyaknya pilihan yang ada, tantangan muncul dalam hal pengelompokan dan rekomendasi lagu yang relevan bagi pendengar.

Fitur audio merupakan parameter penting dalam analisis musik. Penelitian sebelumnya menunjukkan bahwa pengelompokan lagu berdasarkan fitur-fitur ini dapat menghasilkan kategori yang mencerminkan emosi tertentu, seperti bahagia, sedih, atau santai. Sebagai contoh, sebuah studi yang dilakukan pada tahun 2023 mengungkapkan bahwa klusterisasi lagu-lagu populer di *Spotify* dapat dilakukan berdasarkan suasana hati menggunakan algoritma *K-means*, dengan hasil yang menunjukkan distribusi anggota yang merata di setiap klaster (Nuriska et al., 2023).

Dengan meningkatnya jumlah data musik yang tersedia di platform streaming seperti *spotify*, penting untuk menerapkan metode analisis yang efisien dan akurat. Penggunaan teknik *machine learning* seperti *K-means* tidak hanya meningkatkan efisiensi dalam pengolahan data tetapi juga akurasi dalam pengelompokan. Hal ini sangat relevan mengingat kompleksitas preferensi pengguna yang beragam dan dinamis (Reyhan Jarsi Yoga et al., 2024).

Algoritma *K-means* dipilih dalam penelitian ini karena sederhana namun efektif dalam mengelompokkan data berdasarkan kesamaan fitur. *K-means* bekerja dengan membagi data ke dalam sejumlah kluster yang ditentukan sebelumnya, di mana setiap kluster memiliki pusat (*centroid*) yang mewakili kumpulan data tersebut. Proses iterasinya, yang melibatkan perhitungan jarak *Euclidean* antara data dan *centroid*, memastikan bahwa setiap data ditempatkan di kluster yang paling sesuai. *K-means* cocok untuk analisis data berukuran besar, seperti data lagu di *Spotify*, karena efisiensinya dalam menangani banyak fitur secara bersamaan. Penelitian sebelumnya telah menunjukkan bahwa algoritma ini dapat memberikan hasil yang baik dalam pengelompokan data audio, seperti yang ditunjukkan oleh (Marlia et al., 2024), di mana *K-means* berhasil mengidentifikasi pola-pola unik dalam fitur audio lagu populer.

Kelemahan utama algoritma K-Means terletak pada pemilihan pusat kluster awal secara acak, yang sering kali menyebabkan hasil klusterisasi tidak stabil dan suboptimal. Untuk mengatasi hal ini, K-Means++ hadir dengan proses inisialisasi pusat kluster yang lebih cerdas melalui probabilitas berbobot, sehingga mempercepat konvergensi dan meningkatkan kualitas hasil klusterisasi. Dibandingkan dengan K-Means, algoritma K-Means++ lebih efektif dalam menangani data besar dan kompleks serta terbukti menghasilkan kluster yang lebih stabil dan optimal dalam berbagai kasus (Putra Nugraha et al., 2024). Oleh karena itu, penelitian ini memilih K-Means++ untuk memastikan hasil klusterisasi lagu yang lebih akurat berdasarkan fitur audio di *Spotify*.

Spotify juga menyediakan *API* yang memungkinkan peneliti mengakses data musik secara langsung. Data ini mencakup berbagai atribut audio yang relevan untuk analisis klusterisasi. Dengan memanfaatkan data tersebut, penelitian ini bertujuan untuk melakukan klusterisasi lagu di platform *spotify* berdasarkan fitur audio menggunakan algoritma *K-means*. Penelitian ini diharapkan dapat memberikan kontribusi terhadap

pemahaman lebih lanjut mengenai preferensi musik pengguna serta meningkatkan pengalaman mendengarkan mereka melalui rekomendasi yang lebih tepat sasaran (Laili Musyarofah et al., n.d.).

Hasil dari klasterisasi ini tidak hanya bermanfaat bagi pengguna dalam menemukan musik yang sesuai dengan selera mereka tetapi juga dapat digunakan oleh *spotify* untuk mengembangkan sistem rekomendasi yang lebih canggih. Dengan memahami pola-pola dalam data audio dan bagaimana pengguna berinteraksi dengan lagu-lagu tertentu, *spotify* dapat menyajikan rekomendasi yang lebih personal dan relevan bagi setiap pengguna (Marlia et al., 2024).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, dapat diidentifikasi beberapa permasalahan utama yang menjadi fokus penelitian ini:

- 1) Sejauh mana algoritma *K-means* dan *K-means++* efektif untuk melakukan klasterisasi lagu berdasarkan fitur audio pada dataset dengan ukuran besar dan karakteristik yang kompleks?
- 2) Bagaimana fitur audio seperti *danceability*, *energy*, *valence*, *loudness*, *tempo*, *speechiness*, *acousticness*, *instrumentalness*, dan *liveness* dapat dimanfaatkan untuk menghasilkan klaster musik yang informatif dan relevan bagi analisis data musik?

1.3 Tujuan Penelitian

Berdasarkan pada rumusan masalah di atas, maka tujuan dari penelitian ini adalah sebagai berikut:

- 1) Menganalisis efektivitas algoritma *K-means* dan *K-Means++* dalam melakukan klasterisasi lagu pada dataset *spotify* yang besar dan kompleks, termasuk identifikasi kendala dan peluang peningkatan akurasi klasterisasi.

- 2) Mengoptimalkan penggunaan fitur audio seperti seperti *danceability*, *energy*, *valence*, *loudness*, *tempo*, *speechiness*, *acouticness*, *instrumentalness*, dan *liveness* untuk menghasilkan klaster musik yang informatif dan relevan, sehingga dapat memberikan wawasan mendalam terhadap pola dan karakteristik musik.

1.4 Batasan Masalah

Untuk menjaga fokus dan keterbatasan penelitian, beberapa batasan yang diberlakukan dalam penelitian ini adalah:

- 1) Penelitian ini akan menggunakan dataset lagu dari *spotify*, dengan fokus pada fitur audio. Dataset akan dibatasi pada jumlah data yang representatif untuk menganalisis performa algoritma *K-means* dan *K-means++*, tetapi tidak mencakup seluruh katalog *spotify*.
- 2) Hanya algoritma *K-means* dan *K-means++* yang akan dianalisis dalam penelitian ini untuk klasterisasi lagu.
- 3) Penelitian hanya memanfaatkan beberapa fitur audio, seperti seperti *danceability*, *energy*, *valence*, *loudness*, *tempo*, *speechiness*, *acouticness*, *instrumentalness*, dan *liveness* untuk klasterisasi. Fitur-fitur lain yang tersedia di dataset *spotify* tidak akan dianalisis lebih lanjut.
- 4) Penelitian ini akan mengevaluasi hasil klasterisasi berdasarkan metrik internal seperti *silhouette score*, tanpa melibatkan analisis preferensi pengguna secara langsung.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

- 1) Memberikan panduan dan rekomendasi tentang bagaimana algoritma *K-means* dapat diterapkan secara efektif pada dataset musik besar,

termasuk langkah-langkah *preprocessing* dan evaluasi untuk hasil yang optimal.

- 2) Menunjukkan bagaimana fitur dapat dimanfaatkan untuk menghasilkan klaster yang relevan dan memberikan wawasan mendalam tentang pola dan karakteristik musik.

