

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dialog atau percakapan merupakan salah satu bentuk interaksi manusia yang penting dalam kehidupan sehari-hari. Interaksi manusia seringkali tidak selalu berjalan lancar karena berbagai faktor seperti perbedaan pendapat, ekspektasi yang tidak sesuai dengan kenyataan, atau bisa juga dipengaruhi oleh suasana hati seseorang (Dwi Septria dkk., 2019).

Suasana hati (*mood*) didefinisikan pada kondisi emosional yang cenderung bertahan lebih lama dan memiliki intensitas relatif lebih rendah daripada emosi, serta tidak selalu dipicu oleh suatu peristiwa tertentu (Reeve, 2018; Hockenbury & Hockenbury, 2010). Keadaan ini dapat digambarkan melalui penilaian diri (self-report) atau observasi perilaku dan ekspresi nonverbal. Misalnya, individu dapat merasa murung atau gembira sepanjang hari meski tidak mengetahui penyebabnya secara pasti (Planalp, 1999). Suasana hati berperan sebagai latar yang memengaruhi cara seseorang memersepsikan dan mengekspresikan emosi. Ketika seseorang berada dalam suasana hati yang positif, munculnya emosi menyenangkan (misalnya senang atau optimis) lebih memungkinkan, sebaliknya suasana hati negatif cenderung memicu emosi seperti sedih atau marah (Ekman, 1992).

Dalam konteks *Speech Emotion Recognition (SER)*, suasana hati dapat memengaruhi intonasi serta ciri akustik seseorang, sehingga sistem otomatis perlu membedakan ciri-ciri yang berasal dari *mood* jangka panjang dan emosi intens yang muncul secara khusus. Ciri akustik merujuk pada karakteristik atau fitur suara yang

dapat diukur dan dianalisis, seperti tinggi nada (*pitch*), intensitas (*loudness*), kualitas suara (*timbre*), tempo, durasi, serta pola frekuensi tertentu.

Emosi adalah perasaan yang kuat yang dimiliki setiap individu terhadap sesuatu. Emosi juga didefinisikan sebagai reaksi timbal balik terhadap tindakan orang lain atau peristiwa yang dialami oleh pemilik emosi (Akbar & Lawi, 2021). Emosi dapat diklasifikasikan menjadi kecemasan, kebosanan, ketidakpuasan, depresi, jijik, kekecewaan, ketakutan, kebahagiaan, ketidakpedulian, ironi, kegembiraan, netral, panik, larangan, kejutan, kesedihan, stres, malu, kelelahan, stres dan khawatir (Prasetio & Kurniawan, 2017). Saat ini, proses klasifikasi emosi dapat dilakukan melalui pemanfaatan teknologi yang disebut dengan *Speech Emotion Recognition (SER)*.

Speech Emotion Recognition (SER) adalah teknologi yang dirancang untuk mengidentifikasi kondisi emosional atau fisik pembicara dari sinyal ucapannya. Teknologi ini telah menarik banyak peneliti pada saat ini karena pentingnya teknologi ini dalam banyak aplikasi, termasuk *e-learning*, keamanan, kesehatan, sistem penerjemah otomatis dan robotika (Idris, 2015). Penelitian dalam model klasifikasi dalam bidang SER telah banyak dilakukan, tetapi hanya sedikit diantara banyak penelitian bidang SER yang menghasilkan pekerjaan yang efektif dalam memprediksi adanya emosi melalui ucapan. Ada empat model klasifikasi yang sering digunakan model HMM, Model GMM, Model SVM dan Model DL (Deep Learning) (Triandi dkk., 2022). Model DL merupakan model yang cukup populer khususnya pada proses klasifikasi. Salah satu algoritma *deep learning* yang paling sering digunakan yaitu *Convolutional Neural Network (CNN)*. Selain itu MFCC

(Mel-Frequency Cepstral Coefficients) adalah metode yang populer untuk ekstraksi fitur pada audio, terutama dalam pengenalan suara (Fitri Widodo and Fadlil, 2019).

Convolutional Neural Network (CNN) adalah salah satu metode *deep learning* dari pengembangan *Multi Layer Perceptron* (MLP) yang didesain untuk mengolah data dua dimensi. CNN termasuk dalam jenis *Deep Neural Network* karena dalamnya tingkat jaringan dan banyak diimplementasikan dalam data citra (Emanuella dkk., 2021). Algoritma CNN dipilih karena memiliki kemampuan ekstraksi fitur yang tinggi (Listyarini & Anggoro, 2021). Hal tersebut dibuktikan dengan penelitian terkait sebelumnya yaitu pada penelitian identifikasi emosi melalui suara menggunakan *Support Vector Machine* dan *Convolutional Neural Network* yang menunjukkan bahwa metode CNN menghasilkan akurasi sebesar 73%, sedangkan metode SVM menghasilkan akurasi sebesar 63%. Pada penelitian tersebut juga didapatkan hasil nilai presisi dan recall menggunakan CNN dari hamper tiap emosi naik. Sedangkan Ketika menggunakan SVM, nilai presisi dan recall masih banyak yang turun. Oleh karena itu, metode CNN memiliki kinerja yang lebih baik jika dibandingkan dengan metode SVM (Galang, 2021).

Berdasarkan penelitian sebelumnya, pengenalan emosi berdasarkan suara pada dataset SAVEE menggunakan algoritma *deep learning* CNN menunjukkan akurasi akhir dari model sebesar 88% pada data latih dan 52% pada data tes yang berarti model mengalami *overfitting*. Hal tersebut disebabkan oleh tidak seimbangnya kelas emosi pada dataset sehingga membuat model akan cenderung memprediksi kelas yang labelnya lebih banyak. Selain itu, kurangnya heterogenitas dari dataset yang membuat karakter kelas emosi lebih berbeda dari yang lainnya sehingga

mampu mengurangi bias pada model agar tidak membuat model *overfitting*. Pengembangan lebih lanjut dari penelitian ini yaitu dengan lebih mendalami salah satu metode ekstraksi fitur yang digunakan dan menggunakan dataset lain yang memiliki jumlah data yang lebih banyak (Rendi Nurcahyo & Mohammad Iqbal, 2022).

Berdasarkan latar belakang, penelitian ini akan memfokuskan pada analisis klasifikasi emosi berdasarkan suara manusia dengan menggunakan dataset CREMA-D. Dataset CREMA-D memiliki berbagai variasi nada suara, dan situasi yang menciptakan lingkungan penelitian yang sangat relevan dengan pengenalan emosi berdasarkan suara. Sehingga penelitian ini memberikan lebih banyak variasi emosi dan intensitas suara. Teknik augmentasi data yang lebih beragam, seperti *gaussian noise*, *time stretching*, dan *pitch shifting*, diharapkan akan memperkaya *data training*.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka dapat dirumuskan masalah dalam penelitian ini, yaitu:

1. Bagaimana cara untuk mengklasifikasikan emosi berdasarkan suara manusia dengan menerapkan algoritma *Convolutional Neural Network*.
2. Bagaimana performa algoritma *Convolutional Neural Network* dalam mengklasifikasikan emosi berdasarkan suara manusia.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini, yaitu:

1. Mengklasifikasikan emosi berdasarkan suara manusia menggunakan algoritma *Convolutional Neural Network*.
2. Menganalisis performa algoritma *Convolutional Neural Network* dalam mengklasifikasikan emosi berdasarkan suara manusia.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat berkontribusi dalam ilmu pengetahuan sebagai sumber referensi mengenai pengklasifikasian emosi berdasarkan suara manusia menggunakan algoritma *Deep Learning Convolutional Neural Network*. Selain itu, penelitian yang dilakukan dapat menjadi acuan untuk pihak atau peneliti yang lain dengan kasus yang serupa.

Lebih jauh, hasil penelitian ini memiliki berbagai manfaat praktis yang dapat dirasakan langsung. Sistem pengenalan emosi berbasis suara dapat digunakan dalam aplikasi layanan pelanggan untuk mendeteksi emosi pengguna, sehingga dapat memberikan tanggapan yang lebih tepat dan meningkatkan pengalaman pelanggan. Selain itu, teknologi ini dapat diterapkan dalam perangkat asisten virtual seperti *Google Assistant*, *Siri*, atau *Alexa*, yang membuat interaksi lebih manusiawi dan responsif terhadap keadaan emosional pengguna. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi pada ilmu pengetahuan tetapi juga menawarkan manfaat praktis yang dapat meningkatkan interaksi manusia dan teknologi dalam berbagai konteks.

1.5 Batasan Masalah

Terdapat beberapa batasan masalah dalam penelitian ini, sebagai berikut:

1. Format file suara yang digunakan adalah wave (*.wav).

2. Data suara yang digunakan merupakan dataset CREMA-D.
3. Proses klasifikasi emosi yang dilakukan menggunakan algoritma *deep learning Convolutional Neural Network*.

1.6 Sistematika Penulisan

Pembahasan pada penelitian ini terbagi menjadi lima BAB, masing masing BAB diuraikan sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, metodologi penelitian dan sistematika penulisan pada penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas mengenai teori-teori yang berhubungan dengan analisis sentimen yang menjadi pokok utama penelitian serta *state of the art* dari penelitian terkait sebelumnya.

BAB III METODOLOGI PENELITIAN

Bab ini berisi gambaran secara umum mengenai penelitian yang dilakukan dan penjelasan tentang apa saja yang dilakukan pada penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas tentang implementasi sesuai dengan metodologi penelitian yang digunakan, serta pengujian untuk mendapatkan hasil yang sesuai dengan tujuan penelitian.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran berdasarkan temuan penelitian serta garis besar metode penelitian yang digunakan. Kesimpulan merupakan hasil akhir penelitian, sedangkan saran meliputi rekomendasi yang sesuai dengan keterbatasan penelitian yang dilakukan.