Klasifikasi Chord Musik Menggunakan Gabungan Fitur Domain Waktu, Frekuensi, dan MFCC

Ni Made Anita Widyastini[a1], I Gede Arta Wibawa[a2], I Putu Satwika[a3]

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana, Bali
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]widyastini.2308561130@student.unud.ac.id
[2]gede.arta@unud.ac.id
[3]satwika@unud.ac.id

Abstract

This study presents a chord classification system that combines audio features from three different domains: time, frequency, and Mel-Frequency Cepstral Coefficients (MFCC). The purpose is to improve the accuracy of identifying musical chords from audio signals, which often contain overlapping sounds and instrument variations. The dataset used consists of major and minor chord audio clips sourced from Kaggle. Each audio file undergoes preprocessing, including resampling and signal normalization, followed by feature extraction from the three domains. The extracted features are then merged into a single vector and classified using the Random Forest algorithm. The model is evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Results show that the model performs well in detecting major chords (F1-score 0.83), but has lower recall for minor chords (F1-score 0.68). The overall accuracy is 77%, indicating that combining features from multiple domains enhances classification performance. This method shows potential for future development in audio signal analysis and music recognition systems.

Keywords: *Audio Classification, Chord Recognition, Feature Extraction, MFCC, Machine Learning*

1. Pendahuluan

Musik tidak hanya menjadi sarana hiburan, tetapi juga menyimpan informasi kompleks yang bisa dianalisis secara digital. Salah satu elemen penting dalam struktur musik adalah chord, yaitu kumpulan nada yang dimainkan secara bersamaan dan membentuk harmoni. Seiring berkembangnya teknologi, analisis terhadap chord menjadi semakin relevan, terutama dalam bidang seperti edukasi musik berbasis aplikasi, pelabelan otomatis lagu, hingga sistem pengenalan musik. Namun, mengenali jenis chord dari sinyal audio secara otomatis masih menjadi tantangan karena sifat sinyal musik yang dinamis, memiliki berbagai instrumen, serta sering kali mengandung tumpang tindih suara.

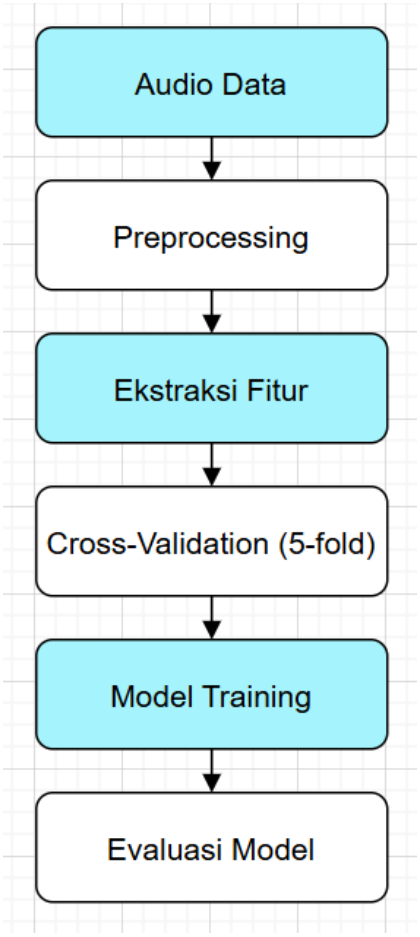
Selama ini, berbagai pendekatan telah digunakan untuk mengklasifikasikan sinyal musik, salah satunya adalah ekstraksi fitur dari domain tertentu seperti waktu, frekuensi, atau waktu-frekuensi[4]. Pendekatan berbasis Mel Frequency Cepstral Coefficients (MFCC) misalnya, cukup umum digunakan karena mampu menangkap karakteristik spektral yang menyerupai persepsi manusia terhadap suara [2]. Di sisi lain, fitur dari domain waktu seperti *Zero Crossing Rate* dan domain frekuensi seperti Spectral Centroid juga memiliki kontribusi penting dalam memahami struktur sinyal audio. Sayangnya, penggunaan fitur dari satu domain saja seringkali tidak cukup untuk menghasilkan klasifikasi yang akurat dalam kasus chord yang tumpang tindih atau kompleks. Melihat keterbatasan tersebut, penelitian ini menawarkan solusi berupa penggabungan fitur dari tiga domain sekaligus: waktu, frekuensi, dan MFCC. Tujuannya adalah untuk menciptakan representasi fitur yang lebih kaya dan informatif, sehingga dapat meningkatkan performa sistem klasifikasi chord musik. Pendekatan ini diharapkan mampu menangkap berbagai aspek dari sinyal audio yang sebelumnya kurang tergali jika hanya menggunakan satu jenis fitur saja.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Beberapa studi sebelumnya telah mencoba kombinasi fitur audio untuk tujuan klasifikasi musik, di mana mereka menggabungkan fitur waktu dan frekuensi untuk klasifikasi instrumen musik dan berhasil meningkatkan akurasi secara signifikan[3]. Namun, masih jarang ditemukan studi yang secara eksplisit mengintegrasikan ketiga domain tersebut secara bersamaan dalam konteks klasifikasi chord. Oleh karena itu, penelitian ini memiliki kontribusi pada pengembangan metode yang lebih lengkap dalam pemrosesan dan klasifikasi data audio musik.

2. Metode Penelitian

Penelitian ini dilakukan untuk mengembangkan sistem klasifikasi chord musik berbasis fitur audio yang diambil dari domain waktu, frekuensi, dan waktu-frekuensi. Prosesnya dimulai dari pengumpulan data hingga evaluasi model. Proses ini disusun secara logis agar hasil yang diperoleh sesuai dengan tujuan penelitian.



Gambar 1. 6 Tahapan Proses

2.1. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu *dataset* file audio chord musik yang diperoleh secara daring pada kaggle. Struktur *dataset* terdiri dari dua folder utama, yaitu *major* dan *minor* yang masing-masing berisi subfolder berdasarkan nama chord. Seperti, di dalam folder *major* terdapat folder C, D, E, dan sebagainya. Setiap folder di dalam *dataset* berisi file audio dalam format .wav dengan durasi pendek (sekitar 2–5 detik), di mana masing-masing file mewakili satu jenis chord. Cara pengambilan data dilakukan dengan mengunduh dan menyusun ulang file audio tersebut ke dalam direktori lokal, lalu dikelompokkan berdasarkan nama label chord untuk memudahkan pelabelan saat proses ekstraksi fitur.

2.2. *Preprocessing*

Sebelum dilakukan klasifikasi, data perlu melalui beberapa tahapan *preprocessing*, antara lain :

- *Resampling* : Semua audio disamakan sampling rate-nya saat dibaca menggunakan librosa agar fitur yang diekstrak memiliki konsistensi.
- Normalisasi panjang sinyal : Jika terdapat perbedaan panjang sinyal audio, sinyal dipotong agar tetap sama.
- Ekstraksi fitur menggunakan MFCC (13 koefisien) dari domain waktu-frekuensi, Chroma STFT (12 fitur) dari domain frekuensi, dan Spectral Centroid (1 fitur) dari domain frekuensi. Semua Ekstraksi fitur kemudian dirata-ratakan (*mean*) dan digabungkan menjadi satu vektor berdimensi 26 fitur.

2.3. Pembagian Data

Setelah semua fitur diekstrak, data dibagi menjadi dua bagian yaitu data latih dan data uji. Pembagian dilakukan menggunakan teknik k-fold *cross validation* dengan nilai k = 5, untuk menghindari *overfitting* dan memberikan penilaian yang lebih umum terhadap performa model. Dalam setiap iterasi, data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian, dan proses ini diulang sebanyak lima kali dengan kombinasi data yang berbeda.

2.4. Pelatihan Model

Model klasifikasi yang digunakan dalam penelitian ini adalah Random Forest Classifier. Model ini dipilih karena mampu menangani fitur yang kompleks [5], tidak sensitif terhadap normalisasi, dan cukup andal dalam data dengan banyak dimensi. Model ini dilatih menggunakan parameter class_weight='balanced' untuk mengatasi kemungkinan jumlah data antar chord yang tidak merata.

2.5. Evaluasi Model

Untuk mengukur seberapa baik model dalam mengklasifikasikan chord dari data audio, digunakan beberapa matriks evaluasi yang umum dalam bidang mesin, khususnya klasifikasi. Metriks ini memberikan gambaran menyeluruh tentang performa model terhadap data uji yang belum pernah dilihat sebelumnya. Berikut adalah penjelasan dari masing-masing ukuran kinerja:

- **Akurasi (Accuracy) :**
  Akurasi menunjukkan seberapa sering model berhasil menebak label chord dengan tepat.
- **Presisi (Precision) :**
  Mengukur ketepatan model saat memprediksi suatu label tertentu. Presisi tinggi menunjukkan bahwa ketika model memprediksi sebuah chord, kemungkinan besar prediksi tersebut benar.
- **Recall :**
  Mengukur seberapa banyak label yang sebenarnya benar berhasil dikenali oleh model.
- **F1-Score :**
  Metrik ini sering digunakan sebagai ukuran utama dalam klasifikasi, terutama ketika terdapat ketidakseimbangan jumlah data antar kelas[**1**].
- ***Confusion matrix* :**
  Menampilkan perbandingan antara label asli dan prediksi model dalam bentuk tabel. *Confusion matrix* memberikan informasi rinci tentang kesalahan dan keberhasilan prediksi pada masing-masing kelas chord[7]. Visualisasi ini juga mempermudah dalam mengevaluasi model, terutama jika jumlah kelas cukup banyak.

3. Hasil dan Diskusi

Implementasi pipeline pada penelitian klasifikasi chord musik ini mencakup beberapa tahapan sistematis, mulai dari pengambilan data hingga prediksi akhir. Proses dimulai dengan membaca *dataset* yang terdiri dari berbagai folder, di mana setiap folder mewakili satu jenis chord. Setiap

p-ISSN: 2986-3929
e-ISSN: 3032-1948

file audio di dalam folder tersebut kemudian diekstraksi fiturnya menggunakan tiga jenis fitur utama: MFCC, chroma, dan spectral centroid. MFCC digunakan untuk menangkap karakteristik spektrum suara berdasarkan persepsi pendengaran manusia, sementara chroma menangkap informasi terkait pitch atau nada, dan spectral centroid merepresentasikan kecerahan atau pusat frekuensi dari sinyal. Ketiga fitur ini kemudian digabungkan menjadi satu vektor fitur untuk setiap file. Setelah proses ekstraksi selesai, data dibagi menjadi data pelatihan dan data pengujian menggunakan metode pembagian 80:20. Model klasifikasi yang digunakan adalah Random Forest dengan pengaturan class_weight='balanced' untuk mengatasi kemungkinan distribusi kelas yang tidak merata. Model ini kemudian dilatih menggunakan data pelatihan, dan hasil prediksi diuji pada data pengujian.

3.1. Hasil Pelatihan dan Evaluasi Model

	precision	recall	f1-score	support
Major	0.74	0.94	0.83	98
Minor	0.87	0.55	0.68	74
accuracy			0.77	172
macro avg	0.80	0.75	0.75	172
weighted avg	0.79	0.77	0.76	172

Gambar 2. Hasil Pelatihan

Setelah proses pelatihan model selesai, dilakukan evaluasi untuk mengetahui seberapa baik model dalam mengenali jenis chord. Dari hasil pengujian, model menunjukkan performa yang cukup baik terutama pada kelas *Major*. Nilai precision untuk kelas ini mencapai 0,74, yang berarti bahwa sebagian besar prediksi chord *major* yang dilakukan oleh model sudah benar. Bahkan, recall-nya tergolong tinggi yaitu sebesar 0,94, menunjukkan bahwa hampir semua data dengan label *major* berhasil dikenali. Kombinasi ini menghasilkan nilai F1-score sebesar 0,83, yang mencerminkan keseimbangan yang cukup solid antara ketepatan dan kelengkapan prediksi model terhadap kelas *major*.

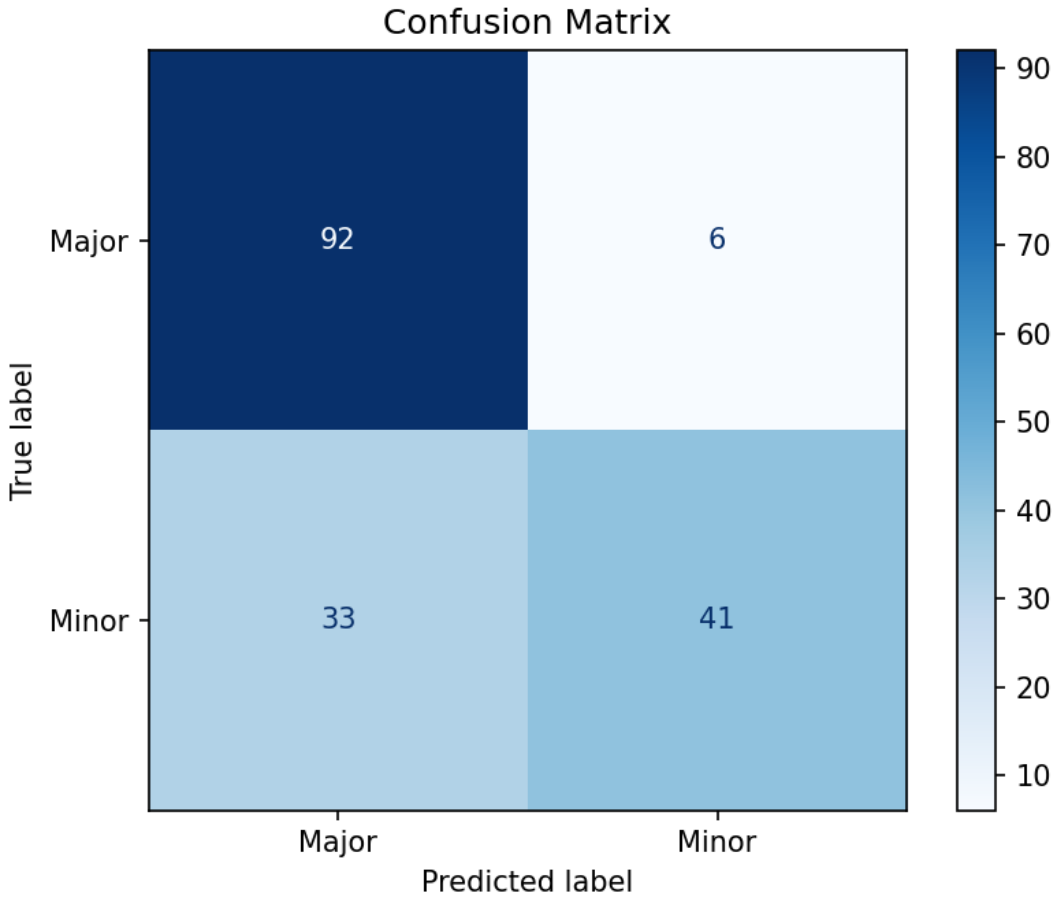
Sebaliknya, untuk kelas *Minor*, meskipun precision-nya justru lebih tinggi, yakni sebesar 0,87, tetapi recall-nya jauh lebih rendah, yaitu hanya 0,55. Artinya, model memang cukup yakin saat memprediksi chord *minor*, namun masih sering keliru dalam mendeteksi data *minor* yang sebenarnya. Hal ini berdampak pada F1-score *minor* yang hanya mencapai 0,68, menandakan bahwa performa model dalam mengenali chord *minor* masih kurang konsisten jika dibandingkan dengan kelas *major*.

Secara keseluruhan, akurasi model tercatat sebesar 0,77 atau 77%, yang berarti sebagian besar prediksi sudah tepat. Nilai rata-rata F1-score tanpa mempertimbangkan jumlah data di tiap kelas (macro average) adalah 0,75, sementara nilai rata-rata berbobot (weighted average) yang memperhitungkan jumlah data per kelas sedikit lebih tinggi, yaitu 0,76. Hal ini menunjukkan bahwa model sedikit lebih optimal pada kelas yang memiliki data lebih banyak, namun secara umum sudah cukup seimbang.

3.2 Visualisasi dan Interpretasi Hasil

Confusion matrix digunakan untuk memvisualisasikan kinerja model dalam membedakan chord *Major* dan *Minor*. Dari hasil yang ditampilkan, model mampu mengenali chord *Major* dengan sangat baik, terbukti dari 92 data yang diklasifikasikan dengan benar dan hanya 6 yang salah. Sebaliknya, pada chord *Minor*, model hanya mengklasifikasikan 41 data dengan tepat, sedangkan 33 lainnya keliru diprediksi sebagai *Major*.

p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 3. *Confusion matrix*

Warna biru tua yang dominan pada diagonal menunjukkan akurasi yang tinggi, khususnya pada kelas *Major*. Namun, warna yang lebih terang di baris *Minor* mengindikasikan adanya kebingungan model dalam mengenali chord ini. Hasil ini sejalan dengan metrik evaluasi sebelumnya, di mana recall untuk *Minor* masih tergolong rendah. Untuk meningkatkan akurasi, perlu dilakukan perbaikan seperti penambahan data *Minor* atau peningkatan fitur yang lebih mampu membedakan karakteristik antar chord[6].

4. Kesimpulan

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa penggabungan fitur dari domain waktu, domain frekuensi, dan MFCC mampu meningkatkan performa klasifikasi chord musik secara signifikan. Pendekatan ini berhasil membentuk representasi fitur audio yang lebih lengkap dibanding hanya menggunakan satu jenis fitur saja.

Melalui tahapan penelitian yang sistematis, mulai dari pengumpulan dan *preprocessing* data, hingga ekstraksi fitur serta pelatihan model, peneliti mampu membangun sistem klasifikasi yang cukup andal. Model Random Forest yang digunakan menunjukkan performa yang baik, terutama pada jenis chord *major*, dengan nilai F1-score mencapai 0,83. Pengujian dilakukan dengan teknik k-fold *cross validation* untuk memastikan hasil yang diperoleh tidak hanya berlaku pada satu subset data saja. Hasil evaluasi menunjukkan bahwa model mencapai akurasi keseluruhan sebesar 77%, yang berarti mayoritas prediksi yang dihasilkan sudah sesuai dengan label aslinya. Visualisasi melalui *confusion matrix* juga menguatkan hasil evaluasi, di mana chord *major* diklasifikasikan dengan sangat baik, sedangkan chord *minor* masih perlu perhatian lebih karena tingkat recall-nya masih rendah. Hal ini menunjukkan bahwa masih ada peluang untuk

p-ISSN: 2986-3929
e-ISSN: 3032-1948

meningkatkan kualitas klasifikasi pada kelas *minor*, misalnya dengan penambahan data atau penyempurnaan fitur[8]. Dengan demikian, dapat disimpulkan bahwa pendekatan penggabungan fitur dari tiga domain dalam klasifikasi chord musik memberikan hasil yang menjanjikan dan dapat dijadikan dasar untuk pengembangan sistem pengenalan musik yang lebih akurat di masa mendatang.

Daftar Pustaka

[1] Chicco, D., Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 6 (2020). https://doi.org/10.1186/s12864-019-6413-7

[2] Gao, W., Liu, H., Wang, X., & Sun, Y. (2022). *An enhanced feature fusion method for music signal classification*. Journal of Intelligent & Fuzzy Systems, 43(1), 715–726. https://doi.org/10.3233/JIFS-212133

[3] Hossain, M. M., Hasan, M. R., & Hossain, M. A. (2021). *Musical instrument classification using combined time and frequency domain features*. International Journal of Computer Applications, 183(8), 25–30. https://doi.org/10.5120/ijca2021921266

[4] H. Zhang, Y. Wang, and C. Tang, “Deep learning based audio classification: A review,” *Neurocomputing*, vol. 426, pp. 1–17, Jan. 2021. https://doi.org/10.1016/j.neucom.2020.10.002

[5] M. Zhang, Y. Chen, Y. Su, and D. Zhang, “Music genre classification with improved neural networks,” *IEEE Access*, vol. 8, pp. 183744–183754, 2020. https://doi.org/10.1109/ACCESS.2020.3029383

[6] Q. Kong, Y. Xu, W. Wang, and M. D. Plumbley, “Sound event detection of weakly labelled data with CNN-transformer and automatic threshold optimization,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 1626–1637, 2021. https://doi.org/10.1109/TASLP.2021.3076736

[7] S. Tharwat, “Classification assessment methods,” *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2021. https://doi.org/10.1016/j.aci.2018.08.003

[8] T. Nakashika, “Chord recognition using deep recurrent neural networks and data augmentation,” *Acoust. Sci. Technol.*, vol. 41, no. 2, pp. 312–320, Mar. 2020. https://doi.org/10.1250/ast.41.312