BAB I
PENDAHULUAN

1.1 Latar Belakang Penelitian

Kemajuan teknologi kecerdasan buatan (*Artificial Intelligence*) dalam bidang pemrosesan suara telah menghasilkan inovasi yang signifikan salah satunya adalah kemampuan AI untuk meniru suara manusia secara sangat akurat melalui teknik *voice cloning* atau *deepfake audio*. Teknologi ini memungkinkan sebuah klip suara pendek saja untuk dikloning menjadi suara yang sangat menyerupai manusia asli, dengan intonasi, aksen, dan karakteristik suara yang hampir tak dapat dibedakan. Namun, keberhasilan ini memunculkan risiko baru yang nyata bagi keamanan digital dan integritas komunikasi[1]. Meskipun terdengar sangat natural, suara hasil rekayasa AI masih menyimpan perbedaan tertentu dibandingkan suara manusia asli, yang umumnya muncul pada aspek prosodi, karakteristik temporal, serta distribusi spektral suara, yang tidak selalu dapat ditiru secara sempurna oleh model *voice cloning* modern.

Di Indonesia, ancaman tersebut telah berkembang menjadi persoalan yang berdampak nyata secara ekonomi. *Indonesia Anti-Scam Centre* lembaga yang dibentuk atas inisiasi Otoritas Jasa Keuangan bersama Satuan Tugas Pemberantasan Aktivitas Keuangan Ilegal mencatat sebanyak 343.402 laporan penipuan transaksi keuangan dengan total kerugian mencapai Rp7,8 triliun sejak mulai beroperasi pada 22 November 2024 hingga 11 November 2025 [3]. Di antara modus yang digunakan, pemanfaatan kecerdasan buatan dalam bentuk *voice cloning* dan *deepfake* telah teridentifikasi sebagai salah satu instrumen penipuan yang digunakan para pelaku untuk menyamarkan identitas dan memanipulasi korban. OJK sendiri memperingatkan bahwa *voice cloning* dan *deepfake* suara telah menjadi salah satu modus utama kejahatan digital, seperti penipuan melalui telepon atau WhatsApp yang meniru suara teman, kolega, atau pejabat institusi keuangan untuk mengelabui korban agar melakukan transfer dana [2].

Kasus-kasus ini menunjukkan bahwa risiko bukan hanya dari aspek teknis namun juga dari aspek sosial dan ekonomi di mana individu maupun institusi dapat dirugikan secara signifikan karena rekaman suara yang tampak valid ternyata hasil rekayasa AI. Seperti disebutkan oleh pakar keamanan siber dari Palo Alto *Networks*

Indonesia, teknologi *deepfake suara* diprediksi akan menjadi ancaman siber yang semakin dominan pada 2025 karena kemudahan penggunaannya dan dampak ekonominya [4].

Dari sisi teknis, banyak penelitian terdahulu yang mencoba mengembangkan sistem deteksi audio palsu menggunakan metode tradisional seperti ekstraksi fitur MFCC dan algoritma *machine learning* klasik [5]. Namun pendekatan ini kerap menghadapi keterbatasan ketika diterapkan pada suara yang dihasilkan oleh model *TTS* atau *voice-cloning* generasi terbaru, yang memiliki artefak berbeda dan kompleksitas yang lebih tinggi. Maka dari itu pendekatan *Self-Supervised Learning (*SSL) muncul sebagai solusi yang menjanjikan salah satunya adalah model wav2vec 2.0 [6] yang mampu mempelajari representasi suara dari data mentah tanpa perlu anotasi besar. Varian lanjutannya, yaitu XLSR-Wav2Vec2 (*Cross-Lingual Speech Representation*) dapat menangani banyak bahasa sekaligus termasuk bahasa dengan sumber daya rendah [7], kemudian diklasifikasikan dengan *Convolutional Neural Network* (CNN) sebagai *classifier head* dua kelas (Indo-AD). CNN dipilih karena efektif dalam menangkap pola spasial dan lokal pada embedding akustik hasil *Self-Supervised Learning* (SSL) yang direpresentasikan dalam bentuk matriks fitur berdimensi tinggi, yang secara struktural menyerupai representasi spektral sinyal audio. CNN efektif dalam mendeteksi anomali akustik pada sinyal audio dibandingkan model linear biasa[8].

Meskipun berbagai upaya pengembangan sistem deteksi suara AI telah dilakukan, terdapat kesenjangan penelitian yang signifikan dalam studi-studi terdahulu. Sebagian besar penelitian masih sangat bergantung pada metode ekstraksi fitur konvensional seperti MFCC dan spektrogram yang dikombinasikan dengan algoritma *machine learning* klasik. Metode tersebut seringkali gagal menangkap artefak akustik halus pada suara sintetis generasi terbaru yang dihasilkan oleh teknologi *voice cloning* modern. Selain itu, fokus penelitian sebelumnya mayoritas masih tertuju pada bahasa berdaya sumber tinggi seperti Bahasa Inggris dan Arab, sementara eksplorasi pada Bahasa Indonesia sebagai *low-resource language* masih sangat terbatas[7].

Penelitian ini bertujuan untuk mengisi kesenjangan tersebut melalui pengembangan sistem deteksi *audio deepfake* berbahasa Indonesia dengan model

Indo-AD, Model ini dipilih karena kemampuannya dalam mempelajari pola akustik kompleks dari data mentah tanpa memerlukan anotasi dalam jumlah besar, serta dukungannya terhadap berbagai bahasa termasuk Bahasa Indonesia yang tergolong *low-resource language*. Maka dari itu, penulis mengajukan proposal penelitian berjudul **"Deteksi Suara Kecerdasan Buatan Berbahasa Indonesia Dengan *Cross-Lingual Speech Representation Wav2vec2* dan *Convolutional Neural Network*".** Pendekatan ini diharapkan tidak hanya menghasilkan sistem deteksi yang akurat dan efisien, tetapi juga memberikan kontribusi ilmiah dalam pengembangan teknologi keamanan suara berbasis *self-supervised learning* untuk Bahasa Indonesia, sekaligus menjadi pijakan awal bagi penelitian lanjutan di bidang deteksi suara sintetis.

1.2 Perumusan Masalah Penelitian

Bagaimana kinerja metode *Self-Supervised Learning* berbasis XLSR-Wav2Vec2 yang dikombinasikan dengan *Convolutional Neural Network* dalam mendeteksi suara manusia asli dan suara kecerdasan buatan berbahasa Indonesia?

1.3 Tujuan Penelitian

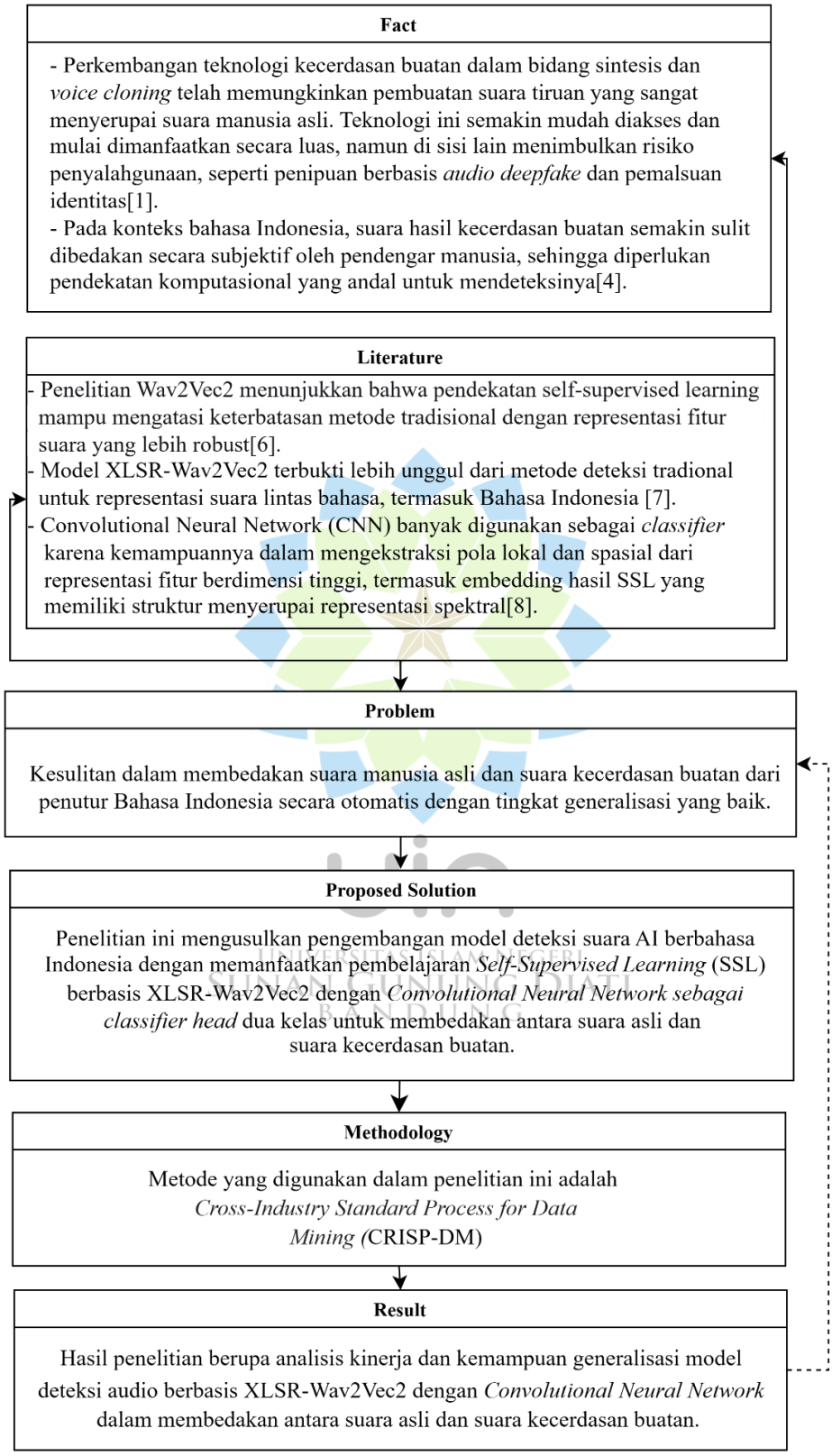
Menganalisis dan mengevaluasi kinerja metode *Self-Supervised Learning* XLSR-Wav2Vec2 yang dikombinasikan dengan *Convolutional Neural Network* dalam mendeteksi suara manusia asli dan suara kecerdasan buatan berbahasa Indonesia.

1.4 Batasan Masalah Penelitian

Penelitian ini dibatasi oleh beberapa ruang lingkup berikut:

1) Bahasa yang digunakan dalam penelitian ini dibatasi pada Bahasa Indonesia, baik untuk data suara asli maupun suara kecerdasan buatan.
2) Model *Self-Supervised Learning* yang digunakan dalam penelitian ini dibatasi pada arsitektur XLSR-Wav2Vec2 yang dikombinasikan dengan *Convolutional Neural Network* sebagai *classifier head* dua kelas.
3) Dataset yang digunakan terdiri dari suara manusia asli hasil perekaman langsung serta suara kecerdasan buatan yang dihasilkan melalui proses *voice cloning* menggunakan layanan publik, dengan sumber utama berasal dari platform *ElevenLabs*, *MinimaxAudio* dan *FishAudio* sehingga variasi jenis model kecerdasan buatan yang diuji masih terbatas.

1.5 Kerangka Pemikiran Penelitian



Gambar 1.1 Kerangka Pemikiran.

Kerangka pemikiran pada Gambar 1.1, ini menggambarkan hubungan antara fakta, literatur, permasalahan, solusi yang diusulkan, metodologi, hingga hasil akhir yang diharapkan. Dengan memahami alur ini, pembaca dapat melihat secara jelas bagaimana pendekatan penelitian diusulkan untuk menjawab permasalahan deteksi audio deepfake berbahasa Indonesia secara sistematis dan terukur.

Penelitian terdahulu banyak berfokus pada pengembangan sistem deteksi audio palsu menggunakan metode tradisional seperti ekstraksi fitur MFCC dan algoritma *machine learning* klasik [5]. Namun, pendekatan tersebut mulai menunjukkan keterbatasan ketika diterapkan pada suara sintetis generasi baru yang dihasilkan oleh teknologi *voice cloning* dan *Text-to-Speech* modern, karena artefak akustiknya semakin halus dan sulit dibedakan dari suara manusia asli. Literatur terbaru menunjukkan bahwa pendekatan *Self-Supervised Learning* (SSL), khususnya Wav2Vec2, mampu mengatasi keterbatasan tersebut melalui representasi fitur suara yang lebih *robust* dibandingkan metode konvensional [6]. Dengan Metode XLSR-Wav2Vec2, terbukti unggul secara empiris dalam menangani representasi suara lintas bahasa termasuk bahasa yang tergolong *low-resource* seperti Bahasa Indonesia [7].

Berdasarkan fakta dan tinjauan literatur tersebut, muncul permasalahan utama, yaitu bagaimana mendeteksi perbedaan antara suara asli manusia dan suara kecerdasan buatan dalam Bahasa Indonesia yang memiliki karakteristik fonetik spesifik dan belum banyak dieksplorasi dalam penelitian terdahulu. Perbedaan ini secara teoritis tercermin pada karakteristik prosodi, pola temporal, dan distribusi spektral suara, yang dapat dipelajari secara implisit melalui embedding hasil pendekatan *Self-Supervised Learning*.

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan pengembangan sistem deteksi suara kecerdasan buatan menggunakan pendekatan *Self-Supervised Learning* dengan model XLSR-Wav2Vec2, yang dikombinasikan dengan *Convolutional Neural Network* (CNN) sebagai *classifier head* dua kelas.

Tahapan penelitian dilakukan menggunakan metodologi CRISP-DM, yang meliputi pemahaman masalah, pemahaman data, persiapan data, pemodelan, evaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*, Confusion matrik hingga dokumentasi dan penyimpanan model terbaik.

Melalui pendekatan tersebut, penelitian ini diharapkan menghasilkan model deteksi *deepfake* suara yang akurat, efisien, dan adaptif terhadap variabilitas penutur Bahasa Indonesia. Selain berkontribusi terhadap penguatan keamanan digital berbasis audio di Indonesia, penelitian ini diharapkan menjadi fondasi ilmiah bagi pengembangan teknologi dan riset lanjutan di bidang deteksi suara sintesis.

1.6 Sistematika Penulisan

Sistematika penulisan laporan memuat sistematika penulisan laporan tugas akhir dengan memberikan gambaran kandungan setiap bab, urutan penulisan, serta keterkaitan antara satu bab dengan bab lainnya dalam sebuah laporan tugas akhir. Berikut sistematika penulisan laporan tugas akhir.

BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang, perumusan masalah, tujuan penelitian, batasan masalah, manfaat, kerangka pemikiran penelitian, dan sistematika penulisan.

BAB II KAJIAN LITERATUR

Pada bab ini membahas terkait literatur atau penelitian terdahulu, konsep konsep, teori-teori, model, dan rumus yang menjadi landasan dalam proses analisis permasalahan dengan topik masalah yang diambil.

BAB III METODOLOGI PENELITIAN

Metodologi penelitian berisi penjelasan langkah-langkah dan teknik yang diterapkan dalam penelitian, diuraikan secara sistematis dan terstruktur.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan dua hal utama, yang pertama adalah pemaparan tentang temuan atau hasil penelitian berdasarkan langkah-langkah penelitian yang telah dilakukan. Selanjutnya adalah pembahasan hasil atau temuan penelitian sebagai jawaban terhadap rumusan masalah penelitian.

BAB V SIMPULAN DAN SARAN

Bab ini berfokus pada penarikan kesimpulan dari hasil penelitian yang diperoleh serta menjawab pertanyaan penelitian atau rumusan masalah. Selain itu, bab ini juga memberikan saran untuk penelitian selanjutnya yang dapat dilakukan agar meningkatkan kualitas dari penelitian tersebut.