

# **BAB I**

## **PENDAHULUAN**

Pada bab ini akan menjelaskan mengenai hal-hal yang melatar belakangi penelitian ini beserta rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian.

### **1.1 Latar Belakang**

Musik telah lama dikenal sebagai salah satu bentuk seni yang paling kuat dalam menyampaikan dan membangkitkan emosi. Salah satu elemen yang paling penting ketika mendengarkan sebuah musik adalah emosi yang diekspresikan musik tersebut. Melalui kombinasi elemen seperti melodi, harmoni, dan ritme, musik dapat menciptakan suasana hati yang mendalam, memengaruhi perasaan pendengarnya, dan menyampaikan pesan emosi yang kompleks [1]. Di era digital saat ini, volume besar data musik yang tersedia secara daring semakin secara pesat dan majemuk sehingga menjadikan tantangan tersendiri untuk bisa mengkategorikan dengan label-label tersendiri agar bisa membedakan dari yang lain. Oleh karena itu, peluang klasifikasi musik semakin menjadi sangat jelas [2].

Perkembangan pada *machine learning* yang menjadi sangat cepat untuk zaman digitalisasi saat ini serta informasi yang tersebar luas memungkinkan kemampuan untuk bisa menganalisis, memahami, dan mengenali emosi yang menjadi perhatian di berbagai bidang [3]. Untuk saat ini para peneliti yang meneliti pengenalan emosi musik memfokuskan pada dua aspek yaitu, pengoptimalan ekstraksi fitur pada musik dan mengembangkan performa penggolongan dari pengenalan emosi. Akurasi dari ekstraksi emosional pada fitur musik dipengaruhi dari hasil akurasi dan efisiensi klasifikasi. Pengenalan emosi musik melibatkan kedua psikologi musik serta ilmu komputer yang dimana menjadi tantangan tersendiri [4].

Dalam konteks analisis emosi musik, aspek audio seperti melodi, tempo, dan dinamika sering kali menjadi fokus utama. Teknik seperti Mel-Frequency Cepstral Coefficients (MFCC) telah digunakan untuk mengekstraksi fitur audio yang relevan. Oleh karena itu, analisis emosi musik yang efektif seharusnya tidak hanya memperhitungkan fitur audio. Kemajuan dalam bidang *deep learning*, khususnya penggunaan Convolutional Neural Networks (CNN) telah memungkinkan analisis yang lebih komprehensif terhadap data yang kompleks seperti musik. CNN menggabungkan

kekuatan Convolutional Neural Networks (CNN) dalam mengekstraksi fitur spasial dari audio. Dengan demikian, CNN menawarkan solusi ideal untuk menangkap pola spasial dari audio, serta dinamika temporal dari melodi [5].

Dalam hasil pemrosesan dari penggunaan algoritma Convolutional Neural Network (CNN), sinyal audio dari musik atau lagu tersebut akan di ekstrak terlebih dahulu dengan menggunakan pendekatan Mel-Frequency Cepstral Coefficient (MFCC) untuk menangkap karakteristik suara atau audio yang penting untuk pengenalan suara [6]. Dalam ekstraksi fitur audio pada musik bertujuan untuk mendapatkan *valence* dan *arousal* dalam ruang dimensional. Dimensi-dimensi tersebut memengaruhi ransangan yang mendasari pengaruh respon pada suasana hati. [7] Implementasi CNN pada audio tersebut akan berfungsi untuk melatih model yang bisa mengkombinasikan kemampuan pemrosesan spasial dan temporal dalam menangani data dari audio tersebut. Setelah itu CNN merupakan alat analisis yang populer dalam penanganan data audio serta dapat menangkap *spectrogram* dari audio yang dimana membantu mempelajari pola dan struktur lokal secara efisien [8].

CNN dipilih untuk analisis sinyal audio karena kemampuannya mengintegrasikan analisis spasial pada audio dengan CNN, menjadikannya lebih unggul dibandingkan metode lain. Pada sinyal audio, CNN dapat mengenali pola frekuensi dan tekstur dari representasi seperti *mel-spectrogram*, sedangkan metode lain seperti LSTM murni atau Recurrent Neural Network (RNN) cenderung kurang efisien dalam menangkap fitur spasial ini serta CNN merupakan algoritma berbasis neural network yang paling menjanjikan dalam hal klasifikasi karena memiliki akurasi yang tinggi untuk mendeteksi pola dan mengklasifikasikan fitur-fitur [9].

Dibandingkan dengan model non-neural seperti Random Forest atau Support Vector Machine (SVM), CNN memiliki keunggulan signifikan dalam memahami data multimodal, karena model tradisional tersebut tidak dirancang untuk menangani pola spasial dan temporal secara bersamaan [7]. Selain itu, metode seperti *Transformer* yang lebih canggih juga dapat digunakan untuk teks, tetapi cenderung membutuhkan lebih banyak data dan sumber daya komputasi dibandingkan LSTM dalam pengaturan tertentu. Dengan memanfaatkan CNN untuk sinyal audio, model ini menjadi pilihan

yang efisien dan efektif untuk tugas-tugas seperti klasifikasi emosi musik, di mana informasi spasial dan temporal dari kedua jenis data harus diintegrasikan dengan baik.

Dengan fitur audio, penelitian ini diharapkan mampu memberikan pemahaman yang lebih mendalam tentang bagaimana musik dalam menyampaikan emosi kepada pendengarnya. Hasil dari klasifikasi nantinya akan bisa membantu membangun referensi terhadap pengembangan sistem rekomendasi musik dan lainnya [10]. Implementasi CNN juga diharapkan dapat menjadi pendekatan baru yang lebih efektif dan efisien dalam menganalisis emosi dalam musik, serta membuka potensi aplikasi praktis di berbagai bidang seperti pengembangan rekomendasi musik yang lebih personal, terapi berbasis musik, hingga analisis preferensi musik pengguna secara otomatis.

## 1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini mencakup pertanyaan-pertanyaan yang ingin dijawab sehubungan dengan analisis emosi dalam musik. Beberapa pertanyaan yang dapat diajukan antara lain,

1. Bagaimana cara mengekstraksi fitur audio dari suatu musik untuk analisis emosi?
2. Bagaimana menerapkan *Convolutional Neural Network* untuk pemrosesan klasifikasi emosi pada musik?
3. Bagaimana tingkat akurasi klasifikasi emosi musik yang diperoleh dengan menggunakan metode *Convolutional Neural Network*?

## 1.3 Batasan Masalah

Dalam penelitian ini, terdapat beberapa batasan yang perlu ditetapkan agar fokus penelitian tetap terjaga. Batasan masalah tersebut antara lain,

1. Dataset yang digunakan adalah DEAM (Database for Emotional Analysis of Music) yang terdiri dari musik populer dengan anotasi emosi berdasarkan *valence* dan *arousal*.
2. Ekstraksi fitur audio hanya akan menggunakan MFCC sebagai fitur dasar, dengan eksplorasi penambahan fitur musik seperti augmentasi untuk informasi harmonik.
3. Model emosi yang digunakan merupakan model emosi dengan 4 kuadran *valence* dan *arousal* serta penyederhanaan menjadi 2 kuadran *valence* atau *arousal*.

4. Evaluasi dilakukan pada level sekmen dengan agregasi untuk prediksi level lagu, menggunakan metrik akurasi, akurasi *balanced*, dan *F1-score* untuk menangani ketidakseimbangan lagu.

#### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan metode ekstraksi fitur yang efektif dari audio musik untuk analisis emosi.
2. Mengimplementasikan Convolutional Neural Network (CNN) untuk klasifikasi emosi dalam musik berdasarkan kombinasi fitur audio.
3. Mengevaluasi akurasi dan efektivitas model CNN dalam mengidentifikasi emosi yang terkandung dalam musik.
4. Menyediakan wawasan mengenai hubungan antara fitur audio dalam menyampaikan emosi dalam musik.

#### 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Menjadi referensi untuk studi lanjutan tentang analisis emosi dalam musik dan penerapan teknik deep learning dalam pemrosesan sinyal audio.
2. Mengetahui kinerja Convolutional Neural Network (CNN) untuk fitur audio.
3. Mengetahui klasifikasi emosi lagu dengan pendekatan fitur audio sebagai data input.
4. Memberikan *framework* metodologi yang komprehensif untuk ekstraksi fitur audio dan implementasi *deep learning* dalam analisis musik yang dapat diadaptasi untuk domain pemrosesan audio lainnya.
5. Memberikan kontribusi terhadap pengembangan sistem kecerdasan buatan yang lebih canggih dalam analisis dan pemahaman emosi, serta membuka peluang penelitian lebih lanjut dalam interaksi manusia dan mesin melalui musik.