

BAB I

PENDAHULUAN

1.1 Latar Belakang

Emosi merupakan komponen penting dalam komunikasi manusia karena berperan dalam mempengaruhi cara seseorang berinteraksi, mengambil keputusan, serta merespons suatu situasi. Dalam konteks perkembangan teknologi modern, kemampuan sistem untuk mengenali emosi menjadi kebutuhan yang semakin penting, terutama pada aplikasi seperti virtual assistant, layanan kesehatan mental, pendidikan jarak jauh, hingga sistem customer service otomatis. Berbagai kajian terbaru menunjukkan bahwa pengenalan emosi dari suara (Speech Emotion Recognition / SER) terus berkembang karena potensi penerapannya yang luas dalam berbagai domain [1].

Suara merupakan modalitas yang kaya akan informasi emosional karena mengandung elemen prosodik seperti intonasi, nada, ritme, dan energi yang relatif sulit dimanipulasi serta tetap muncul meskipun konten linguistik tidak disampaikan secara eksplisit. Studi sistematis menunjukkan bahwa fitur prosodik dan spektral seperti Mel Frequency Cepstral Coefficient (MFCC), pitch, formant, dan energi memiliki peran penting dalam mengenali emosi manusia secara akurat [1]. Meskipun demikian, tantangan seperti variasi karakteristik penutur, perbedaan bahasa, serta kondisi lingkungan yang bising masih menjadi permasalahan utama dalam pengembangan sistem SER [4].

Dalam beberapa tahun terakhir, pendekatan deep learning semakin mendominasi penelitian di bidang SER karena kemampuannya dalam mempelajari pola data yang kompleks dan non-linear secara otomatis. Salah satu metode deep learning yang banyak digunakan adalah Convolutional Neural Network (CNN). CNN terbukti efektif dalam mengekstraksi fitur penting dari representasi sinyal suara, khususnya ketika data suara diubah ke dalam bentuk representasi spektrum seperti MFCC. Penelitian-penelitian terbaru menunjukkan bahwa CNN mampu menghasilkan performa klasifikasi emosi yang lebih baik dibandingkan metode machine learning konvensional, terutama pada dataset berskala menengah hingga besar [2].

Keunggulan utama CNN terletak pada kemampuannya melakukan ekstraksi fitur secara otomatis melalui proses konvolusi, sehingga ketergantungan terhadap perancangan fitur secara manual dapat dikurangi. Dalam konteks SER, CNN mampu menangkap pola temporal dan spektral dari sinyal suara dengan lebih efektif, terutama ketika dikombinasikan dengan fitur MFCC serta turunannya seperti delta dan delta-delta [3]. Pendekatan ini dinilai lebih adaptif dalam menghadapi variasi data suara dibandingkan metode berbasis fitur statis.

Tantangan lain yang banyak dibahas dalam literatur adalah perbedaan performa model pada berbagai jenis korpus suara. Beberapa penelitian menunjukkan bahwa model yang dilatih menggunakan dataset tertentu, seperti RAUDES2, tidak selalu memberikan hasil optimal ketika diuji pada dataset lain (cross-corpus evaluation). Hal ini menunjukkan pentingnya proses pelatihan dan

evaluasi model yang tepat agar sistem SER dapat bekerja secara andal pada kondisi dunia nyata [3], [4].

Dalam domain aplikasi, deteksi emosi berbasis suara mulai dimanfaatkan untuk pemantauan kesehatan mental, deteksi stres, serta penilaian kondisi psikologis pengguna. Tinjauan literatur terbaru menegaskan bahwa SER memiliki potensi besar sebagai alat bantu non-invasif bagi profesional kesehatan, dengan syarat model yang digunakan memiliki tingkat akurasi yang memadai dan mampu bekerja secara stabil pada berbagai kondisi lingkungan [4].

Berdasarkan uraian tersebut, penelitian ini berfokus pada pengembangan sistem deteksi emosi manusia berbasis suara menggunakan metode Convolutional Neural Network (CNN). Penelitian ini bertujuan untuk mengevaluasi performa CNN dalam mengklasifikasikan emosi suara manusia berdasarkan fitur MFCC, serta mengimplementasikan model yang dihasilkan ke dalam sebuah aplikasi berbasis web. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem SER yang lebih akurat dan aplikatif untuk penggunaan di dunia nyata.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana tahapan pengolahan sinyal suara yang meliputi preprocessing dan ekstraksi fitur berbasis Mel Frequency Cepstral Coefficient (MFCC) dalam mendeteksi emosi manusia melalui suara?

2. Bagaimana penerapan metode Convolutional Neural Network (CNN) dalam melakukan klasifikasi emosi suara manusia berdasarkan fitur MFCC?
3. Sejauh mana tingkat akurasi model Convolutional Neural Network (CNN) dalam mengklasifikasikan emosi manusia ke dalam empat kelas emosi, yaitu marah, senang, sedih, dan netral?
4. Bagaimana kinerja model CNN dalam mendeteksi emosi manusia berdasarkan suara ditinjau dari hasil evaluasi menggunakan confusion matrix dan classification report?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Penelitian ini dibatasi pada deteksi emosi manusia berdasarkan sinyal suara, tanpa mempertimbangkan aspek non-verbal lainnya seperti ekspresi wajah, gerakan tubuh, atau teks ucapan.
2. Emosi yang diklasifikasikan dalam penelitian ini dibatasi pada empat kelas emosi, yaitu marah (angry), senang (happy), sedih (sad), dan netral (neutral).
3. Fitur yang digunakan dalam penelitian ini adalah Mel-Frequency Cepstral Coefficients (MFCC) serta fitur turunannya yaitu delta (turunan pertama) dan delta-delta (turunan kedua). Kombinasi fitur ini digunakan untuk menangkap karakteristik spektral dan dinamika perubahan sinyal suara yang berkaitan dengan ekspresi emosi.

4. Metode klasifikasi yang digunakan dalam penelitian ini dibatasi pada algoritma Convolutional Neural Network (CNN), tanpa melakukan perbandingan dengan algoritma klasifikasi lain seperti SVM atau Random Forest.
5. Dataset suara yang digunakan dalam penelitian ini berasal dari dataset publik RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song), dengan data suara yang difilter sesuai dengan empat kelas emosi yang diteliti.
6. Proses pelatihan dan pengujian model dilakukan menggunakan bahasa pemrograman Python melalui platform Google Colab, sedangkan implementasi sistem deteksi emosi dilakukan menggunakan framework Streamlit.
7. Evaluasi kinerja model dibatasi pada pengukuran akurasi, confusion matrix, dan classification report, tanpa melakukan pengujian lanjutan seperti cross-validation atau optimasi hiperparameter secara mendalam.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menganalisis proses pengolahan sinyal suara manusia yang meliputi tahap preprocessing dan ekstraksi fitur spektral berbasis Mel Frequency Cepstral Coefficient (MFCC) untuk mendeteksi emosi berdasarkan suara.
2. Menerapkan algoritma Convolutional Neural Network (CNN) dalam melakukan klasifikasi emosi suara manusia ke dalam empat kelas emosi,

yaitu marah (angry), senang (happy), sedih (sad), dan netral (neutral), berdasarkan fitur MFCC, delta MFCC, dan delta-delta MFCC.

3. Mengukur kinerja model CNN dalam mendeteksi emosi suara manusia menggunakan dataset RAVDESS dengan menggunakan metrik evaluasi berupa akurasi, confusion matrix, dan classification report.
4. Mengimplementasikan model deteksi emosi berbasis suara ke dalam sebuah aplikasi berbasis web menggunakan framework Streamlit, sehingga sistem dapat menerima input suara dari pengguna dan menampilkan hasil prediksi emosi secara langsung.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1.5.1 Manfaat Teoritis

1. Menambah wawasan dan kontribusi keilmuan dalam bidang Speech Emotion Recognition (SER), khususnya terkait penerapan algoritma Convolutional Neural Network (CNN) untuk klasifikasi emosi berbasis suara.
2. Memberikan pemahaman mengenai pemanfaatan fitur spektral berbasis Mel Frequency Cepstral Coefficient (MFCC), termasuk delta MFCC dan delta-delta MFCC, dalam meningkatkan performa sistem deteksi emosi suara.

3. Menjadi referensi dan bahan acuan bagi peneliti selanjutnya yang ingin mengembangkan, membandingkan, atau mengevaluasi metode deep learning dalam pendeteksian emosi manusia melalui sinyal suara.

1.5.2 Manfaat Praktis

1. Menghasilkan sebuah model deteksi emosi berbasis suara menggunakan CNN yang mampu mengklasifikasikan emosi manusia secara otomatis dengan tingkat akurasi yang baik.
2. Menjadi dasar pengembangan aplikasi deteksi emosi berbasis suara yang dapat diterapkan pada berbagai bidang, seperti layanan pelanggan (call center), asisten virtual, sistem e-learning, serta pendukung analisis kondisi psikologis atau kesehatan mental.
3. Memberikan gambaran implementasi nyata model deep learning untuk pengenalan emosi suara dalam bentuk aplikasi berbasis web menggunakan framework Streamlit, sehingga dapat dimanfaatkan sebagai media pembelajaran maupun pengembangan sistem serupa di masa mendatang.

1.6 Sistematika Penulisan

Adapun sistematika penulisan laporan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN, Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA, Bab ini membahas penelitian terdahulu, teori pendukung mengenai emosi, pengolahan sinyal suara, ekstraksi fitur (MFCC, chroma, dan lain-lain), serta teori algoritma Support Vector Machine (SVM).

BAB III METODE PENELITIAN, Bab ini menjelaskan metode yang digunakan dalam penelitian, meliputi tahapan penelitian, dataset, preprocessing, ekstraksi fitur, pemilihan parameter, perancangan model SVM, serta metode evaluasi.

BAB IV HASIL DAN PEMBAHASAN, Bab ini berisi hasil implementasi sistem, visualisasi proses ekstraksi fitur, hasil pelatihan dan pengujian model, analisis performa model, serta pembahasan terhadap temuan penelitian.

BAB V PENUTUP, Bab ini memuat kesimpulan dari penelitian serta saran untuk pengembangan penelitian selanjutnya.

