

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Pasar musik rekaman global mengalami pertumbuhan yang signifikan sebesar 10,2% pada tahun 2023, di mana jumlah pelanggan layanan streaming berbayar kini telah melampaui angka 700 juta orang [18]. Fenomena ini mempertegas betapa pentingnya peran Music Genre Classification (MGC) dalam lingkup Music Information Retrieval (MIR). Penggunaan MGC secara otomatis sangat krusial untuk meningkatkan pengalaman pengguna melalui sistem rekomendasi yang lebih cerdas dan pengorganisasian perpustakaan musik yang lebih terstruktur [15].

Metode MGC telah mengalami evolusi dari teknik tradisional, seperti penggunaan Mel-Frequency Cepstral Coefficients (MFCCs) dan Support Vector Machines (SVMs) [7], menuju model pembelajaran mendalam (deep learning) yang lebih canggih. Model-model tersebut meliputi Convolutional Neural Networks (CNNs) [5] serta Recurrent Neural Networks (RNNs), khususnya arsitektur Long Short-Term Memory (LSTM) yang terbukti sangat efektif dalam mengenali pola temporal pada musik.

Seiring dengan kemajuan teknologi, baru-baru ini telah dikembangkan model fondasi audio berbasis self-supervised learning yang secara umum dikenal sebagai Audio Large Language Models (LLMs), seperti HuBERT, WavLM, dan Wav2Vec 2.0 [16]. Ketiga model ini dipilih dalam penelitian ini karena mereka mewakili berbagai pendekatan inovatif dalam pembelajaran self-supervised: HuBERT menggunakan masked prediction, WavLM mengintegrasikan denoising dan pelatihan yang sadar akan pembicara (speaker-aware), sementara Wav2Vec 2.0 mengandalkan contrastive learning.

Penelitian ini bertujuan untuk melakukan analisis komparatif antara model berbasis LSTM yang telah dioptimalkan dengan tiga model LLM audio yang disesuaikan (fine-tuned) untuk tugas MGC. Studi ini diharapkan dapat memberikan wawasan mendalam mengenai keseimbangan antara spesialisasi arsitektur dan generalisasi model dalam merancang sistem klasifikasi genre musik yang lebih efektif.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka permasalahan yang diangkat dalam penelitian ini sebagai berikut:

1. Bagaimana performa model LSTM domain-spesifik dalam mengklasifikasikan genre musik menggunakan representasi Mel-Spectrogram pada dataset GTZAN?
2. Bagaimana performa model Audio Large Language Models (HuBERT, WavLM, dan WAV2Vec 2.0) setelah dilakukan fine-tuning untuk tugas klasifikasi genre musik?
3. Apakah terdapat perbedaan yang signifikan antara performa model LSTM dengan model Audio LLM dalam tugas Music Genre Classification (MGC) berdasarkan metrik Accuracy, Precision, Recall, dan F1-Score?
4. Model manakah yang paling efektif dan efisien untuk digunakan dalam sistem klasifikasi genre musik otomatis antara model LSTM domain-spesifik dan model Audio LLM?

## 1.3 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

- a. Mengoptimalkan dan mengevaluasi performa arsitektur Neural Network berbasis Long Short-Term Memory (LSTM) dengan input Mel-spectrogram untuk mencapai akurasi maksimal dalam klasifikasi genre musik pada dataset GTZAN
- b. Mengimplementasikan proses fine-tuning pada tiga model Audio Large Language Models (LLMs) HuBERT, WavLM, dan Wav2Vec 2.0—

menggunakan data audio mentah (raw audio) untuk melihat sejauh mana model generalis ini dapat beradaptasi pada tugas spesifik musik

- c. Melakukan analisis komparatif yang mendalam antara model khusus domain (LSTM) dan model generalis (LLMs) guna menentukan arsitektur mana yang paling efektif dalam menangkap fitur akustik dan temporal yang unik dari setiap genre musik
- d. Mengidentifikasi kekuatan dan kelemahan dari masing-masing pendekatan, terutama dalam hal kemampuan generalisasi, risiko overfitting, serta akurasi klasifikasi pada genre-genre yang memiliki karakteristik serupa

#### **1.4 Batasan Penelitian**

Penelitian ini memiliki batasan sebagai berikut:

- a. Penelitian ini terbatas pada penggunaan dataset GTZAN, yang terdiri dari 1.000 klip audio berdurasi 30 detik yang mencakup 10 genre musik (Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, dan Rock)
- b. Analisis hanya dibatasi pada perbandingan antara arsitektur Long Short-Term Memory (LSTM) yang dioptimalkan dengan tiga varian Audio Large Language Models (LLMs), yaitu HuBERT, WavLM, dan Wav2Vec 2.0
- c. Untuk model LSTM, input yang digunakan adalah fitur Mel-spectrogram yang telah diekstraksi dari sinyal audio.
- d. Untuk model LLM, proses fine-tuning dilakukan langsung menggunakan audio mentah (raw waveform)

#### **1.5 Jadwal Pengerjaan**

Tabel berikut merangkum seluruh rangkaian tahapan penelitian, proses submisi jurnal ilmiah, serta penyusunan laporan akhir yang direncanakan selama enam bulan

**Tabel 1.1. Jadwal Pengerjaan**

No	Kegiatan	Bulan					
		1	2	3	4	5	6
1	Bimbingan, Penentuan Topik, dan Latar Belakang	■					
2	Studi Literatur dan Pengumpulan Referensi	■	■				
3	Finalisasi Eksperimen, Pengumpulan Data dan menyusun draft jurnal		■	■			
4	Sidang Pleno			■			
5	Submit Jurnal dan Monitoring Review			■	■		
6	Penyusunan Laporan Akhir (Bab 1-5)				■	■	
7	Revisi Laporan dan Persiapan Sidang					■	■
8	Sidang Tugas Akhir						■