

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Penelitian tentang kecerdasan buatan dengan Algoritma *Convolutional Neural Network* atau yang biasa disingkat CNN dalam beberapa waktu terakhir terus berkembang khususnya pada teknologi pengenalan emosi dari ucapan. Seperti yang umumnya diketahui, *speech* atau ucapan adalah suatu hal yang penting bagi manusia dalam berkomunikasi karena menggambarkan emosi dari manusia [1]. Dalam pendekatan menggunakan teknologi, terdapat sebuah sistem yang dapat mengenali emosi manusia yang disebut *Speech Emotion Recognition* (SER). Pada penerapannya, *SER* memiliki ruang lingkup yang cukup luas seperti membantu dalam bidang psikologi, robotik, *mobile services*, dan lain sebagainya [1].

Selain itu, emosi adalah hal yang tidak mudah untuk dianalisis oleh sistem karena manusia memiliki karakteristik yang bervariasi dalam berbicara seperti cara berartikulasi [2], latar belakang kultur pembicara [3], kecepatan berbicara, tinggi rendahnya suara dan lain sebagainya. Keberadaan *noise* atau kebisingan pada sebuah audio juga dapat menyulitkan sistem untuk mengenali emosi dari ucapan.

Berdasarkan beberapa penelitian yang ada, Algoritma CNN adalah salah satu algoritma yang dapat mendeteksi emosi manusia dengan cukup baik. Contohnya seperti pada penelitian yang menggabungkan algoritma CNN dengan algoritma lain seperti LSTM [4],[5] ataupun kombinasi algoritma yang sama namun memiliki dimensi yang berbeda seperti CNN 1 dan 2 dimensi [6] yang menunjukkan bahwa CNN dapat melakukan prediksi emosi dari manusia. Selain itu juga, pendekatan lain dilakukan untuk meningkatkan akurasi dari algoritma CNN dengan dilakukannya *augmentation* [1], [4] pada *dataset* sehingga membuat *dataset* menjadi lebih variatif dan membuat model CNN terhindar dari masalah seperti *overfitting*.

Berdasarkan beberapa penelitian tersebut, peneliti terinspirasi dari penelitian yang menggabungkan perbedaan dimensi dari data yang diekstrak ke dalam algoritma CNN yaitu 1 dan 2 dimensi [6] sehingga kemudian dapat memprediksi emosi dari manusia berdasarkan suara. Maka dari itu, penelitian ini diharapkan dapat memberikan hasil yang dapat berkontribusi pada topik *Speech Emotion Recognition* dan juga sebagai salah satu pilihan arsitektur serta sebagai bahan referensi untuk penelitian-penelitian di masa mendatang.



## 1.2 Rumusan Masalah

Berdasar latar belakang masalah seperti dijelaskan pada 1.1, maka beberapa rumusan masalah yang ada pada penelitian ini adalah:

1. Bagaimana implementasi kombinasi dari CNN 1 dan 2 dimensi dapat melakukan prediksi terhadap emosi dari ucapan manusia?
2. Seberapa akurat model dapat memprediksi emosi manusia berdasarkan *dataset* yang ada dari ketiga algoritma CNN 1, 2 dan kombinasi?

## 1.3 Batasan Masalah

Peneliti menetapkan batasan-batasan masalah agar dapat mendapat hasil yang diinginkan. Adapun batasan-batasan masalah adalah sebagai berikut.

1. Penelitian ini hanya menggunakan Algoritma *Convolutional Neural Network* (CNN) yang diimplementasikan pada kasus *Speech Emotion Recognition* (SER)
2. Menggunakan Algoritma CNN 1 dan 2 dimensi serta kombinasi dari keduanya
3. Nilai akurasi yang dipakai adalah hasil akurasi validasi untuk menghindari *overfitting* yang berlebihan
4. Menggunakan 3 *Dataset* publik yaitu *Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)* [7], *The Audio, Speech, and Vision Processing Lab Emotional Sound database (ASVP-ESD)* [8], *Toronto emotional speech set (TESS)* [9]
5. Emosi yang diklasifikasikan berjumlah 6 yaitu senang, marah, sedih, jijik, takut, dan netral
6. Menggunakan fitur ekstraksi yang berbasis *fourier transform*

## 1.4 Tujuan Penelitian

1. Penelitian ini bertujuan untuk mengimplementasikan kombinasi dari hasil ekstraksi fitur yang memiliki 1 dan 2 dimensi ke dalam algoritma CNN 1 dan 2 dimensi.
2. Penelitian ini bertujuan untuk mengevaluasi keakuratan dari algoritma CNN

1, 2 dan kombinasi.

3. Penelitian ini juga diharapkan dapat menjadi pilihan arsitektur untuk digunakan pada bidang penelitian *Speech Emotion Recognition* di masa yang akan datang

#### 1.5 Manfaat Penelitian

Manfaat penelitian bagi peneliti, ilmu pengetahuan adalah sebagai berikut:

1. Bagi penulis  
Penulis dapat memahami lebih dalam tentang bagaimana implementasi dari model CNN 1 dan 2 dimensi yang dikombinasikan pada kasus *Speech Emotion Recognition*
2. Bagi ilmu pengetahuan  
Penelitian ini dapat memberikan manfaat bagi peneliti dan penelitian selanjutnya pada topik *Speech Emotion Recognition* sebagai referensi perbandingan algoritma dan juga untuk memperkaya literatur

#### 1.6 Sistematika Penulisan

Dalam melakukan penelitian ini sistematika penulisan agar mudah dibaca dan dipahami sehingga dibagi menjadi beberapa bab yaitu sebagai berikut

**BAB I PENDAHULUAN**, berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan masalah, manfaat masalah dan sistematika penulisan

**BAB II TINJAUAN PUSTAKA**, berisi tinjauan pustaka dari penelitian-penelitian sebelumnya yang berasal dari jurnal maupun buku, dasar-dasar teori yang digunakan peneliti sebagai pendukung dalam penelitian ini

**BAB III METODE PENELITIAN**, berisi objek penelitian yang akan diteliti, Alur penelitian yang menjelaskan bagaimana penelitian ini dikerjakan secara bertahap beserta penjelasan untuk setiap alurnya, lalu terdapat juga bagian alat dan bahan yang mendeskripsikan alat dan bahan apa saja yang digunakan pada penelitian ini.

**BAB IV HASIL DAN PEMBAHASAN**, pada bab ini peneliti melakukan pembahasan yang lebih teknis seperti implementasi apa saja yang dilakukan pada

saat melakukan penelitian, dan juga penjelasan hasil yang didapatkan dari implementasi seperti melakukan testing pada model, dan juga dilakukan evaluasi sebagai hasil akhir yang mengukur model yang telah dibangun

**BAB V PENUTUP**, berisi kesimpulan dan saran yang dapat peneliti rangkum selama proses penelitian sehingga penelitian ini dapat dikembangkan lagi lebih lanjut.

