

Segmentasi Pengguna Spotify Berdasarkan Preferensi Musik dengan Algoritma K-Means Clustering

Kadek Bisma Dharmasena^{a1}, Cokorda Pramatha^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana

Jln. Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, 08261, Bali, Indonesia

¹bisma2412@gmail.com

²cokorda@unud.ac.id

Abstract

Music streaming platforms like Spotify have become integral to the daily lives of millions globally, offering personalized listening experiences. However, managing a vast music catalog to present relevant content to each user remains a challenge. This study explores the application of the K-means clustering algorithm to segment Spotify users based on their music preferences. The goal is to group users into clusters with similar tastes to enhance targeted marketing and user engagement. We utilized a secondary dataset of trending Spotify songs and their attributes from 2023. Through data preprocessing, feature selection, and normalization, we prepared the data for clustering. The optimal number of clusters was determined using the Elbow Method, resulting in six distinct clusters. Each cluster represents unique music preferences, analyzed through metrics such as danceability, energy, and popularity. The findings demonstrate that K-means clustering effectively identifies user segments, providing insights for improving personalized recommendations and marketing strategies. This research underscores the potential of machine learning in optimizing user experiences on music streaming platforms.

Keywords: Spotify, Elbow Method, K-Means Clustering, Music, User Segmentation

1. Pendahuluan

Platform streaming musik seperti Spotify telah menjadi bagian integral dari kehidupan sehari-hari bagi jutaan pengguna di seluruh dunia. Dengan akses tak terbatas ke jutaan lagu dari berbagai genre dan artis, Spotify menawarkan pengalaman mendengarkan musik yang disesuaikan dengan preferensi individual pengguna. Namun, dalam mengelola katalog musik yang luas ini, Spotify dihadapkan pada tantangan untuk menyajikan konten yang relevan dan menarik bagi setiap pengguna. Salah satu cara untuk mengatasi tantangan ini adalah dengan menggunakan teknik segmentasi pengguna. Segmentasi pengguna memungkinkan Spotify untuk memahami preferensi musik dari setiap pengguna dan menyajikan konten yang lebih sesuai dengan selera mereka. Dalam artikel ini, kami akan mengeksplorasi penerapan algoritma K-Means Clustering untuk melakukan segmentasi pengguna Spotify berdasarkan preferensi musik mereka. K-Means Clustering adalah salah satu teknik yang populer dalam analisis data yang memungkinkan kita untuk mengelompokkan data ke dalam kategori yang berbeda berdasarkan kemiripan karakteristik. Penerapan algoritma K-Means Clustering pada data preferensi musik pengguna Spotify diharapkan akan memungkinkan kita untuk mengidentifikasi kelompok-kelompok pengguna dengan preferensi musik yang serupa. Dengan demikian, artikel ini tidak hanya akan menjelaskan konsep dasar dari algoritma K-Means Clustering, tetapi juga akan menunjukkan bagaimana teknik ini dapat diterapkan secara praktis untuk meningkatkan pengalaman pengguna dalam platform streaming musik seperti Spotify.

2. Metode Penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian kali ini adalah dataset sekunder yang berupa dataset lagu-spotify beserta atributnya yang trending pada tahun 2023 dari website Kaggle.

2.2. Spotify

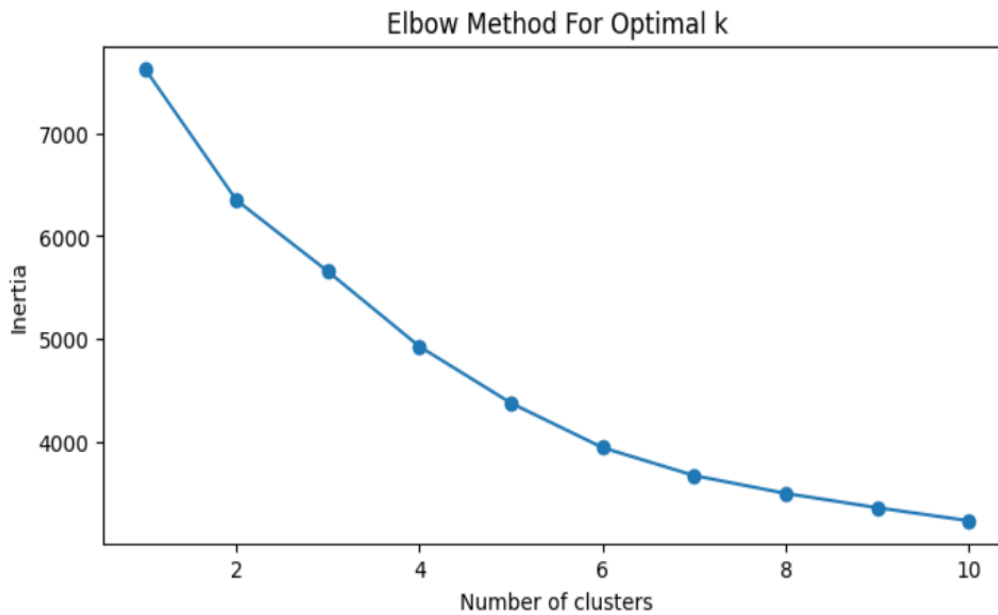
Spotify adalah layanan streaming musik yang populer di seluruh dunia, menawarkan akses ke jutaan lagu dan podcast dari berbagai genre dan artis. Diluncurkan pada tahun 2008, Spotify telah menjadi platform utama bagi pecinta musik yang ingin menikmati beragam pilihan musik kapan saja dan di mana saja. Pengguna dapat mendengarkan musik secara gratis dengan iklan atau berlangganan layanan premium untuk menikmati pengalaman bebas iklan, kualitas audio lebih baik, dan fitur tambahan lainnya. Spotify dikenal karena kemampuannya untuk menyediakan pengalaman mendengarkan yang personal. Dengan menggunakan algoritma canggih, Spotify menganalisis kebiasaan mendengarkan penggunanya dan memberikan rekomendasi musik yang disesuaikan dengan preferensi masing-masing individu. Fitur seperti "Discover Weekly" dan "Release Radar" membantu pengguna menemukan musik baru yang sesuai dengan selera mereka. Selain itu, Spotify memungkinkan pengguna untuk membuat, berbagi, dan mengikuti playlist, serta melihat apa yang sedang didengarkan teman-teman mereka. Ini menambahkan elemen sosial yang memperkaya pengalaman mendengarkan musik. Setiap akhir tahun, fitur "Spotify Wrapped" memberikan ringkasan visual tentang kebiasaan mendengarkan pengguna selama tahun tersebut, menyoroti artis, lagu, dan genre yang paling sering diputarkan. Dengan berbagai fitur inovatif dan beragam pilihan musik, Spotify terus menjadi platform yang menarik bagi berbagai segmen pengguna dengan preferensi musik yang berbeda-beda.

2.3. Preprocessing Data

Tahap preprocessing dilakukan agar data menjadi lebih siap untuk diolah dan menjadi lebih mudah untuk digunakan. Dalam penelitian ini, tahap preprocessing mencakup penyamaan tipe data, pemilihan fitur, lalu normalisasi. Pada tahap ini tidak dilakukan data cleaning, karena berdasarkan fitur yang dipilih tidak ada data kosong ataupun data yang tidak terdapat nilai. Fitur dari setiap lagu yang dipilih untuk penelitian ini adalah 'bpm', 'danceability_%', 'valence_%', 'energy_%', 'acousticness_%', 'instrumentalness_%', 'liveness_%', dan 'speechiness_%'. Sedangkan untuk fitur popularitasnya terdapat 'songs', 'streams', 'in_spotify_playlists', dan 'in_spotify_charts'. Lalu normalisasi perlu dilakukan terlebih dahulu untuk menyamakan rentang nilai atau domain setiap atribut sehingga mencapai rentang nilai 0, 1 sebelum melanjutkan ke tahap klusterisasi [1].

2.4. Elbow Method

Elbow Method merupakan metode yang digunakan dalam analisis kluster, khususnya K-Means Clustering, untuk menentukan jumlah kluster yang optimal. Ilustrasi nilai K pada metode Elbow dengan K-Means adalah grafik hubungan antara kluster dengan penurunan kesalahan, di mana nilai K yang meningkat akan membuat grafik tersebut menurun secara perlahan hingga mencapai nilai K yang stabil [2]. Metode ini melibatkan penerapan algoritma K-Means dengan berbagai jumlah kluster (K) yang berbeda dan menghitung Within-Cluster Sum of Squares (WCSS) untuk setiap K. WCSS adalah jumlah kuadrat jarak dari setiap titik data ke pusat kluster terdekatnya, yang mengukur seberapa baik klustering tersebut meminimalkan variabilitas dalam kluster. Hasil WCSS kemudian dipetakan dalam grafik dengan sumbu y sebagai WCSS dan sumbu x sebagai jumlah kluster K [3]. Bentuk grafik ini biasanya menunjukkan penurunan tajam di awal, diikuti dengan penurunan yang semakin landai, membentuk bentuk mirip siku atau "elbow." Titik di mana penurunan WCSS mulai melambat secara signifikan menandakan jumlah kluster yang optimal. Dengan memilih K pada titik "elbow" ini, kita dapat menentukan jumlah kluster yang memberikan keseimbangan terbaik antara kompleksitas model dan kemampuan untuk menjelaskan variabilitas data. Berikut ini merupakan grafik yang didapat dari elbow method.



Gambar 1. Grafik Elbow Method

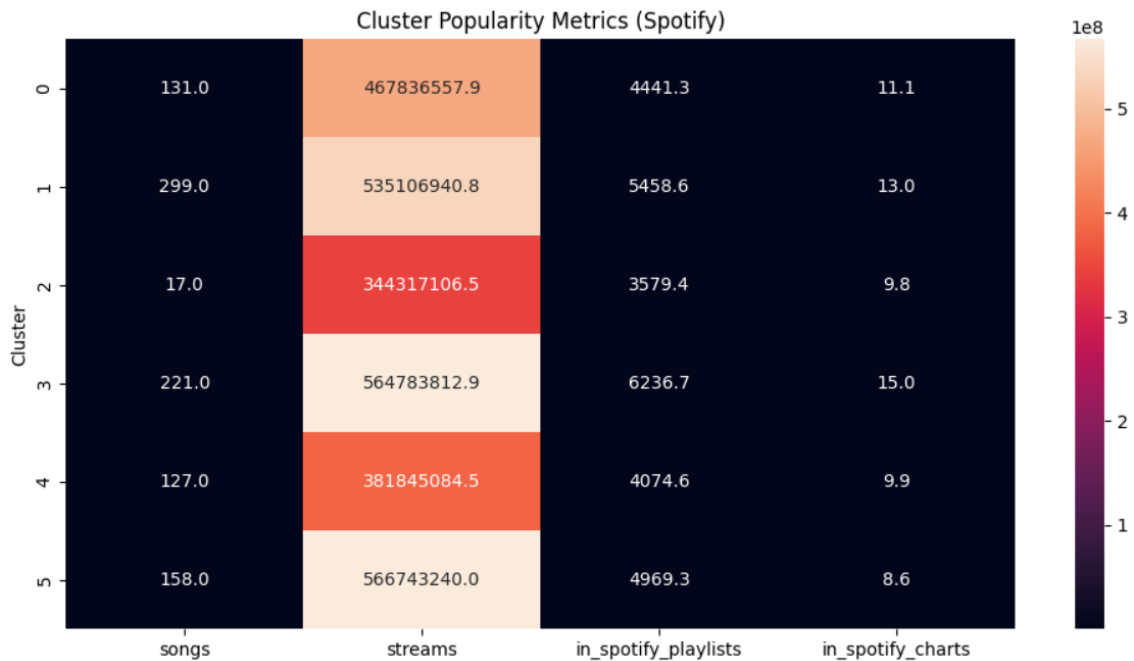
2.5. Algoritma K-Means Clustering

Klasterisasi merujuk pada proses mengelompokkan entitas seperti rekaman, pengamatan, atau objek-objek, berdasarkan kesamaan mereka untuk membentuk kelas atau kelompok [4]. Salah satu contoh algoritma klasterisasi adalah K-Means Clustering. Algoritma K-Means Clustering beroperasi dengan cara memisahkan data ke dalam kelompok-kelompok yang berbeda, tergantung pada kesamaan atribut yang dimiliki [5]. Tujuannya adalah untuk meminimalkan jumlah variabilitas dalam setiap kelompok, dengan menempatkan titik pusat (centroid) yang mewakili setiap kelompok sedekat mungkin dengan anggota-anggotanya. Cara kerja algoritma ini dimulai dengan inisialisasi, di mana titik-titik awal dipilih secara acak sebagai pusat kelompok. Setiap titik data dalam dataset kemudian diberikan label yang sesuai dengan kelompok terdekat (berdasarkan jarak Euclidean, misalnya) dari titik pusat yang diinisialisasi sebelumnya. Selanjutnya, pusat baru untuk setiap kelompok dihitung sebagai rata-rata dari semua titik data yang termasuk dalam kelompok tersebut. Langkah ini diulangi sampai tidak ada perubahan dalam label kelompok dari satu iterasi ke iterasi berikutnya, atau hingga mencapai jumlah iterasi maksimum yang ditentukan sebelumnya. Algoritma K-Means Clustering cenderung cepat dan efisien dalam menangani dataset besar, meskipun sensitif terhadap inisialisasi titik-titik pusat awalnya. Oleh karena itu, seringkali disarankan untuk menjalankan algoritma beberapa kali dengan inisialisasi yang berbeda dan memilih hasil terbaik berdasarkan evaluasi. Evaluasi hasil Klasterisasi dapat dilakukan menggunakan metrik seperti inersia (jumlah total kuadrat jarak antara setiap titik data dan pusat kelompoknya), serta metrik eksternal seperti indeks Inertia atau indeks Silhouette untuk mengevaluasi kualitas klasterisasi. Dengan memahami langkah-langkah dasar dan cara kerja Algoritma K-Means Clustering, peneliti dapat menggunakannya secara efektif untuk menganalisis dan mengelompokkan data, termasuk dalam kasus segmentasi pengguna Spotify berdasarkan preferensi musik.

3. Hasil dan Pembahasan

Penelitian ini akan menguji beberapa fitur dari tiap lagu dan membandingkannya dengan tingkat popularitasnya, lalu model akan menciptakan klasterisasi terhadap lagu-lagu tersebut dengan algoritma K-Means Clustering. Setelah tahap preprocessing dilakukan, dan hasil dari Elbow Method didapatkan, peneliti mengambil keputusan akan menggunakan klaster atau 'K' sebanyak 6, lalu menjalankan klasterisasi berdasarkan fitur-fitur yang disimpan dalam variabel "features".

Berikut ini adalah gambaran heatmap dari hasil klusterisasi yang dibandingkan dengan popularity metrics.



Gambar 2. Heatmap Klusterisasi

Lalu berikut ini adalah data rata rata dari setiap klaster yang dihasilkan, beserta jumlah lagu yang ada pada tiap klaster.

Tabel 1. Hasil Rata-rata Klusterisasi

Cluster	in_spotify_playlists	in_spotify_streams	bpm	Danceability_%	valence_%
0	4441.297710	11.106870	4.678366e+08	116.969466	67.656489
1	5458.645485	13.030100	5.351069e+08	116.066890	77.200669
2	3579.352941	9.823529	3.443171e+08	122.764706	60.352941
3	6236.696833	14.968326	5.647838e+08	134.972851	58.443439
4	4074.582677	9.889764	3.818451e+08	127.314961	74.181102
5	4969.259494	8.626582	5.667432e+08	118.158228	53.879747

energy_%	acousticness_%	instrumentalness_%	liveness_%	speechiness_%	songs
73.305344	17.000000	0.068702	43.473282	8.068702	131.0
71.337793	19.882943	0.314381	12.829431	6.989967	299.0
58.647059	31.000000	57.411765	14.117647	5.411765	17.0
66.778281	13.425339	0.963801	13.990950	6.330317	221.0
62.559055	26.259843	0.078740	15.118110	31.346457	127.0
41.930380	68.259494	1.297468	16.291139	6.556962	158.0

Dari tabel 1 dapat dilihat bahwa tiap kluster mempunyai variasi nilai yang beragam dan signifikan dalam preferensi musik di setiap klaster. Klaster 0 memiliki danceability dan energy yang tinggi dengan acousticness rendah, menunjukkan preferensi untuk musik yang energik dan mudah

untuk menari. Klaster 1, yang merupakan klaster terbesar, menyukai musik upbeat dan positif dengan nilai danceability dan valence yang tinggi. Klaster 2, yang paling sedikit jumlah lagunya, cenderung menyukai musik instrumental dengan instrumentalness tinggi dan valence rendah. Klaster 3 memiliki popularitas tertinggi dalam metrik Spotify dan menunjukkan preferensi untuk musik dengan tempo cepat namun tidak selalu upbeat. Klaster 4 menyukai musik akustik yang tetap bisa digunakan untuk menari, dengan nilai acousticness dan danceability yang tinggi. Terakhir, klaster 5 menunjukkan preferensi untuk musik yang lebih tenang dan akustik, dengan acousticness tertinggi dan energy terendah. Data ini membantu memahami preferensi musik dari setiap klaster dan dapat digunakan untuk strategi pemasaran dan pengembangan konten yang lebih tepat sasaran.

4. Kesimpulan

Berdasarkan penelitian ini, dapat diperoleh kesimpulan bahwa segmentasi pengguna spotify berdasarkan preferensi musiknya dengan metode K-Means Clustering telah berhasil dilakukan. Dengan memahami preferensi musik dari setiap klaster, data tersebut dapat digunakan baik untuk strategi pemasaran maupun meningkatkan pengalaman pengguna yang lebih relevan, menarik dan tepat sasaran

Daftar Pustaka

- [1] I. G. A. M. Pratama, L. G. Astuti, I. M. Widiartha, I. G. N. C. P. Anom, C. R. A. Pramarta, and I. D. M. B. A. Darmawan, "Diagnosis Penyakit Ginjal Kronis dengan Algoritma C4.5, K-Means dan BPSO," 2022.
- [2] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," in IOP Conference Series: Materials Science and Engineering, Institute of Physics Publishing, Apr. 2018. doi: 10.1088/1757-899X/336/1/012017.
- [3] D. M. Saputra, D. Saputra, and L. D. Oswari, "Advances in Intelligent Systems Research," 2020.
- [4] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," Jurnal Teknologi dan Sistem Informasi (JTSI), vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [5] C. Hafidz Ardana et al., "Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering," 2024.

Halaman ini sengaja dibiarkan kosong