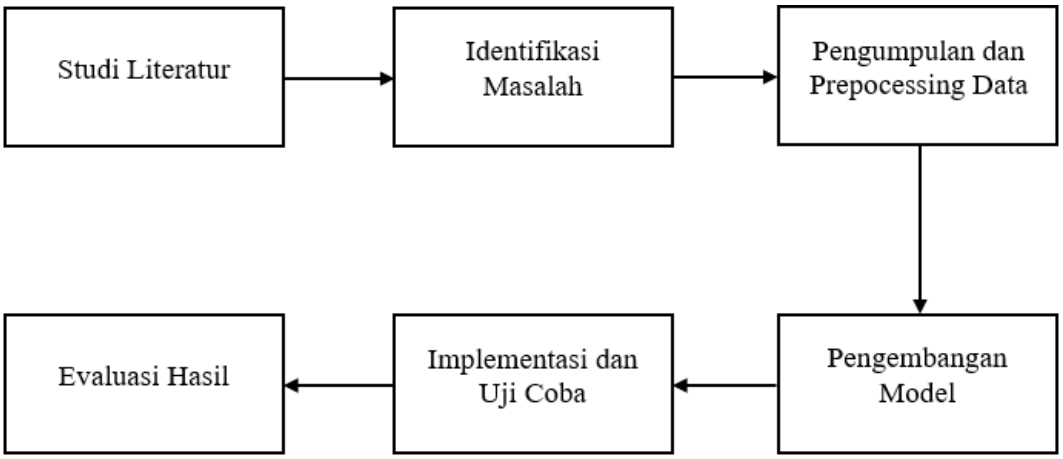
BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Metodologi penelitian diperlukan sebagai kerangka dan panduan dalam melakukan proses penelitian, sehingga penelitian yang dilakukan menjadi lebih terarah, teratur, dan sistematis (Harman, 2019). Berikut merupakan beberapa tahapan pada penelitian ini dapat dilihat pada Gambar 3.2 berikut ini.



Gambar 3.1 Tahapan Penelitian

3.1.1 Studi Literatur

Studi literatur dilakukan dengan melakukan analisa terhadap beberapa penelitian terkait, dengan mengkaji dari beberapa sumber seperti buku, jurnal, serta laporan penelitian yang berkaitan dengan klasifikasi emosi berdasarkan suara, klasifikasi suara, dan algoritma *Convolutional Neural Network*. Tahap ini bertujuan untuk mengetahui aspek teoritis dan aspek praktis.

3.1.2 Identifikasi Masalah

Identifikasi masalah bertujuan untuk mendefinisikan permasalahan yang akan diteliti dan permasalahan ini diuraikan pada bagian latar belakang.

3.1.3 Pengumpulan dan Preprocessing Data

Pengumpulan data bertujuan untuk memperoleh informasi yang dibutuhkan guna mencapai tujuan penelitian (Afrianto, 2020). Sumber data dalam penelitian ini adalah dataset yang berasal dari Kaggle atau *Crowd-sourced Emotional Multimodal Actors Dataset* (CREMA-D). Dataset ini merupakan dataset berlabel yang digunakan untuk mempelajari pengenalan ekspresi dan emosi. Data kemudian diolah agar dapat dimasukkan ke dalam model dengan lebih baik, sehingga meningkatkan performa model yang dibangun. Pengolahan data ini juga mencakup ekstraksi fitur untuk menangkap sifat spektral dari sinyal audio, yang penting untuk analisis emosi.

3.1.4 Keterkaitan Suasana Hati

Dalam penelitian ini, suasana hati (*mood*) memiliki peran penting dalam klasifikasi emosi dari suara. Suasana hati seseorang dapat mempengaruhi cara individu mengekspresikan emosi melalui intonasi, nada, dan tempo bicara. Suasana hati yang positif cenderung memunculkan intonasi yang lebih ceria, tempo bicara yang lebih cepat, serta nada suara yang lebih tinggi dan dinamis. Sebaliknya, suasana hati negatif biasanya ditandai dengan intonasi yang lebih datar, tempo yang lambat, serta nada yang lebih rendah dan monoton. Oleh karena itu, dalam proses *preprocessing data* khususnya pada tahap ekstraksi fitur, ciri-ciri akustik seperti tinggi nada (*pitch*), intensitas suara (*loudness*), kualitas suara (timbre), dan pola

tempo dianalisis secara khusus untuk membedakan emosi yang terpengaruh oleh suasana hati jangka panjang dibandingkan emosi intens yang muncul secara spontan.

3.1.5 Pengembangan Model

Pengembangan model dilakukan dengan merancang model klasifikasi untuk selanjutnya dilakukan pelatihan model menggunakan dataset yang sudah diolah dan memastikan bahwa model tersebut dapat mengidentifikasi emosi berdasarkan suara dengan baik.

3.1.6 Implementasi dan Uji Coba

Menerapkan model yang sudah dilatih dan melakukan uji coba terhadap model untuk memastikan bahwa model tersebut dapat mengidentifikasi emosi berdasarkan suara dengan baik.

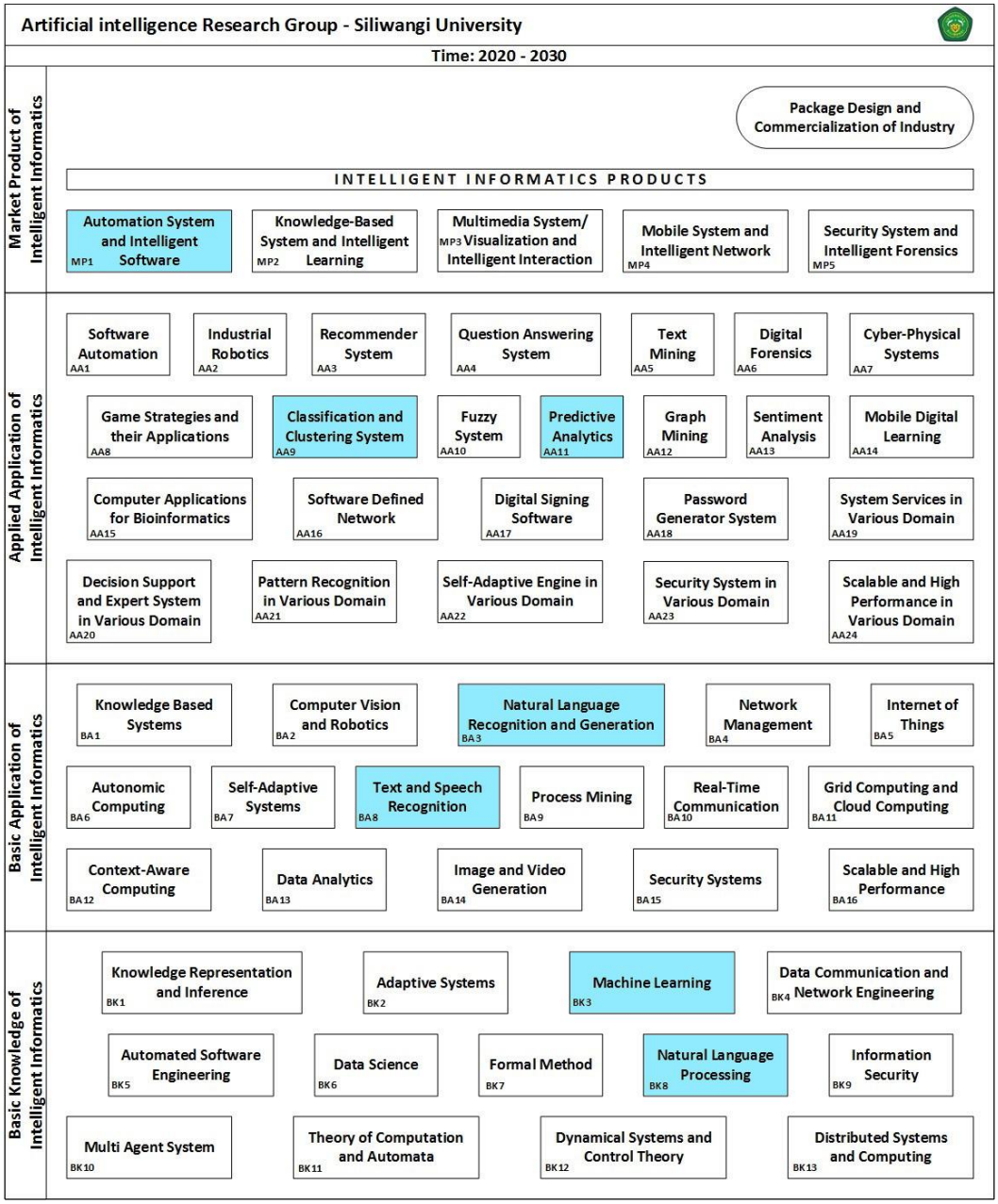
3.1.7 Evaluasi Hasil

Evaluasi hasil meliputi proses menguji performa dari algoritma *Convolutional Neural Network* menggunakan *confusion matrix* dan memberikan gambaran hasil klasifikasi emosi berbasis suara.

3.2 Peta Jalan (*Road Map*) Penelitian

Peta jalan atau roadmap adalah sebuah konsep pengaturan arah dalam penelitian yang bertujuan untuk menguraikan perjalanan dan fokus yang akan diambil dalam pengembangan kecerdasan buatan (*The Purpose and Goal of a Roadmap: A Comprehensive Guide*, 2024). Penelitian ini merujuk pada roadmap yang dikembangkan oleh Kelompok Keahlian Informatika dan Sistem Inteligen

(KK-ISI) dari Universitas Siliwangi untuk rentang waktu 2020 hingga 2030 yang ditunjukkan pada gambar 3.2.



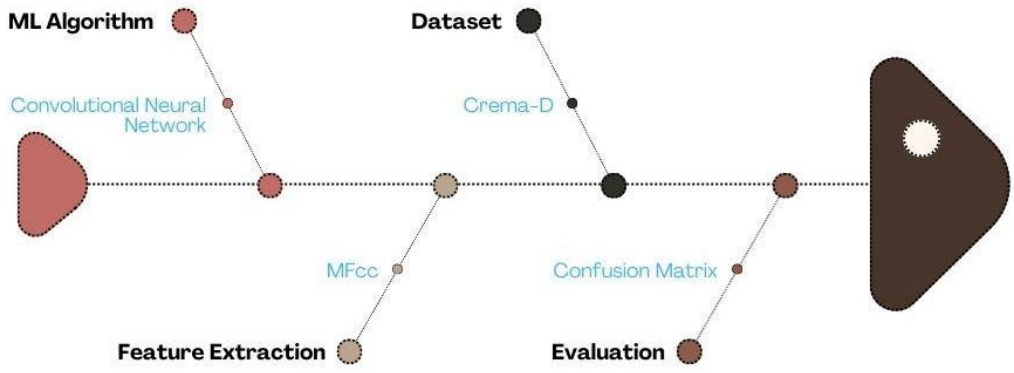
Gambar 3.2 Roadmap Penelitian (KK-ISI Universitas Siliwangi, 2020)

Berdasarkan Gambar 3.2 penelitian ini melibatkan beberapa kajian utama yang ditandai dengan kolom berwarna biru. Kajian tersebut mencakup *Automation*

Systems and Intelligent Software, Classification and Clustering System, Predictive Analytics, Text and Speech Recognition, Natural Language Recognition and Generation, serta *Machine Learning* dan *Natural Language Processing*. Penelitian yang berjudul "Klasifikasi Emosi Berdasarkan Suara pada Dataset CREMA-D Menggunakan *Convolutional Neural Network*" mengadopsi metodologi terstruktur untuk menganalisis klasifikasi emosi dari suara. Penelitian ini dimulai dengan studi literatur, diikuti oleh pengumpulan dan *preprocessing data* menggunakan dataset CREMA-D dari Kaggle. Selanjutnya, penelitian mengembangkan model klasifikasi yang dilatih menggunakan dataset yang telah diolah, dan akhirnya mengimplementasikan serta menguji model tersebut menggunakan *Confusion Matrix* untuk menilai performanya.

3.3 Fishbone Diagram

Berikut merupakan diagram *fishbone* pada penelitian ini



Gambar 3.3 Fishbone Diagram (Kumah dkk., 2024)

Gambar 3.3 menunjukan beberapa faktor kunci pada penelitian yang ditandai dengan teks berwarna merah diantaranya yaitu *Feature Extraction, Dataset, ML Algorithm, Evaluation* Masing – masing bagian pada tulang memiliki keterangan sebagai berikut:

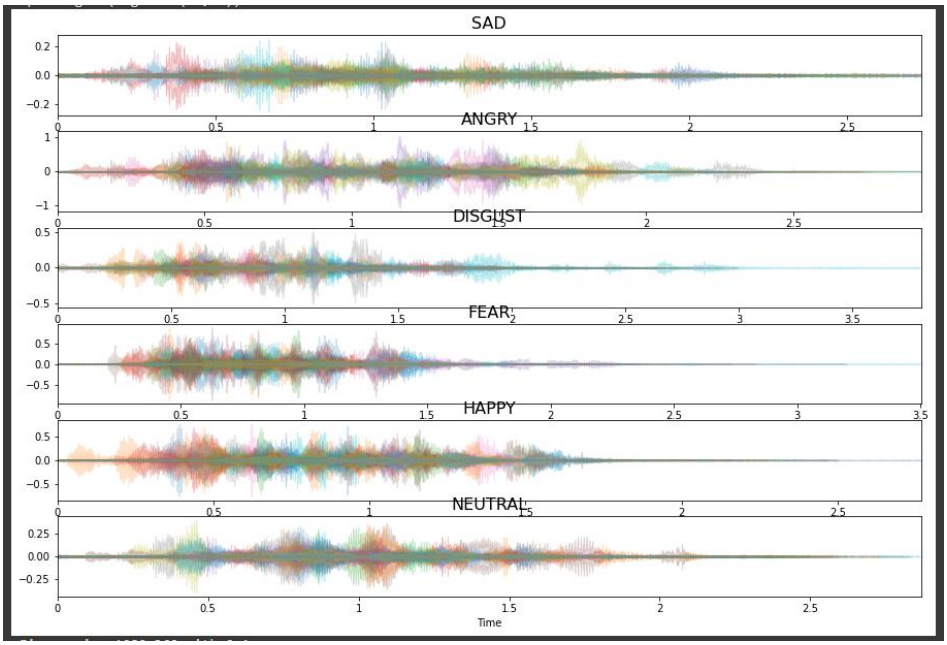
1) *Machine Learning Algorithm*

   Proses klasifikasi pada penelitian ini menggunakan model *Convolutional Neural Network*. Algoritma *Convolutional Neural Network* (CNN) dipilih karena kemampuannya dalam menangkap pola-pola kompleks dari data audio. CNN memiliki keunggulan dalam mengekstraksi fitur-fitur penting secara otomatis, sehingga sangat efektif dalam pengenalan emosi melalui sinyal audio. Arsitektur CNN pada penelitian ini terdiri dari beberapa *layer* konvolusional, *pooling*, dan *fully connected* yang menggunakan fungsi aktivasi ReLU dan softmax pada tahap akhir. Optimasi parameter dilakukan dengan menggunakan metode *gradient descent*.

2) Dataset

   CREMA-D (*Crowd-sourced Emotional Multimodal Actors Dataset*) adalah kumpulan rekaman audio yang telah dilabeli dengan emosi yang sesuai (Cao dkk., 2014). Dataset ini dikembangkan oleh para peneliti untuk tujuan penelitian dalam pengenalan emosi melalui suara. Dataset ini mencakup rekaman dari aktor yang mengekspresikan berbagai emosi dengan kalimat-kalimat tertentu. Dataset ini terdiri dari 7,442 klip audio dari 91 aktor yang mengekspresikan 12 kalimat target dengan enam emosi yang berbeda

(marah, jijik, takut, bahagia, netral, sedih) pada empat tingkat intensitas yang berbeda.



Gambar 3.4 Representasi Visual Berbagai Emosi (Ismail, 2023)

Gambar 3.4 menunjukkan representasi visual dari sinyal audio untuk beberapa emosi yang berbeda, yaitu S*ad, Angry*, *Disgust*, *Fear*, *Happy*, dan *Neutral*. Setiap grafik menggambarkan bentuk gelombang suara (*waveform*) yang menunjukkan perubahan amplitudo suara terhadap waktu. Untuk emosi Sad, sinyal suara terlihat lebih halus dengan fluktuasi yang stabil, mencerminkan pola bicara yang lebih lambat dan rendah energi. Emosi *Angry* menunjukkan perubahan amplitudo yang lebih besar dan tajam, mencerminkan intensitas suara yang lebih tinggi dengan energi yang kuat dan sering kali tidak teratur. *Disgust* memiliki bentuk gelombang yang mirip dengan *Angry*, namun dengan fluktuasi yang lebih teratur dan tidak sebesar pada *Angry*, mencerminkan suara yang keras tetapi lebih terkontrol. *Fear*

menampilkan pola gelombang yang lebih halus dan mirip dengan *Sad*, menunjukkan suara yang tertekan dan cemas, dengan sedikit lonjakan energi. Pada emosi *Happy*, terdapat variasi amplitudo yang lebih cerah dan teratur, menggambarkan percakapan yang lebih ceria dan penuh energi. Sementara itu, *Neutral* menunjukkan pola suara yang paling stabil dengan sedikit fluktuasi, mencerminkan suara yang datar dan tanpa perubahan emosi yang signifikan. Secara keseluruhan, gambar ini menggambarkan perbedaan karakteristik suara yang terkait dengan berbagai emosi.

3) *Feature Extraction*

Ekstraksi fitur audio merupakan tahapan penting karena CNN membutuhkan input yang terstruktur untuk mengenali pola emosi dari sinyal suara. Salah satu fitur utama yang digunakan adalah *Mel-Frequency Cepstral Coefficients* (MFCC). MFCC merupakan fitur yang mengekstraksi karakteristik spektral dari sinyal audio berdasarkan persepsi pendengaran manusia. Secara lebih spesifik, MFCC mengubah sinyal suara menjadi representasi numerik yang menangkap informasi penting.

Proses ekstraksi MFCC melibatkan beberapa tahapan, salah satunya adalah penggunaan *Fast Fourier Transform* (FFT). FFT adalah algoritma matematis yang digunakan untuk mengubah sinyal audio dari domain waktu menjadi domain frekuensi. Dengan FFT, sinyal audio kompleks dipecah menjadi komponen frekuensi penyusunnya, memungkinkan identifikasi spektrum frekuensi secara jelas. Selanjutnya, spektrum ini dipetakan ke dalam skala Mel, yaitu skala frekuensi yang mengikuti sensitivitas

pendengaran manusia. Proses ini menghasilkan koefisien MFCC yang merepresentasikan karakteristik utama sinyal audio dalam bentuk yang mudah diolah oleh CNN untuk klasifikasi emosi.

4) *Evaluation*

Evaluasi hasil dilakukan dengan tujuan menganalisis performa algoritma klasifikasi yang digunakan (Yoga Siswa, 2023). Proses analisis hasil dilakukan dengan menggunakan Confusion Matrix. Confusion matrix memberikan gambaran lengkap mengenai kinerja model dengan menunjukkan jumlah prediksi yang benar dan salah yang dibuat oleh model dibandingkan dengan nilai sebenarnya.

5) Tantangan Implementasi

Dalam proses implementasi penelitian ini, terdapat beberapa tantangan utama yang harus dihadapi. Salah satu tantangan terbesar adalah adanya *noise* pada data audio, yang dapat mengganggu proses ekstraksi fitur dan menurunkan akurasi model. *Noise* ini dapat berasal dari berbagai sumber, seperti kebisingan latar belakang atau kualitas perekaman yang buruk. Selain itu, variasi kualitas rekaman juga menjadi kendala, karena perbedaan dalam perangkat perekam atau kondisi lingkungan dapat menghasilkan data yang tidak konsisten. Tantangan lain yang cukup signifikan adalah keterbatasan daya komputasi, terutama ketika harus melatih model *Convolutional Neural Network* (CNN) pada dataset yang besar. Keterbatasan ini mempengaruhi waktu pelatihan dan kapasitas untuk mengoptimalkan model dengan kompleksitas tinggi.

6) Strategi Mengatasi *Overfitting*

Untuk mengatasi masalah *overfitting*, penelitian ini menerapkan beberapa teknik regulasi. Salah satunya adalah penggunaan *dropout*, di mana sejumlah neuron dalam jaringan di-nonaktifkan secara acak selama pelatihan untuk mencegah model terlalu bergantung pada pola tertentu. Selain itu, diterapkan pula regularisasi L2, yang berfungsi membatasi besarnya bobot dalam jaringan sehingga mendorong model untuk tetap sederhana dan menghindari kompleksitas yang berlebihan. Augmentasi data juga digunakan untuk memperkaya variasi data pelatihan, seperti dengan menambahkan noise buatan, melakukan *time shifting*, atau mengubah *pitch* dari rekaman audio. Teknik-teknik ini secara kolektif membantu meningkatkan generalisasi model, seperti yang dijelaskan dalam artikel *Tensorflow Core* yang berjudul *Overfit and Underfit*. Teknik-teknik ini secara kolektif membantu meningkatkan generalisasi model, seperti yang dijelaskan dalam artikel *Tensorflow Core* yang berjudul *Overfit and Underfit*., tetapi juga tetap handal pada data yang sebelumnya tidak terlihat.