Noted: Jika hanya 1 tidak perlu menggunakan Daftar Lampiran

BAB I PENDAHULUAN

1.1 Latar Belakang

Komunikasi merupakan proses fundamental dalam interaksi manusia, berfungsi sebagai medium pertukaran informasi, ide, dan emosi yang memungkinkan terbentuknya pemahaman bersama dan kohesi sosial. Namun, efektivitas proses ini seringkali terhambat oleh adanya kesenjangan bahasa. Salah satu kesenjangan yang paling signifikan dan berdampak luas terhadap inklusivitas sosial di Indonesia adalah antara komunitas Teman Dengar (orang yang dapat mendengar) dan Teman Tuli (orang dengan disabilitas pendengaran).

Besarnya populasi yang menghadapi tantangan ini di Indonesia sendiri tercermin dalam berbagai data statistik. Meskipun terdapat variasi antar survei, sebuah tantangan tersendiri dalam perencanaan kebijakan yang akurat, semua data yang tersedia mengonfirmasi adanya populasi signifikan yang membutuhkan perhatian. Survei Sosial Ekonomi Nasional (Susenas) pada tahun 2019 memperkirakan terdapat sekitar 23,3 juta penyandang disabilitas di Indonesia[1][2]. Data yang lebih baru dari BPS (Badan Pusat Statistik) pada tahun 2022 mencatat angka 22,5 juta jiwa[3][4], sementara Survei Kesehatan Indonesia (SKI) oleh Kementerian Kesehatan pada tahun 2023 melaporkan prevalensi disabilitas pendengaran secara spesifik sebesar 0,4% dari populasi[5]. Terlepas dari perbedaan angka, data ini secara kolektif menggarisbawahi skala masalah dan menegaskan urgensi pengembangan solusi yang dapat diakses dan diskalakan.

Hambatan yang dialami oleh komunitas Tuli di Indonesia tidak hanya terbatas pada interaksi personal, tetapi merambah ke berbagai aspek krusial dalam kehidupan yang menciptakan siklus marginalisasi sistemik[6]. Akses terhadap layanan publik seringkali menjadi tantangan besar. Berbagai laporan menyoroti kesulitan yang dihadapi Teman Tuli saat mengakses layanan perbankan, mendaftar ke institusi pendidikan, atau bahkan melaporkan tindak kejahatan ke pihak kepolisian, akibat minimnya ketersediaan JBI (Juru Bahasa Isyarat) profesional[7]. Di fasilitas kesehatan, ketiadaan informasi yang aksesibel dan JBI dapat menyebabkan kesalahpahaman dalam diagnosis dan pengobatan, yang berisiko terhadap kesehatan mereka[8]. Situasi ini diperburuk di sektor pendidikan, di mana banyak anak Tuli tidak mendapatkan akses pendidikan yang layak. Kurikulum yang tidak diadaptasi, kurangnya guru yang kompeten dalam bahasa isyarat, serta fasilitas yang tidak memadai menjadi penghalang utama, yang pada akhirnya membatasi capaian akademik dan peluang mereka di masa depan[8]. Keterbatasan dalam pendidikan ini secara langsung berdampak pada akses ke dunia kerja, di mana penyandang

disabilitas rungu sering menghadapi diskriminasi dan stigma, serta kurangnya akomodasi yang layak dari perusahaan.

Sebagai solusi komunikasi alami dan bahasa ibu bagi komunitas Tuli, bahasa isyarat memegang peranan sentral. Di Indonesia, terdapat dua sistem yang utama: SIBI (Sistem Isyarat Bahasa Indonesia) dan Bahasa Isyarat Indonesia (BISINDO)[9]. Perbedaan keduanya sangat fundamental dan menjadi inti dari diskursus hak linguistik komunitas Tuli. SIBI adalah sistem yang distandarisasi dan dipromosikan oleh pemerintah untuk digunakan di lingkungan pendidikan formal. Strukturnya direkayasa untuk mengikuti tata bahasa lisan Indonesia, menjadikannya sebuah sistem artifisial yang dibuat oleh orang dengar untuk orang Tuli. Sebaliknya, BISINDO adalah bahasa yang tumbuh dan berkembang secara organik di dalam komunitas Tuli, dengan struktur visual-spasial dan tata bahasa yang unik serta kaya akan ekspresi wajah dan gerak tubuh[9]. Karena sifatnya yang alamiah, BISINDO lebih banyak digunakan dalam percakapan sehari-hari dan dianggap sebagai representasi otentik dari budaya Tuli. Banyak aktivis dan organisasi Tuli, seperti GERKATIN (Gerakan untuk Kesejahteraan Tunarungu Indonesia), memperjuangkan pengakuan BISINDO karena SIBI dianggap sebagai bentuk pemaksaan linguistik yang tidak melibatkan partisipasi komunitas Tuli dalam pengembangannya[10]. Oleh karena itu, pemilihan BISINDO sebagai fokus dalam pengembangan teknologi pendukung komunikasi ini bukan sekadar pilihan teknis, melainkan sebuah keputusan yang berlandaskan pada pemahaman budaya dan keberpihakan pada bahasa yang paling otentik bagi penggunanya, sebuah langkah penting menuju inklusivitas sejati.

Dalam beberapa tahun terakhir, kemajuan pesat di bidang teknologi, khususnya *Computer Vision* dan *Deep Learning*, telah membuka peluang baru untuk menjembatani kesenjangan komunikasi ini. Teknologi pengenalan bahasa isyarat atau SLR (*Sign Language Recognition*) adalah teknologi yang dirancang untuk mengenali dan menerjemahkan bahasa isyarat ke dalam bentuk teks atau suara secara otomatis. Sistem ini bertujuan untuk menjembatani komunikasi antara pengguna bahasa isyarat, seperti Teman Tuli, dengan orang yang tidak memahami bahasa isyarat. Proses pengenalannya dimulai dari pengambilan data gerakan tangan, ekspresi wajah, dan postur tubuh menggunakan kamera atau sensor. Data ini kemudian diolah menggunakan metode ekstraksi fitur, salah satunya melalui teknologi seperti MediaPipe Holistic yang mampu mendeteksi titik-titik kunci tubuh secara real-time. Karena bahasa isyarat bersifat dinamis dan sekuensial, digunakan model deep learning seperti RNN (*Recurrent Neural Network*) atau LSTM (*Long Short-Term Memory*) untuk mempelajari pola gerakan dari waktu ke waktu. Hasil pengenalan kemudian diklasifikasikan menjadi teks atau suara agar dapat dipahami oleh lawan bicara. SLR ini memungkinkan mesin untuk "melihat" dan menginterpretasikan gerakan isyarat visual, kemudian menerjemahkannya ke dalam bentuk teks atau suara[11]. Perkembangan ini menawarkan potensi luar biasa untuk

menciptakan alat bantu komunikasi yang dapat diakses, terjangkau, dan efektif, sehingga dapat digunakan dalam berbagai situasi sehari-hari tanpa memerlukan kehadiran JBI manusia secara fisik.

Tugas Akhir ini mengusulkan pengembangan sistem deteksi BISINDO secara *real-time* dengan memanfaatkan kombinasi teknologi mutakhir, yaitu *MediaPipe* dan LSTM (*Long Short-Term Memory*). Pendekatan ini dipilih setelah analisis mendalam terhadap berbagai metode yang ada. Berbeda dengan pendekatan tradisional yang menggunakan CNN(*Convolutional Neural Network*) untuk memproses seluruh *frame* gambar, yang secara komputasi sangat berat, sistem ini menggunakan *MediaPipe* untuk mengekstraksi fitur *landmark* atau *keypoint* dari tubuh, wajah, dan tangan. Data *keypoint* yang berupa urutan koordinat ini kemudian dianalisis oleh model LSTM, sebuah arsitektur jaringan saraf yang dirancang khusus untuk memahami data sekuensial atau deret waktu, seperti gerakan dinamis dalam bahasa isyarat. Kombinasi ini terbukti akurat juga menjadikannya pilihan ideal untuk aplikasi *real-time* yang responsif.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, masalah-masalah yang akan dipecahkan dalam Tugas Akhir ini dirumuskan sebagai berikut:

1. Bagaimana merancang dan mengimplementasikan sebuah *pipeline* pengolahan data yang efektif untuk mengekstraksi fitur-fitur *keypoint* spasial-temporal dari video gerakan BISINDO secara real-time menggunakan *MediaPipe*?
2. Bagaimana membangun dan melatih model klasifikasi berbasis LSTM (*Long Short-Term Memory*) yang mampu mengenali dan membedakan beberapa kosakata dinamis dalam BISINDO berdasarkan *sekuens* fitur *keypoint* yang telah diekstraksi?
3. Bagaimana mengevaluasi kinerja sistem yang dikembangkan secara kuantitatif, baik dari segi akurasi klasifikasi (menggunakan metrik seperti *accuracy, precision, recall,* dan *confusion matrix*) maupun dari segi performa *real-time* (latensi dan *frames per second*)?

1.3 Tujuan

Berdasarkan latar belakang yang telah diuraikan, tujuan dari Tugas Akhir ini adalah sebagai berikut:

1. Mengimplementasikan *pipeline* pengolahan data dari akuisisi video mentah hingga menjadi vektor fitur sekuensial yang siap digunakan oleh model *deep learning*, dengan memanfaatkan pustaka OpenCV dan *MediaPipe*.

2. Membangun, melatih, dan melakukan *tuning* pada model *deep learning* dengan arsitektur LSTM untuk melakukan klasifikasi terhadap kosakata BISINDO yang telah ditentukan.
3. Melakukan pengujian dan analisis komprehensif terhadap sistem untuk mengukur tingkat akurasi klasifikasi dan kelayakan performa untuk penggunaan secara *real-time*.

1.4 Cakupan Pengerjaan

Untuk memastikan pengerjaan Tugas akhir ini tetap terarah, kami tetapkan batasan-batasan masalah sebagai berikut:

1. Fokus Sistem: Tugas ini secara eksklusif berfokus pada pengembangan dan pengujian mesin deteksi (*detection engine*) bahasa isyarat secara *real-time*. Tugas ini tidak mencakup pengembangan antarmuka pengguna grafis (*Graphical User Interface*), aplikasi *front-end* (baik berbasis web maupun *mobile*), maupun sistem *back-end* yang terintegrasi dengan API (Application Programming Interface). Sistem yang dibangun bersifat *standalone*.
2. Metodologi: Sistem ini menggunakan kombinasi teknologi *MediaPipe* untuk proses ekstraksi fitur dan LSTM (*Long Short-Term Memory*) untuk klasifikasi sekuens gerakan. Metode berbasis CNN (*Convolutional Neural Network*) untuk ekstraksi fitur langsung dari gambar mentah, sebagaimana diusulkan dalam proposal, tidak digunakan dalam implementasi akhir ini.
3. Bahasa Isyarat: Sistem dirancang dan dilatih secara spesifik untuk mendeteksi Bahasa Isyarat Indonesia (BISINDO). Sistem ini tidak dirancang untuk mengenali SIBI (Sistem Isyarat Bahasa Indonesia) atau bahasa isyarat dari negara lain.
4. Kosakata: Deteksi terbatas pada sejumlah kata atau kalimat pendek yang bersifat dinamis, yang umum digunakan dalam percakapan BISINDO sehari-hari dan juga mencakup deteksi abjad namun tidak di satukan oleh sistem *Holistic*.
5. Arah Komunikasi: Sistem ini beroperasi secara satu arah, yaitu menerjemahkan gerakan isyarat visual yang ditangkap oleh kamera menjadi keluaran berupa teks. Tidak ada fungsionalitas untuk komunikasi dua arah, seperti konversi dari teks ke visualisasi isyarat.
6. Kondisi Pengujian: Evaluasi performa dan akurasi sistem dilakukan dalam kondisi lingkungan yang relatif terkontrol. Hal ini mencakup pencahayaan yang memadai, latar belakang yang tidak terlalu ramai atau kompleks, dan posisi pengguna yang menghadap langsung ke kamera.

1.5 Tahapan Pengerjaan

Metodologi pengerjaan tugas ini mengadopsi model pengembangan perangkat lunak *waterfall* yang dimodifikasi untuk mengakomodasi sifat iteratif dari penelitian *machine learning*.

1. Studi Literatur: Tahap awal ini melibatkan analisis mendalam terhadap penelitian-penelitian terdahulu di bidang pengenalan bahasa isyarat. Fokus utama adalah pada sistem yang menggunakan pendekatan *deep learning*. Kajian pustaka secara spesifik mendalami teknologi kunci yang akan digunakan, yaitu *MediaPipe* untuk ekstraksi fitur kerangka dan arsitektur LSTM (*Long Short-Term Memory*) untuk pemodelan data sekuensial.
2. Pengumpulan dan Persiapan Data: Tahap ini berfokus pada pembuatan dataset primer. Proses ini mencakup perekaman serangkaian video untuk setiap kosakata BISINDO yang telah ditentukan sebagai target deteksi. Video-video ini kemudian diorganisir dan diproses untuk menjadi data mentah yang akan digunakan dalam tahap pelatihan dan pengujian model.
3. Perancangan Sistem dan Model: Pada tahap ini, arsitektur sistem secara keseluruhan dirancang, mulai dari alur data (dari input kamera hingga *output* prediksi) hingga perancangan arsitektur spesifik dari model LSTM. Perancangan model mencakup penentuan jumlah lapisan (layer), jumlah neuron di setiap lapisan, fungsi aktivasi yang digunakan, dan parameter lainnya.
4. Implementasi dan Pelatihan: Tahap ini merupakan realisasi dari perancangan ke dalam bentuk kode program. Implementasi dilakukan menggunakan bahasa pemrograman Python dan memanfaatkan pustaka utama seperti OpenCV, MediaPipe, TensorFlow/Keras, dan Scikit-learn. Setelah sistem diimplementasikan, proses pelatihan model dilakukan pada data yang telah disiapkan. Tahap ini juga mencakup proses optimasi melalui *hyperparameter tuning* untuk mendapatkan performa model terbaik.
5. Pengujian dan Evaluasi: Setelah model dilatih, sistem diuji secara menyeluruh menggunakan data uji yang sebelumnya tidak pernah digunakan dalam proses pelatihan. Kinerja sistem dievaluasi secara kuantitatif menggunakan berbagai metrik standar seperti akurasi, presisi. Hasil pengujian dianalisis secara mendalam untuk memahami kekuatan dan kelemahan sistem.