

**KLASTERISASI LAGU PADA PLATFORM *SPOTIFY*
BERDASARKAN FITUR AUDIO MENGGUNAKAN
ALGORITMA *K-MEANS* DAN *K-MEANS++***

SKRIPSI SARJANA INFORMATIKA

Oleh

Ramadanti Indah Safitri

217064516064



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL**

2024

**KLASTERISASI LAGU PADA PLATFORM *SPOTIFY* BERDASARKAN
FITUR AUDIO MENGGUNAKAN ALGORITMA *K-MEANS* DAN *K-
MEANS++***

SKRIPSI

Diajukan kepada:

Universitas Nasional Jakarta

Untuk memenuhi Salah Satu Persyaratan dalam

Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:

RAMADANTI INDAH SAFITRI

NIM. 217064516064

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA

UNIVERSITAS NASIONAL

JAKARTA

2024

HALAMAN PENGESAHAN

TUGAS AKHIR

**“KLASTERISASI LAGU PADA PLATFORM SPOTIFY
BERDASARKAN FITUR AUDIO MENGGUNAKAN
ALGORITMA *K-MEANS* DAN *K-MEANS++*”**



Ramadanti Indah Safitri

217064516064

Dosen Pembimbing 1



(Sari Ningsih, S.Si., M.M)

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

**“KLASTERISASI LAGU PADA PLATFORM SPOTIFY BERDASARKAN
FITUR AUDIO MENGGUNAKAN ALGORITMA *K-MEANS* DAN *K-MEANS++*”**

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 4 Maret 2024



Ramadanti Indah Safitri

217064516064

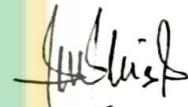
LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

**“KLASTERISASI LAGU PADA PLATFORM SPOTIFY
BERDASARKAN FITUR AUDIO MENGGUNAKAN
ALGORITMA K-MEANS DAN K-MEANS++”**

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2024-2025 pada tanggal 25 Februari Tahun 2025.

Dosen Pembimbing 1



Sari Ningsih, S.Si., M.M

NIDN : 0302066701

Ketua Program Studi



Ratih Titi Komala Sari, ST., MM., MMSI

NIDN : 0301038302

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Ramadanti Indah Safitri
NPM : 217064516064
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Informatika
Tanggal Sidang : 25 Februari 2025

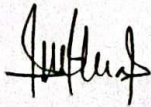
JUDUL DALAM BAHASA INDONESIA :

**“KLASTERISASI LAGU PADA PLATFORM SPOTIFY BERDASARKAN
FITUR AUDIO MENGGUNAKAN ALGORITMA K-MEANS DAN K-
MEANS++”**

JUDUL DALAM BAHASA INGGRIS :

**“CLUSTERING SONGS ON SPOTIFY PLATFORM BASED ON AUDIO
FEATURES USING K-MEANS AND K-MEANS++ ALGORITHMS”**

TANDA TANGAN DAN TANGGAL

Sari Ningsih, S.Si., M.M.	Ratih Titi Komala Sari, ST., MM., MMSI	Ramadanti Indah Safitri
TGL : 4 Maret 2025	TGL : 4 Maret 2025	TGL : 4 Maret 2025
		

KATA PENGANTAR

Segala puji hanya milik Allah Subhanahu Wa Ta'ala atas segala nikmat dan kasih sayang Nya yang telah memudahkan penulis untuk menyelesaikan skripsi yang berjudul “KLAUSTERISASI LAGU PADA PLATFORM *SPOTIFY* BERDASARKAN FITUR AUDIO MENGGUNAKAN *ALGORITMA K-MEANS* DAN *K-MEANS++*”. Semoga shalawat dan salam senantiasa terlimpah kepada Nabi Muhammad Sallallahu ‘Alaihi wa Sallam. Dan semoga kita semua mendapat syafaatnya di hari kiamat nanti, Aamiin.

Penulis mengucapkan rasa syukur dan terima kasih yang tak terhingga kepada semua pihak-pihak yang selalu memberikan bantuan dan motivasi kepada penulis untuk dapat menyelesaikan skripsi ini. Ucapan ini penulis sampaikan kepada:

1. Mamah Eka Defiyanti dan Ayah Sanusi, S.E serta keluarga penulis yang selalu memberikan doa, dukungan fisik maupun mental tanpa henti. Kehadiran dan kasih sayang mereka memberikan penulis kekuatan luar biasa untuk terus berjuang dalam menyelesaikan penelitian ini.
2. Bapak Dr. Agung Triayudi, S.Kom., M.Kom selaku Dekan FTKI. Ibu Ratih Titi Komala Sari, S.T., M.M., selaku Ketua Program Studi. Bapak Yunan Fauzi Wijaya, S.KOM., MMSI selaku Kepala Lab Networking Database and Cloud.
3. Ibu Sari Ningsih., S.Si., MM selaku Dosen Pembimbing. Ibu Nur Hayati, S.Si., MTI selaku Dosen Penguji 1 dan Bapak Albaar Rubhasy, S.Si, MTI selaku Dosen Penguji 2.
4. Segenap Dosen, Admin, Sekretariat, Laboran, dan Mahasiswa di FTKI Universitas Nasional yang telah memberikan banyak ilmu bermanfaat, serta kesempatan untuk terus berkembang dan belajar selama masa studi saya.
5. Qiqi Shalzabilla Putri yang selalu memberikan bantuan, dukungan, dan kebersamaan selama ini. Merupakan salah satu sumber kekuatan dan

kebahagiaan yang tak ternilai, selalu menemani dikala suka maupun duka setiap langkah penulis.

6. Gusti Rangga Wardhana yang memberikan dukungan, pengertian, dan motivasi. Memberikan penulis keyakinan untuk melanjutkan penelitian disaat penulis sedang tidak yakin pada kemampuan diri sendiri, memberikan kekuatan untuk terus maju, melewati berbagai tantangan dalam penyelesaian penelitian ini.

Akhir kata, penulis mengakui bahwa penulisan pada skripsi ini masih banyak kekurangan. Saya berharap semoga skripsi ini diterima sebagai amal ibadah yang tulus dan bermanfaat di sisi Allah Subhanahu Wa Ta'ala. Semoga karya ini menjadi bagian dari kontribusi yang tak terputus dalam rangka memperkuat dan mengembangkan ilmu pengetahuan, serta melaksanakan tugas sebagai manusia yang berkomitmen.

Jakarta, 08 Desember 2024

Penulis



ABSTRAK

Perkembangan teknologi telah mengubah cara pengguna menikmati musik, dengan platform seperti Spotify menyediakan jutaan lagu dari berbagai genre. Namun, dengan jumlah lagu yang sangat besar, pengelompokan lagu berdasarkan karakteristik audio menjadi tantangan tersendiri. Penelitian ini bertujuan untuk mengelompokkan lagu-lagu di Spotify berdasarkan fitur audio menggunakan algoritma K-Means dan K-Means++. Dataset yang digunakan mencakup atribut seperti danceability, energy, valence, tempo, loudness, speechiness, instrumentality, acousticness, dan liveness. Metode klusterisasi K-Means dan K-Means++ diterapkan untuk membagi lagu ke dalam tiga kluster utama, berdasarkan hasil Elbow Method. Lalu kluster dilabeli dengan Energik, Mellow, dan Ceria berdasarkan analisis hasil rata-rata pada tiap cluster. Evaluasi menggunakan Silhouette Score menunjukkan bahwa K-Means++ menghasilkan klusterisasi yang lebih optimal dibandingkan K-Means, dengan nilai 0.32246467825287023. Kluster Energik menjadi kelompok terbesar (56.5%), diikuti oleh Mellow (24.3%) dan Ceria (19.2%). Visualisasi menunjukkan pemisahan kluster yang cukup jelas, meskipun terdapat sedikit tumpang tindih antara kluster Ceria dan Energik.

Kata Kunci: Klusterisasi Lagu, Fitur Audio, Spotify, K-Means, K-Means++, Perbandingan

ABSTRACT

Technological developments have changed the way users enjoy music, with platforms such as Spotify providing millions of songs from various genres. However, with such a large number of songs, clustering songs based on audio characteristics is a challenge. This research aims to cluster songs on Spotify based on audio features using K-Means and K-Means++ algorithms. The dataset used includes attributes such as danceability, energy, valence, tempo, loudness, speechiness, instrumentalness, acousticness, and liveness. K-Means and K-Means++ clustering methods are applied to divide the songs into three main clusters, based on the Elbow Method results. Then the clusters are labeled with Energetic, Mellow, and Cheerful based on the analysis of the average results in each cluster. Evaluation using Silhouette Score shows that K-Means++ produces more optimal clustering than K-Means, with a value of 0.32246467825287023. The Energetic cluster became the largest group (56.5%), followed by Mellow (24.3%) and Cheerful (19.2%). Visualization shows a fairly clear separation of clusters, although there is a slight overlap between the Cheerful and Energetic clusters.

Keywords: Song Clustering, Audio Features, Spotify, K-Means, K-Means++, Comparison

DAFTAR ISI

KATA PENGANTAR.....	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	4
1.5 Manfaat Penelitian.....	4
BAB II	6
TINJAUAN PUSTAKA	6
2.1 Landasan Teori.....	6
2.1.1 Klasterisasi Data	6
2.1.2 Aplikasi <i>Spotify</i>	7
2.1.3 Algoritma <i>K-means</i>	7
2.1.4 Algoritma <i>K-Means++</i>	10
2.1.5 Google Colab	13
2.1.6 Bahasa Pemrograman Python	14
2.2 Studi Literatur	14
BAB III.....	21
METODOLOGI PENELITIAN	21
3.1 Waktu Penelitian	21
3.2 Alat dan Bahan.....	21
3.2.1 Perangkat Keras.....	21
3.2.2 Perangkat Lunak.....	21
3.3 Penentuan Objek Penelitian.....	22

3.4	Fokus Penelitian	22
3.5	Sumber Data	22
3.6	Desain Penelitian	23
3.6.1	Tahapan Penelitian.....	23
3.6.2	Identifikasi Masalah	23
3.6.3	Studi Literatur.....	24
3.6.4	Pengumpulan Data.....	24
3.6.5	<i>Preprocessing</i>	27
3.6.6	Pengolahan Data Menggunakan Algoritma <i>K-means</i>	30
3.6.7	Pengolahan Data Menggunakan Algoritma <i>K-Means++</i>	31
3.6.8	Evaluasi Hasil Klastering	32
BAB IV		33
HASIL DAN PEMBAHASAN		33
4.1	Perhitungan Manual	33
4.1.1	Algoritma <i>K-Means</i>	35
4.1.2	Algoritma <i>K-Means++</i>	40
4.2	Implementasi Algoritma	48
4.2.1	Algoritma <i>K-Means</i>	48
4.2.2	Algoritma <i>K-Means++</i>	55
4.3	Hasil Perbandingan	60
4.4	Tampilan Interface.....	60
BAB V		61
KESIMPULAN DAN SARAN		61
5.1	Kesimpulan	61
5.2	Saran	62
DAFTAR PUSTAKA		63
LAMPIRAN		66

DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian	23
Gambar 3. 2 Dataset <i>Spotify</i>	24
Gambar 3. 3 Tahapan Preprocessing	27
Gambar 3. 4 Hasil Cek Duplikasi pada Python	28
Gambar 3. 5 Hasil Cek Missing Values pada Python	28
Gambar 3.6 Hasil Normalisasi Data	29
Gambar 3.7 Hasil Cek Outliers pada Python	29
Gambar 3. 8 Tahapan Pengolahan Data Menggunakan Algoritma <i>K-means</i>	30
Gambar 3. 9 Tahapan Pengolahan Data Menggunakan Algoritma <i>K-means++</i>	30
Gambar 4.1 Heatmap Korelasi Fitur	48
Gambar 4.2 Grafik Elbow Method	50
Gambar 4.3 Proses Klasterisasi K-Means	50
Gambar 4.4 Rata-rata fitur tiap klaster K-Means	51
Gambar 4.5 Hasil Silhoutte Score K-Means	51
Gambar 4.6 Visualisasi PCA Berdasarkan Mood	52
Gambar 4.7 Visualisasi Distribusi Cluster Berdasarkan Mood	53
Gambar 4.8 Proses Pemilihan Centroid	55
Gambar 4.9 Proses Klasterisasi K-Means++	55
Gambar 4.10 Rata-rata fitur tiap klaster K-Means++	56
Gambar 4.11 Hasil Silhoutte Score K-Means++	56
Gambar 4.12 Visualisasi PCA Berdasarkan Mood	57
Gambar 4.13 Visualisasi Distribusi Cluster Berdasarkan Mood	58
Gambar 4.14 Tampilan Prediksi Mood Lagu	60

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	14
Tabel 3.1 Waktu Penelitian	21
Tabel 3.2 Atribut Data Spotify	25
Tabel 4.1 Sample Dataset	33
Tabel 4.2 Titik Acak Centroid	35
Tabel 4.3 Jarak Manhattan	36
Tabel 4.4 Penugasan Klaster	37
Tabel 4.5 Memperbarui Posisi Centroid	38
Tabel 4.6 Hasil Konvergen.....	39
Tabel 4.7 Inialisasi Centroid Pertama	40
Tabel 4.8 Jarak Euclidean Kuadrat.....	40
Tabel 4.9 Probabilitas Pemilihan	41
Tabel 4.10 Pemilihan Centroid Baru	42
Tabel 4.11 Titik Centroid	43
Tabel 4.12 Jarak Manhattam	44
Tabel 4.13 Penugasan Klaster	45
Tabel 4.14 Memperbarui Posisi Centroid	46
Tabel 4.15 Hasil Konvergen.....	47
Tabel 4.16 Hasil Lagu Pada Cluster 0 (Energik).....	53
Tabel 4.17 Hasil Lagu Pada Cluster 1 (Mellow).....	54
Tabel 4.18 Hasil Lagu Pada Cluster 2 (Ceria).....	54
Tabel 4.19 Hasil Lagu Pada Cluster 0 (Energik).....	58
Tabel 4.20 Hasil Lagu Pada Cluster 1 (Ceria).....	59
Tabel 4.21 Hasil Lagu Pada Cluster 2 (Mellow).....	59