

BAB I

PENDAHULUAN

A. Latar Belakang

Komunikasi memiliki hubungan yang sangat erat dengan aktivitas manusia. Dalam kehidupan sehari-hari, manusia selalu berinteraksi dengan orang lain untuk memenuhi berbagai kebutuhannya. Selama proses interaksi tersebut, komunikasi menjadi bagian yang tak terpisahkan. Komunikasi sendiri dapat diartikan sebagai segala bentuk perilaku seseorang, baik yang dilakukan secara lisan atau verbal maupun melalui cara non-verbal, yang kemudian memunculkan respons dari individu lain. Salah satu aspek penting dalam komunikasi adalah emosi, yang tidak hanya memengaruhi cara seseorang menyampaikan pesan, tetapi juga bagaimana pesan tersebut diterima dan dipahami oleh orang lain [1].

Emosi merupakan aspek penting dalam komunikasi manusia yang mencerminkan perasaan dan kondisi psikologis seseorang. Kemampuan untuk mengenali emosi memiliki peran signifikan dalam berbagai bidang, seperti psikologi, pendidikan, kesehatan mental, hingga interaksi manusia-mesin. Salah satu metode efektif untuk mendeteksi emosi adalah melalui analisis suara, karena suara dapat menyampaikan informasi secara verbal dan non-verbal [1].

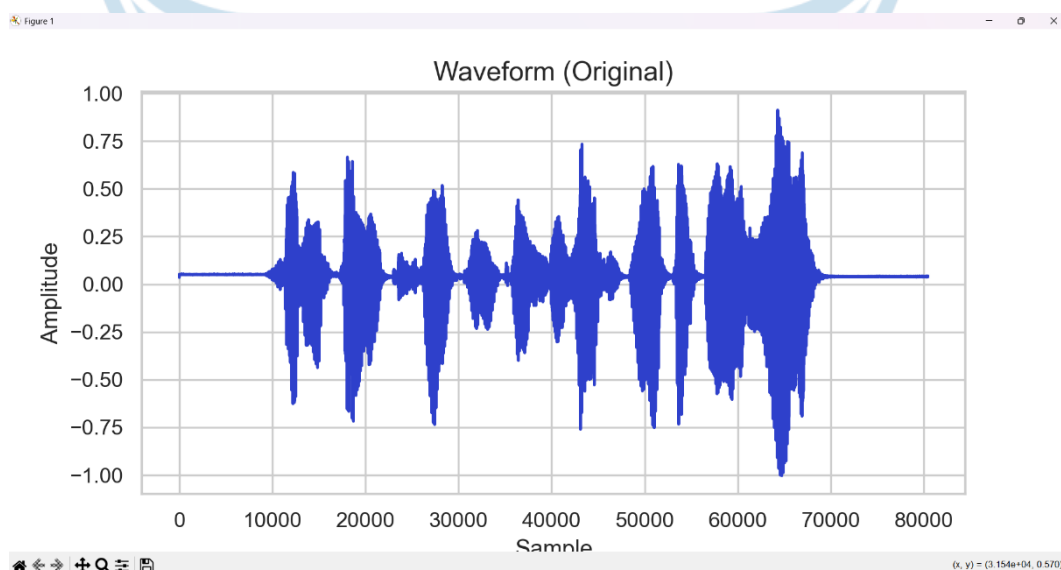
Salah satu aspek utama dalam analisis suara adalah bagaimana emosi disampaikan melalui intonasi dan nada. Menurut penelitian Mehrabian dan Wiener pada tahun 1967 yang tercantum pada penelitian [2] menunjukkan bahwa nada suara sering lebih penting daripada kata-kata dalam komunikasi. Namun, karena emosi dalam suara melibatkan banyak elemen, sulit untuk dijelaskan secara tepat. Dalam mempelajari emosi, fokusnya ada pada dua hal: bagaimana menggambarkan emosi, dan bagaimana emosi diekspresikan dan dipahami manusia [2].

Untuk memahami emosi yang muncul dalam suara, emosi dapat diklasifikasikan ke dalam enam kategori dasar *anger* (kemarahan), *disgust* (jijik), *fear* (ketakutan), *happiness* (kebahagiaan), *sadness* (kesedihan), dan *surprise* (terkejut). Keenam emosi ini bersifat universal, dikenali di berbagai budaya melalui intonasi suara, atau respons fisiologis, dan menjadi landasan penting

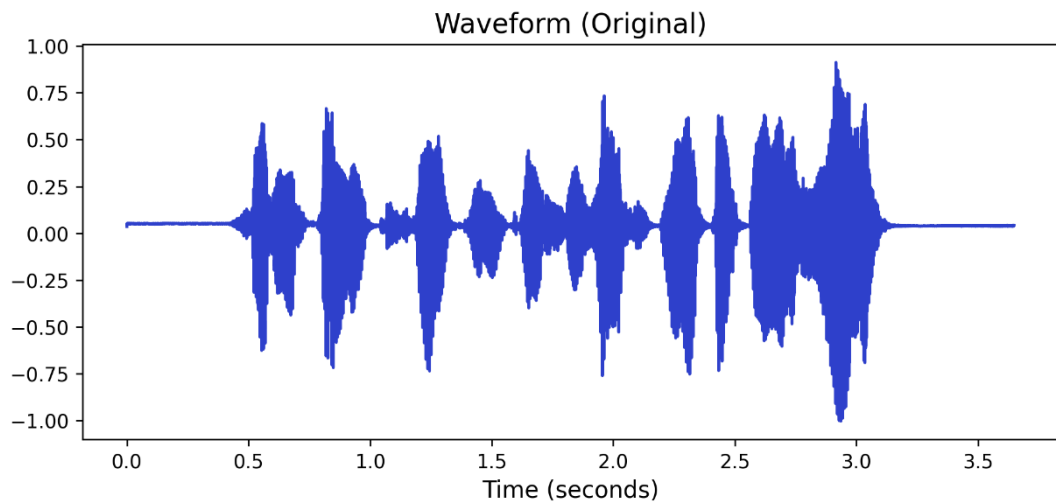
dalam penelitian tentang pengenalan emosi, psikologi, dan interaksi manusia-mesin [3].

Untuk memahami bagaimana emosi diekspresikan dalam suara, dapat dilihat dari dua aspek utama, yaitu komponen verbal dan non-verbal vokal. Verbal sendiri merujuk pada jumlah total kata yang diucapkan, pemilihan kata dan kata-kata emosional. Sedangkan non-verbal vokal merujuk pada berbagai parameter akustik, termasuk frekuensi (*pitch*), amplitudo (energi), dan parameter waktu, yang memberikan wawasan penting tentang ekspresi emosional.

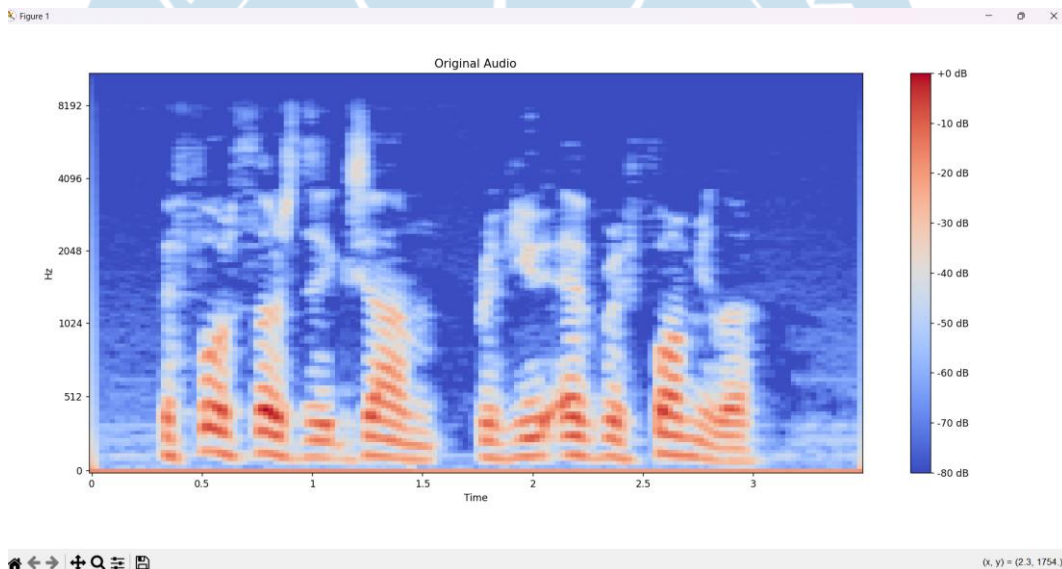
Parameter waktu meliputi durasi vokal (dalam detik), durasi bicara tanpa jeda, serta jumlah dan durasi jeda yang terjadi, seperti ditunjukkan pada Gambar 1.2. Parameter *pitch* mencakup nilai rata-rata dan standar deviasi *pitch* (Hz), frekuensi fundamental rata-rata (F0, Hz), serta nilai *pitch* minimum dan maksimum (Hz), sebagaimana terlihat pada Gambar 1.3. Sementara itu, parameter amplitudo mencakup rata-rata energi (dalam dB), standar deviasi energi (dB), serta energi minimum dan maksimum (dB), seperti pada Gambar 1.1. Ketiga kelompok parameter ini sering digunakan dalam studi akustik untuk memahami dan mengevaluasi aspek nonverbal dari komunikasi vokal [4].



Gambar 1. 1 *Waveform* suara dengan sumbu-Y menunjukkan amplitudo, dan sumbu-X menunjukkan jumlah sampel.



Gambar 1. 2 *Waveform* suara dengan sumbu-Y menunjukkan amplitudo, dan sumbu-X menunjukkan waktu (dalam detik).



Gambar 1. 3 *Spectrogram* suara dengan sumbu-Y menunjukkan frekuensi (Hz), sumbu-X menunjukkan waktu (dalam detik), dan warna menunjukkan energi dari frekuensi.

Untuk mengatasi kompleksitas pengenala emosi dalam suara, penelitian ini menggunakan pendekatan *Deep Learning* dalam pengenalan emosi melalui suara. Model seperti *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) menjadi relevan karena keunggulannya dalam menangkap pola data kompleks. CNN efektif dalam menganalisis fitur spasial dari data spektrogram [5] dan juga dapat digunakan untuk mengekstraksi pola dari audio *waveform* menggunakan *1D convolution* [6], sedangkan LSTM dan arsitektur RNN lainnya unggul dalam memahami pola data sekuensial seperti suara [7].

Dengan pendekatan ini, klasifikasi emosi berbasis suara tidak hanya bertujuan untuk mendeteksi emosi dengan akurasi tinggi tetapi juga dapat digunakan untuk aplikasi praktis, seperti mendukung diagnosis kesehatan mental.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah dalam penelitian ini adalah bagaimana menentukan kombinasi model arsitektur, metode scaling, dan optimizer yang paling optimal dalam meningkatkan akurasi pada tugas klasifikasi emosi berdasarkan data suara.

C. Batasan Masalah

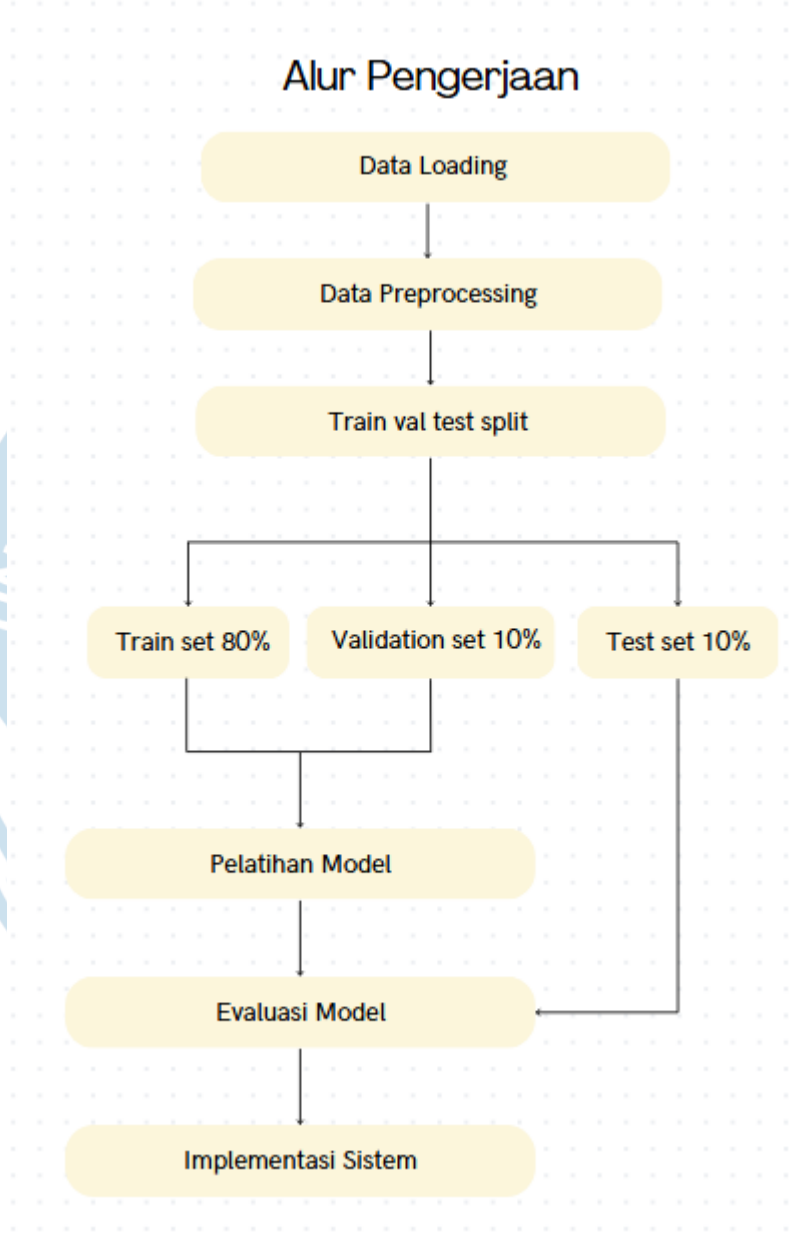
Pada penelitian ini, terdapat beberapa batasan lingkup atau batasan agar tidak terdapat penyimpangan dari rumusan masalah yang dibuat. Berikut merupakan lingkup masalah:

1. Jenis emosi yang akan di klasifikasi hanya empat emosi, yaitu: netral, senang, sedih dan marah.
2. Jenis suara yang digunakan adalah suara manusia.
3. Sumber data yang digunakan adalah Ravdess, Crema-D, Savee dan Tess yang bersifat *open source (kaggle)*.
4. Model klasifikasi yang digunakan adalah CNN, LSTM, BiLSTM dan GRU.
5. *Scaler* yang digunakan adalah *Z-score normalization* atau *standardscaler*, *Min-max* dan *Robust*.
6. *Optimizer* yang digunakan adalah Adam, SGD dan RMSprop.
7. Jumlah *training epochs* ditetapkan sebanyak 50 untuk setiap percobaan, menyesuaikan dengan keterbatasan sumber daya komputasi yang tersedia.
8. Jumlah *Batch* tiap *epochs* ditetapkan sebanyak 32 untuk setiap percobaan, menyesuaikan dengan keterbatasan sumber daya komputasi yang tersedia.
9. Dalam pengujian Sistem, sumber data adalah IEMOCAP yang bersifat *open source (kaggle)*.

D. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan model berbasis *Deep Learning* yang mampu mengklasifikasikan emosi berdasarkan suara manusia dengan performa yang optimal. Untuk mencapai tujuan tersebut, dilakukan serangkaian evaluasi performa berbagai arsitektur model *deep learning* (seperti CNN, LSTM, GRU dan BiLSTM), serta pengujian pengaruh dari metode *feature scaling* (Min-maxScaler, *Z-score normalization* atau standardscaler dan RobustScaler) dan algoritma *optimizer* (Adam, SGD dan RMSprop) dalam meningkatkan akurasi klasifikasi. Hasil akhir dari penelitian ini diharapkan dapat mengidentifikasi kombinasi model, metode *scaling* dan *optimizer* yang paling optimal.

E. Metode Penelitian



Gambar 1. 4 alur pengerjaan

1. *Data loading*

Pada tahap ini data yang sudah melalui proses filtrasi akan dimuat ke dalam sistem.

2. *Preprocessing Data*

Preprocessing data merupakan tahap penting yang dilakukan

sebelum proses pelatihan model. Pada tahap ini, dilakukan beberapa langkah utama untuk memastikan bahwa data yang digunakan benar-benar siap dan relevan untuk pelatihan model *deep learning*.

Langkah pertama dimulai dengan dilakukan proses augmentasi data, yaitu penambahan variasi buatan pada data suara yang ada tanpa perlu menambah data baru. Teknik ini berguna untuk meningkatkan keragaman data dan mengurangi risiko *overfitting*. Beberapa metode augmentasi yang digunakan dalam penelitian ini antara lain *pitch shifting*, penambahan *noise*, *time stretching*, *time shifting*, dan perubahan kecepatan (*change speed*).

Selanjutnya dilakukan ekstraksi fitur dari data suara. Proses ini bertujuan untuk mengubah sinyal suara mentah menjadi representasi numerik yang lebih bermakna dan informatif bagi model. Fitur-fitur yang diekstrak dalam penelitian ini meliputi MFCC (*Mel-Frequency Cepstral Coefficients*), ZCR (*Zero Crossing Rate*), dan RMS (*Root Mean Square Energy*), yang masing-masing menggambarkan karakteristik penting dari sinyal suara seperti nada, energi, dan struktur frekuensi.

3. *Train validation test split*

Data yang telah melalui proses *preprocessing data* dibagi ke dalam tiga bagian: *train set* (80%), *validation set* (10%), dan *test set* (10%). Seluruh data kemudian disimpan dalam format *.npz* agar dapat langsung digunakan pada saat pelatihan model, tanpa perlu dilakukan pemrosesan ulang.

4. Pelatihan Model

Model dilatih menggunakan optimasi seperti Adam atau RmsPprop dengan fungsi loss categorical cross-entropy. Selama proses pelatihan, performa model dipantau pada set validasi. Setelah pelatihan selesai, model dievaluasi pada set pengujian menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score* untuk mengukur akurasi

klasifikasinya.

5. Evaluasi Model

Setelah pelatihan selesai, model dievaluasi menggunakan data *test set* untuk mengukur efektivitasnya dalam mengklasifikasikan emosi. Potensi kendala, seperti ketidakseimbangan data dan kebutuhan komputasi tinggi, diantisipasi dengan menerapkan teknik augmentasi data untuk meningkatkan keragaman input serta memanfaatkan GPU guna mempercepat proses pelatihan.

6. Implementasi Sistem

Setelah melakukan evaluasi berbagai model untuk menentukan performa terbaik, tahap selanjutnya adalah merancang dan membangun tampilan antarmuka website menggunakan Streamlit. Website ini berfungsi sebagai media interaktif di mana pengguna dapat mengunggah input suara dalam format .wav untuk kemudian diproses oleh model terbaik yang telah dipilih. Hasil klasifikasi emosi akan ditampilkan secara langsung di halaman tersebut. Selain fitur input dan output suara, website juga dilengkapi dengan dokumentasi singkat mengenai proses pengembangan model, mulai dari tahap preprocessing, pelatihan, hingga evaluasi, sehingga pengguna dapat memahami latar belakang dan cara kerja sistem yang dibangun.

6. Sistematika Penulisan

BAB I Pendahuluan

Pada bab ini akan berisikan mengenai latar belakang proyek, rumusan masalah, Batasan penelitian, tujuan penelitian, metode penelitian dan sistematika penulisan.

BAB II Tinjauan Pustaka

Pada bab kedua ini akan berisikan penelitian-penelitian sebelumnya yang memiliki keterkaitan dengan penelitian yang dilakukan pada saat ini.

BAB III Landasan Teori

Pada bab ketiga, terdapat penjelasan mengenai kumpulan teori yang akan digunakan selama penelitian yang akan dilakukan. Landasan teori ini akan

menjadi panduan bagi penulis dalam menyelesaikan masalah yang ada dalam penelitian ini.

BAB IV Dataset dan Pengembangan Model

Pada bab keempat, Dataset dan pengembangan dari model.

BAB V Implementasi Model dan Pengujian Sistem

Pada bab kelima akan berisikan Implementasi model dan pengujian sistem.

