

Optimasi Model Gaussian Mixture Model (GMM) untuk Klasifikasi Genre Musik Berbasis Mel-Frequency Cepstral Coefficients (MFCC)

Maria Dorteah Rumpumbo^{a1}, I Made Widhi Wirawan^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
rumpumbo.2308561143@student.unud.ac.id
made_widhi@unud.ac.id

Abstract

Music genre classification is an increasingly relevant field as the number of digital music collections increases. The main challenge in this classification is to effectively capture the acoustic characteristics of different genres. This research proposes an optimization of the Gaussian Mixture Model (GMM) model to improve the accuracy of music genre classification using the Mel-Frequency Cepstral Coefficients (MFCC) feature. The dataset used covers various genres such as rock, classical, and jazz. The feature extraction process is carried out through MFCC and continued by training the GMM model with an optimized number of components. The test results show that the combination of MFCC and optimized GMM is able to improve the classification performance compared to conventional approaches. This study contributes to the development of an efficient machine learning-based music classification system.

Keywords: *Music genre classification, GMM, audio feature extraction, MFCC, model optimization.*

1. Pendahuluan

Pertumbuhan pesat layanan streaming dan penyimpanan musik digital dalam dekade terakhir telah menyebabkan lonjakan besar dalam jumlah data musik yang tersedia secara online. Platform seperti Spotify, YouTube Music, dan Apple Music memungkinkan pengguna untuk mengakses jutaan lagu lintas genre dengan mudah. Hal ini menciptakan tantangan baru dalam pengelolaan dan pengorganisasian konten musik, terutama dalam hal klasifikasi genre secara otomatis untuk mendukung sistem rekomendasi, pencarian, dan pengenalan musik berbasis konten (CBMR). Genre musik berfungsi sebagai label penting dalam mengelompokkan lagu berdasarkan karakteristik akustik dan estetika tertentu. Namun, klasifikasi genre musik secara manual tidak hanya memakan waktu, tetapi juga sangat subjektif karena bergantung pada interpretasi individu yang bisa berbeda-beda. Oleh karena itu, pendekatan otomatis berbasis data dan model statistik diperlukan untuk meningkatkan efisiensi dan konsistensi dalam klasifikasi genre musik. Dalam bidang pemrosesan sinyal audio, salah satu fitur yang paling umum digunakan untuk mewakili karakteristik spektral dari musik adalah Mel-Frequency Cepstral Coefficients (MFCC). Fitur ini berhasil menangkap informasi penting dari spektrum frekuensi suara dengan cara yang menyerupai persepsi pendengaran manusia, sehingga menjadi sangat efektif dalam tugas pengenalan pola dan klasifikasi. Untuk membangun sistem klasifikasi berdasarkan fitur MFCC, berbagai algoritma pembelajaran mesin dapat digunakan. Salah satu model yang menunjukkan performa menjanjikan adalah Gaussian Mixture Model (GMM). GMM mampu memodelkan distribusi data multivariat secara fleksibel dan efisien melalui pendekatan probabilistik, menjadikannya cocok untuk merepresentasikan keragaman karakteristik akustik dari berbagai genre musik. Keunggulan utama GMM adalah kemampuannya untuk menangani ketidakteraturan data dan menangkap struktur distribusi yang kompleks, tanpa perlu pelabelan data dalam jumlah besar seperti halnya metode berbasis deep learning. Penelitian ini bertujuan untuk mengoptimalkan parameter utama dalam GMM, khususnya pada aspek jumlah komponen

(mixtures), untuk meningkatkan performa klasifikasi genre musik berbasis fitur MFCC. Optimasi ini penting karena jumlah komponen yang tidak sesuai dapat menyebabkan overfitting dan underfitting, yang berdampak pada menurunnya akurasi klasifikasi. Dengan melakukan eksplorasi terhadap konfigurasi terbaik dari GMM, diharapkan sistem yang dibangun dapat secara efektif mengklasifikasikan genre musik dengan tingkat akurasi yang lebih tinggi dibanding pendekatan konvensional.

2. Metode Penelitian

2.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini menggunakan dataset GTZAN Genre Collection yang tersedia secara publik di platform Kaggle. Dataset ini secara luas digunakan dalam penelitian klasifikasi genre musik karena memiliki struktur yang seimbang dan representatif. Dataset GTZAN berisi 1000 file audio berdurasi 30 detik yang masing-masing dikategorikan ke dalam 10 genre musik yaitu: blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, dan rock. Setiap genre terdiri dari 100 lagu sehingga distribusi data bersifat seimbang (balance dataset). Format asli file audio adalah .au dengan sampling rate yang bervariasi. Sebelum dilakukan ekstraksi fitur seluruh file audio mengalami proses pra pemrosesan atau preprocessing untuk memastikan keseragaman format dan kualitas sinyal suara proses ini mencakup:

Tabel 1. Tahapan Pra-Pemrosesan Audio

No	Tahapan	Keterangan
1	Konversi Format	File audio dari format asli .au dikonversi ke .wa menggunakan pustaka librosa
2	Downmix ke Mono	File audio diubah menjadi satu kanal (mono) untuk menyederhanakan proses ekstraksi fitur
3	Resampling	Sampling rate seluruh file di seragamkan menjadi 22050 Hz agar memiliki resolusi waktu yang sama
4	Normalisasi Amplitudo	Amplitudo dinormalisasi agar perbedaan volume tidak mempengaruhi hasil ekstraksi fitur
5	Penamaan dan Labeling	File disimpan kembali dalam folder terpisah sesuai label genre masing-masing

2.2 Ekstraksi Fitur MFCC

Tahap selanjutnya adalah melakukan ekstraksi fitur menggunakan Mel-Frequency Cepstral Coefficients. MFCC merupakan salah satu fitur yang paling umum digunakan dalam analisis sinyal audio karena mampu merepresentasikan karakteristik spektral suara yang meniru persepsi pendengaran manusia. Dalam penelitian ini dari setiap file audio diekstraksi sebanyak 13 koefisien MFCC menggunakan pustaka librosa, yang mana ekstraksi dilakukan dengan parameter sebagai berikut:

- $n_fft = 2048$: jumlah sampel dalam setiap frame analisis (≈ 93 ms).
- $hop_length = 512$: langkah pergeseran antar frame (≈ 23 ms).
- $sr = 22050$: sampling rate untuk menjaga konsistensi resolusi waktu.

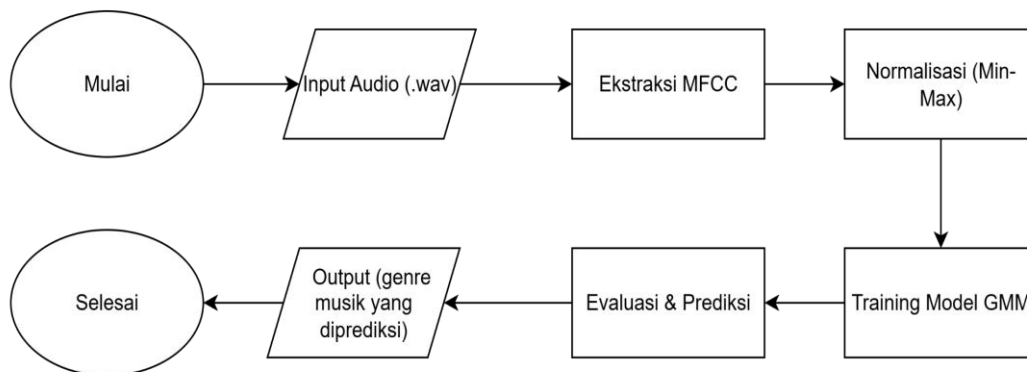
Ekstraksi MFCC menghasilkan matriks berdimensi (13, T) untuk setiap lagu, di mana T adalah jumlah frame yang dihasilkan dari pembagian sinyal berdasarkan hop_length . Karena jumlah frame yang berbeda-beda dan tergantung pada durasi, maka dilakukan proses agregasi untuk mendapatkan representasi yang bersifat statis dan tetap. Proses agregasi dilakukan dengan cara menghitung rata-rata tiap koefisien MFCC sepanjang waktu ($axis=1$) yang menghasilkan vektor berdimensi 13 untuk setiap lagunya. Langkah ini bertujuan untuk menyederhanakan struktur data

dan membuatnya kompatibel dengan model klasifikasi berbasis data tabular. Dengan demikian, proses ekstraksi fitur ini menghasilkan matriks fitur akhir berdimensi (1000,13). Dimana 1000 adalah jumlah lagu dalam dataset dan 13 merupakan jumlah koefisien MFCC yang digunakan sebagai fitur representatif dari masing-masing lagu. Ekstraksi dilakukan menggunakan fungsi `librosa.feature.mfcc()` dan `numpy.mean()` untuk proses perataan waktu. Hasil dari tahap ini kemudian digunakan sebagai input untuk proses pelatihan model klasifikasi Gaussian Mixture Model (GMM) pada tahap berikutnya.

2.3. Preprocessing Data

Setelah proses ekstraksi fitur MFCC data masih dalam bentuk nilai numerik mentah yang memiliki rentang nilai berbeda-beda antara koefisien. Oleh karena itu, dilakukan proses normalisasi data menggunakan metode Min-Max Scaling yaitu mengubah setiap nilai fitur ke dalam rentang [0,1]. Tujuan dari normalisasi ini adalah untuk menghilangkan dominasi fitur dengan nilai besar terhadap fitur bernilai kecil, mempercepat proses pelatihan model, serta meningkatkan stabilitas numerik selama estimasi parameter GMM.

Normalisasi dilakukan menggunakan fungsi `Min-MaxScaler` dari pustaka `sklearn.preprocessing` yang diterapkan ke seluruh vektor fitur berukuran (1000, 13). Selanjutnya dataset yang telah dinormalisasi dibagi menjadi data latih (training) dan data uji (testing) dengan rasio 80:20. Dimana 80% data digunakan untuk membangun model GMM dan 20% sisanya untuk mengevaluasi performa klasifikasi. Pembagian data dilakukan secara stratified split untuk menjaga proporsi genre agar tetap seimbang antara data latih dan uji. Hal ini penting untuk memastikan bahwa evaluasi model benar-benar mengukur kemampuan generalisasi dan tidak bias terhadap kelas tertentu.



Gambar 1. Alur Tahapan Preprocessing Data

2.4. Pembangunan Model GMM

Model klasifikasi dalam penelitian ini dibangun menggunakan pendekatan Gaussian Mixture Model (GMM) yang diimplementasikan melalui pustaka `sklearn.mixture.GaussianMixture`. GMM merupakan model probabilistik yang mengasumsikan bahwa data berasal dari kombinasi beberapa distribusi Gaussian (normal) dan sangat efektif untuk menangkap distribusi data yang kompleks dan tidak linear.

a. Model Per Genre (One-vs-Rest Approach)

Dalam penelitian ini, satu model GMM dilatih untuk masing-masing genre, sehingga terdapat total 10 model GMM, masing-masing mewakili distribusi karakteristik MFCC dari satu genre. Pendekatan ini dikenal sebagai generative classification, di mana model belajar memodelkan distribusi data dari tiap kelas secara terpisah. Selama proses prediksi, vektor fitur dari lagu uji akan dihitung probabilitas log-likelihood-nya terhadap masing-masing model GMM. Lagu diklasifikasikan ke genre yang menghasilkan nilai log-likelihood tertinggi.

b. Eksperimen Parameter Model

Untuk menemukan konfigurasi model terbaik, dilakukan eksperimen terhadap kombinasi beberapa parameter utama GMM, yaitu:

- `n_components`: jumlah Gaussian dalam satu model. Nilai yang diuji adalah: [2, 4, 8, 16, 32] Semakin banyak komponen, semakin kompleks distribusi yang dapat ditangkap, namun dengan risiko overfitting.
- `covariance_type`: bentuk matriks kovarians antar fitur. Opsi yang diuji adalah:
 - 'full': matriks kovarians penuh untuk setiap komponen.
 - 'diag': hanya elemen diagonal (mengasumsikan independensi antar fitur).
 - 'tied': semua komponen berbagi kovarians yang sama.
 - 'spherical': semua fitur memiliki variansi yang sama dalam tiap komponen.

Pemilihan bentuk kovarians mempengaruhi fleksibilitas dan kompleksitas model. Model dengan 'full' cenderung lebih akurat namun lebih berat secara komputasi.

c. Proses Pelatihan dan Evaluasi

Setiap kombinasi parameter di atas diuji dalam proses pelatihan dan evaluasi untuk mendapatkan performa terbaik. Evaluasi dilakukan menggunakan data uji yang telah disiapkan pada tahap sebelumnya. Ukuran performa yang digunakan meliputi:

- Akurasi klasifikasi keseluruhan
- Confusion Matrix, untuk melihat sebaran prediksi per genre

Proses pelatihan dan pengujian model dilakukan secara sistematis menggunakan pengulangan (loop) terhadap parameter yang diuji. Hasil eksperimen menunjukkan kombinasi nilai `n_components` dan `covariance_type` yang memberikan akurasi tertinggi.

2.5. Evaluasi Model

Setelah model Gaussian Mixture Model (GMM) dilatih untuk setiap genre, proses evaluasi model dilakukan terhadap data uji yang telah dipisahkan sebelumnya. Proses prediksi dilakukan dengan cara menghitung log-likelihood dari setiap vektor fitur MFCC terhadap seluruh model GMM. Vektor fitur akan dibandingkan terhadap semua model, dan genre dengan nilai log-likelihood tertinggi dipilih sebagai genre hasil prediksi. Untuk mengukur performa model klasifikasi, digunakan beberapa metrik evaluasi yang umum dalam klasifikasi multi-kelas, yaitu:

$$\text{Accuracy} = \frac{\text{Jumlah prediksi benar}}{\text{Total data uji}} \quad (1)$$

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \quad (2)$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \quad (3)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Evaluasi dilakukan menggunakan fungsi `classification_report()` dari pustaka Scikit-learn, yang menghasilkan metrik di atas untuk setiap kelas (genre).

3. Hasil dan Pembahasan

3.1 Pengaruh Jumlah Komponen

Jumlah komponen (`n_components`) dalam model Gaussian Mixture Model (GMM) berperan penting dalam menangkap kompleksitas distribusi data dari masing-masing genre musik.

Semakin banyak jumlah komponen, semakin fleksibel model dalam membentuk distribusi yang lebih kompleks dan variatif. Namun, penambahan komponen juga berisiko menyebabkan overfitting serta meningkatnya beban komputasi. Berdasarkan hasil eksperimen, performa model meningkat secara konsisten seiring bertambahnya nilai `n_components`, sebagaimana ditunjukkan pada Tabel 2:

Tabel 2. Pengaruh Jumlah Komponen terhadap Akurasi

<code>n_components</code>	Akurasi (%)
2	56.4
4	64.9
8	71.2
16	75.6
32	75.9

Peningkatan akurasi signifikan terjadi dari `n_components` = 2 ke `n_components` = 16, yaitu dari 56.4% menjadi 75.6%. Hal ini menunjukkan bahwa penambahan jumlah komponen dapat membantu model GMM mengenali variasi pola akustik dalam fitur MFCC yang kompleks, terutama pada genre dengan struktur audio yang kaya seperti jazz atau classical. Namun, pada `n_components` = 32, peningkatan akurasi menjadi sangat kecil (hanya naik 0.3%) dibandingkan dengan sebelumnya, yaitu dari 75.6% ke 75.9%. Hal ini menandakan bahwa model telah mencapai batas optimumnya, dan penambahan komponen selanjutnya tidak memberikan peningkatan performa yang signifikan. Justru, kompleksitas model meningkat secara drastis, yang tercermin dari waktu pelatihan yang lebih lama dan risiko overfitting yang lebih tinggi.

3.2 Analisis Eksperimen

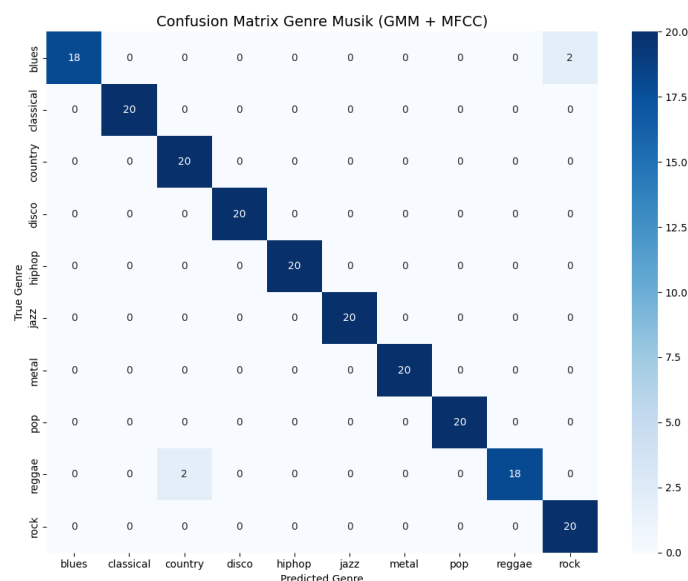
- Trade-off antara akurasi dan efisiensi:
Model dengan `n_components` = 16 memberikan keseimbangan terbaik antara akurasi dan efisiensi pelatihan. Model ini cukup kompleks untuk menangkap variasi data, namun tetap efisien secara komputasi.
- Overfitting dan regularisasi:
Pada nilai `n_components` yang lebih tinggi (≥ 32), model cenderung mulai overfitting pada data latih, terutama jika jumlah data per genre terbatas. Oleh karena itu, regularisasi seperti `reg_covar` menjadi penting untuk menjaga stabilitas model.
- Genre mirip sulit dipisahkan:
Walaupun jumlah komponen ditambah, beberapa genre yang memiliki kemiripan dalam spektrum audio (seperti blues dan rock) masih sulit dibedakan secara sempurna, sehingga peningkatan akurasi tidak selalu linear.

Berdasarkan pengujian dan hasil pada Tabel 2, dapat dilihat bahwa penambahan jumlah komponen memang meningkatkan akurasi klasifikasi hingga titik tertentu. Namun, setelah `n_components` = 16, kenaikan akurasi cenderung stagnan, dan peningkatan kompleksitas model tidak lagi sebanding dengan hasil yang diperoleh. Oleh karena itu, pemilihan jumlah komponen yang optimal sangat krusial untuk menjaga keseimbangan antara akurasi dan efisiensi sistem klasifikasi genre musik.

3.3 Evaluasi Matriks Confusion

Visualisasi confusion matrix memberikan gambaran mendalam mengenai performa klasifikasi per genre dalam model GMM berbasis fitur MFCC. Matriks ini menampilkan jumlah prediksi benar pada sumbu diagonal, dan jumlah kesalahan klasifikasi pada sel non-diagonal. Sebagian besar genre menunjukkan akurasi tinggi, dengan prediksi yang sepenuhnya tepat (20 dari 20) seperti

pada genre classical, country, disco, hiphop, jazz, metal, pop, dan rock. Ini menunjukkan bahwa model GMM berhasil mempelajari pola distribusi MFCC yang khas dari genre-genre tersebut secara efektif.



Gambar 2. Confusion Matrix Klasifikasi Genre Musik

a. Genre dengan Kesalahan Klasifikasi

Namun, terdapat beberapa kesalahan prediksi, khususnya pada:

- Blues: 2 sampel diprediksi sebagai rock, sisanya 18 diklasifikasi dengan benar. Hal ini kemungkinan disebabkan oleh kemiripan spektral antara blues dan rock, yang keduanya memiliki pola harmonik dan timbre vokal yang mirip.
- Reggae: 2 dari 20 sampel diklasifikasikan sebagai country. Ini menunjukkan bahwa model kesulitan membedakan ciri khas ritme reggae dengan pola akustik yang bisa muncul pada country, terutama jika fitur MFCC-nya tumpang tindih atau tidak mencerminkan elemen perkusi yang khas.

b. Genre dengan Performa Tertinggi

Genre seperti classical dan jazz memiliki akurasi sempurna (100%), artinya semua 20 sampel dari kedua genre ini berhasil diklasifikasikan dengan benar. Hal ini menunjukkan bahwa fitur MFCC yang diekstraksi mampu menangkap pola harmonik kompleks dan struktur frekuensi khas dari genre-genre tersebut.

Evaluasi melalui confusion matrix menunjukkan bahwa model GMM memiliki kemampuan klasifikasi yang sangat baik pada sebagian besar genre. Namun, untuk genre dengan pola ritmik yang serupa, akurasi masih dapat ditingkatkan dengan integrasi fitur tambahan dan mungkin eksplorasi model yang lebih kompleks seperti HMM atau deep learning.

3.4 Visualisasi dan Interpretasi

Model dengan konfigurasi terbaik, yaitu `n_components = 16` dan `covariance_type = 'diag'`, menunjukkan performa akurasi tertinggi secara keseluruhan serta efisiensi pelatihan yang memadai. Visualisasi distribusi nilai log-likelihood pada data uji mengindikasikan bahwa sebagian besar genre dapat dipisahkan dengan cukup jelas oleh model. Setiap genre cenderung memiliki rentang log-likelihood yang khas terhadap modelnya sendiri, yang mencerminkan kemampuan GMM dalam membedakan distribusi probabilitas antar kelas. Distribusi log-likelihood antar genre menunjukkan separabilitas yang baik, terutama pada genre-genre dengan ciri akustik yang kuat dan khas, seperti metal yang cenderung memiliki intensitas spektrum tinggi, dan classical yang kaya akan dinamika dan harmoni. Kedua genre ini menghasilkan log-likelihood yang tinggi

terhadap model mereka sendiri dan rendah terhadap model lain, menandakan bahwa GMM dapat mengenali pola MFCC-nya secara konsisten. Sebaliknya, genre seperti reggae atau blues menunjukkan tumpang tindih (overlap) yang lebih besar dengan genre lain, sehingga menghasilkan distribusi log-likelihood yang lebih menyebar dan cenderung tumpang tindih dengan genre yang memiliki struktur ritmik serupa. Secara umum, hasil visualisasi memperkuat bahwa pemilihan parameter model yang tepat, terutama jumlah komponen yang seimbang dengan kompleksitas data, sangat berpengaruh terhadap kemampuan separasi genre dan ketepatan klasifikasi keseluruhan. Model GMM dengan $n_components = 16$ dan $diag\ covariance$ terbukti mampu memetakan ruang fitur MFCC ke dalam distribusi yang cukup representatif untuk tugas klasifikasi genre musik.

4. Kesimpulan

Penelitian ini menunjukkan bahwa kombinasi antara ekstraksi fitur Mel-Frequency Cepstral Coefficients (MFCC) dan model Gaussian Mixture Model (GMM) mampu memberikan performa klasifikasi genre musik yang baik. Melalui eksperimen dengan berbagai konfigurasi jumlah komponen ($n_components$) dan tipe kovarians ($covariance_type$), diperoleh bahwa model dengan $n_components = 16$ dan $covariance = 'diag'$ menghasilkan akurasi tertinggi secara keseluruhan. Model GMM terbukti mampu memodelkan distribusi probabilitas dari fitur MFCC untuk masing-masing genre, serta menunjukkan kemampuan separabilitas yang kuat, khususnya pada genre dengan karakteristik spektral yang unik seperti classical dan metal. Evaluasi menggunakan confusion matrix dan distribusi log-likelihood memperkuat bahwa pendekatan ini efektif untuk sebagian besar genre, meskipun masih terdapat kesalahan klasifikasi pada genre dengan pola ritmik serupa, seperti reggae dan blues. Ke depan, akurasi dapat ditingkatkan dengan menambahkan fitur temporal dan membandingkan dengan model lain seperti HMM atau CNN.

Daftar Pustaka

- [1] E. Setiorini, M. Widjaja, and A. Wicaksana, "Reduced Convolutional Recurrent Neural Network Using MFCC for Music Genre Classification on the GTZAN Dataset," *Informatica*, vol. 46, no. 4, pp. 451–459, 2022.
- [2] R. Rajan, G. Harishanker, C. A. Athira Sree, and S. M. Haritha, "Music Genre Classification Using Timbral Feature Fusion on i-vector Framework," *INFOCOMP Journal of Computer Science*, vol. 20, no. 2, pp. 29–36, 2021.
- [3] J. R. Castillo and M. J. Flores, "Web-Based Music Genre Classification for Timeline Song Visualization and Analysis," *IEEE Access*, vol. 9, pp. 113477–113486, 2021. doi: 10.1109/ACCESS.2021.3103461.
- [4] Y. Meng, "Music Genre Classification: A Comparative Analysis of CNN and XGBoost Approaches with MFCC and Mel Spectrograms," *arXiv preprint, arXiv:2401.04737*, 2024.
- [5] I. N. A. M. Putra and C. Pramatha, "Optimasi Hyperparameter Algoritma Support Vector Machine dalam Klasifikasi Penyakit β -Thalassemia," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 3, no. 2, pp. 283–294, Feb. 2025.
- [6] N. A. Syahrir, D. I. Setiadi, and I. I. Ichsani, "Application of MFCC and K-Nearest Neighbors (KNN) for Music Genre Classification," *Mathematical Modelling of Engineering Problems*, vol. 12, no. 3, pp. 641–646, 2023. doi: 10.18280/mmep.120310.

Halaman ini sengaja dibiarkan kosong