

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Pada penelitian ini, dataset suara berbasis emosi akan dikembangkan secara mandiri untuk memastikan kesesuaian dengan konteks Bahasa Indonesia. Proses pembuatan dataset dimulai dengan menentukan jenis emosi yang akan dideteksi, seperti marah, senang, sedih, dan netral, serta jumlah data yang diperlukan untuk menjaga keseimbangan dataset. Rekaman suara dilakukan dengan mengambil rekaman audio dari konten media social, dengan hasil rekaman disimpan dalam format standar seperti .wav. Setiap rekaman kemudian diberi label sesuai emosi yang diinginkan. Validasi pelabelan juga dilakukan untuk memastikan akurasi dan kualitas dataset yang dihasilkan.

Penelitian ini dilakukan dalam beberapa tahapan utama, mulai dari penyusunan laporan hingga evaluasi data. Setiap tahapan memiliki waktu pelaksanaan dan lokasi yang direncanakan untuk memastikan penelitian berjalan secara sistematis dan terstruktur. Jadwal tahapan penelitian beserta waktu dan lokasinya dapat dilihat pada Tabel 2 berikut:

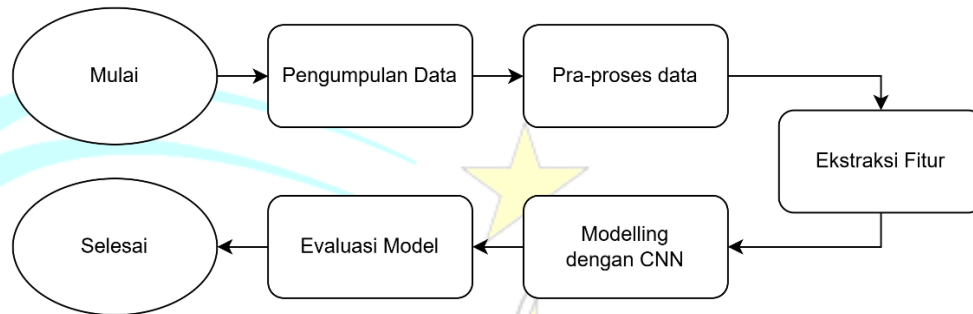
Tabel 1. Jadwal Tahapan Penelitian

No	Kegiatan	Bulan 1				Bulan 2				Bulan 3				Bulan 4			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengumpulan Data	■	■	■	■	■	■	■	■								
2	Pra-proses data					■	■	■	■								
3	Ekstraksi Fitur									■	■	■	■				
4	Modelling													■	■	■	■
5	Evaluasi Data													■	■	■	■

3.2 Prosedur Penelitian

Penelitian ini dirancang dengan tahapan-tahapan yang sistematis dan berurutan untuk mendeteksi emosi pelanggan berbasis suara menggunakan

model *Convolutional Neural Network* (CNN). Tahapan tersebut meliputi pengumpulan dan praproses data, pengolahan data, serta pengujian dan evaluasi model. Diagram alur penelitian ditampilkan pada Gambar 2, dengan penjelasan setiap tahapan berikut ini:



Gambar 1. Alur Penelitian

3.2.1 Pengumpulan Data

Pada tahap ini, data suara dikumpulkan dari berbagai sumber untuk memastikan keberagaman data yang representatif. Sumber data dapat berasal dari podcast berbahasa Indonesia, dan rekaman audio dari konten di sosial media. Data yang diperoleh sering kali memiliki format yang berbeda, sehingga dilakukan konversi ke format .wav. Format ini dipilih karena kompatibilitasnya yang tinggi dengan berbagai alat dan teknik pemrosesan suara, sehingga mempermudah proses analisis pada tahap berikutnya. Rincian dataset dapat dilihat pada Tabel 3.

Tabel 2. Rincian Dataset

kategori	path
marah	.../datasets/marah/audio_1.wav
marah	.../datasets/marah/audio_2.wav
marah	.../datasets/marah/audio_3.wav
...	...
netral	.../datasets/netral/audio_4.wav
netral	.../datasets/netral/audio_5.wav

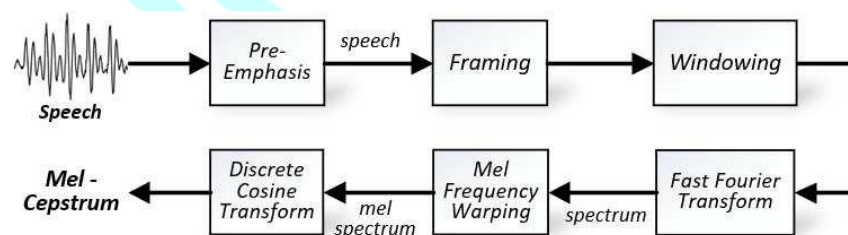
3.2.2 Pra-proses Data

Tahapan ini bertujuan untuk mempersiapkan data suara agar siap digunakan dalam proses analisis dan pelatihan model. Langkah-langkah pra-proses data meliputi:

- 1) Penambahan *White Noise*: Teknik ini menambahkan gangguan acak pada data suara untuk mensimulasikan kondisi dunia nyata dan meningkatkan daya tahan model terhadap noise.
- 2) Modifikasi *Speed and Pitch*: Variasi kecepatan dan nada suara diterapkan untuk menciptakan data tambahan secara sintetis tanpa mengubah karakteristik emosi.
- 3) *Normalization*: Amplitudo data suara diselaraskan untuk memastikan bahwa semua sampel memiliki skala yang seragam, sehingga model dapat mempelajari pola tanpa bias dari perbedaan volume.

3.2.3 Ekstraksi Fitur

Setelah data selesai diproses, langkah berikutnya adalah ekstraksi fitur menggunakan metode *Mel-Frequency Cepstral Coefficient* (MFCC). MFCC digunakan untuk mengubah data suara menjadi representasi numerik yang mencerminkan karakteristik frekuensi suara manusia.



Gambar 2. Alur Ekstraksi Fitur MFCC

Langkah langkah ekstraksi fitur MFCC yang ditunjukkan pada gambar 3 adalah seperti ini:

- 1) *Pre-Emphasis* sebagai langkah pertama bertujuan untuk meningkatkan efisiensi analisis spektral pada sinyal suara

- 2) *Framing* adalah proses merubah sinyal menjadi segmen – segmen untuk proses analisis spektral dan cepstral
- 3) *Windowing* adalah metode untuk mengkoreksi kebocoran sinyal akibat proses *framing*
- 4) *Fast Fourier Transform* (FFT) adalah metode untuk mengubah data dari domain waktu ke domain frekuensi dan sebaliknya. Metode FFT dengan cepat melakukan proses transformasi. Penerapan FFT secara tidak langsung menghasilkan informasi tentang jumlah energi di setiap pita frekuensi.
- 5) *Mel Frequency Warping* merupakan proses penyingkapan frekuensi untuk setiap nada dalam *Hertz* (Hz) pada skala yang disebut "*Mel scale*".
- 6) Langkah terakhir adalah mengembalikan nilai komputasi dari domain frekuensi ke domain waktu. Nilai terakhir yang diperoleh adalah *Mel-Cepstrum*, yang dikonversi dengan metode *Discret-Cosine-Transform*, sebagai metode kebalikan dari *Fourier Transform*. Jumlah koefisien DCT dimasukkan untuk memperoleh *Mel-Cepstrum*.

Fitur ini kemudian menjadi input utama bagi model CNN untuk mengenali pola-pola emosi.

3.2.4 *Modelling*

Pada tahap ini, proses inti dari penelitian dilakukan dengan merancang, mengimplementasikan, dan melatih model *deep learning* berbasis *Convolutional Neural Network* (CNN). CNN dipilih karena kemampuannya dalam menangkap pola spasial dari fitur audio yang telah diekstraksi menggunakan *Mel-Frequency Cepstral Coefficients* (MFCC).

Tahapan *modelling* dimulai dengan proses *feature engineering*, di mana fitur MFCC diekstraksi dari sinyal audio untuk menghasilkan representasi data dalam bentuk matriks dua dimensi. Matriks ini kemudian digunakan sebagai

input ke dalam model CNN. Arsitektur model dirancang dengan beberapa lapisan utama sebagai berikut:

- 1) *Input Layer*: Data hasil ekstraksi MFCC dimasukkan ke dalam *layer input* sebagai representasi fitur audio.
- 2) *Convolutional Layers*: Beberapa lapisan konvolusi diterapkan untuk mengekstraksi pola spasial dari fitur audio, diikuti oleh fungsi aktivasi ReLU untuk menangkap hubungan *non-linear*.
- 3) *Pooling Layers*: Lapisan pooling digunakan untuk mengurangi dimensi data dan menghindari *overfitting*.
- 4) *Fully Connected Layers*: Lapisan *fully connected* berfungsi untuk menggabungkan fitur-fitur yang telah dipelajari dan memetakan ke dalam kategori emosi.
- 5) *Output Layer*: Menggunakan fungsi aktivasi *softmax* untuk menghasilkan probabilitas dari setiap kelas emosi.

Data yang digunakan dibagi menjadi dua subset:

- a. *Data Training*: Untuk melatih model agar dapat mengenali pola emosi.
- b. *Data Testing*: Untuk mengevaluasi kinerja model pada data yang tidak pernah dilihat sebelumnya.

Proses *training* model dilakukan dengan optimasi *hyperparameter*, termasuk pengaturan *learning rate*, jumlah *epoch*, dan ukuran *batch*, untuk mencapai performa terbaik. Evaluasi model dilakukan menggunakan *accuracy* dan *loss*. *accuracy* untuk mengukur tingkat keberhasilan model dalam memprediksi label yang benar, kemudian *loss*, untuk mengevaluasi tingkat kesalahan selama proses pelatihan dan pengujian.

3.2.5 Evaluasi Model

Evaluasi dilakukan untuk mengukur kinerja model dalam mendeteksi emosi suara. Metode evaluasi mencakup tiga metrik utama:

- 1) *Accuracy* menunjukkan persentase prediksi yang benar terhadap seluruh data uji, memberikan gambaran umum tentang kemampuan model dalam mengklasifikasikan emosi dengan benar. Persamaan untuk menghitung *accuracy* dapat dilihat pada Persamaan (5):

$$Accuracy = \left(\frac{TP + TN}{TP + FP + TN + FN} \right) \times 100\% \quad (5)$$

- 2) *Loss* Digunakan untuk mengukur seberapa jauh prediksi model dari nilai sebenarnya. Nilai *loss* dihitung selama proses pelatihan dan pengujian untuk memberikan gambaran tentang tingkat kesalahan model pada setiap *epoch*. Persamaan untuk menghitung *loss* dapat dilihat pada Persamaan (6):

$$Loss = \frac{1}{N} \sum_{i=1}^N L(y_i, \hat{y}_i) \quad (6)$$

Di mana L adalah fungsi *loss* seperti *Mean Squared Error* (MSE) atau *Cross-entropy*, y_i adalah nilai sebenarnya, dan \hat{y}_i adalah prediksi model.

- 3) *Confusion Matrix* digunakan untuk mengevaluasi performa model dengan memberikan gambaran detail tentang prediksi yang benar dan salah untuk setiap kelas emosi. Matriks ini berisi informasi terkait jumlah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dengan menggunakan *confusion matrix*, model dapat dianalisis lebih mendalam untuk melihat pola kesalahan dalam klasifikasi, seperti apakah model sering salah mengklasifikasikan satu emosi menjadi emosi lainnya. Dari hasil *confusion matrix*, *accuracy*, *precision*, *recall*, dan *f1-score* dihitung untuk memberikan analisis yang lebih komprehensif.

Hasil evaluasi dari ketiga metrik ini digunakan untuk menilai keberhasilan model secara keseluruhan, baik dari segi generalisasi pada data baru maupun kesesuaian prediksi dengan label sebenarnya.