

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan pesat dalam bidang kecerdasan buatan, khususnya *Deep Learning*, telah mendorong lahirnya berbagai sistem cerdas yang mampu melakukan pemrosesan informasi secara kompleks dan adaptif. Dua bidang yang mengalami perkembangan signifikan adalah *Computer Vision* yang berfokus pada kemampuan sistem komputer untuk memahami informasi visual, dan *Natural Language Processing* yang berfokus untuk memahami, memproses, serta menghasilkan keluaran dalam bahasa alami [1]. Integrasi antara kedua bidang tersebut membuka peluang baru dalam pengembangan teknologi yang mampu memahami, menghubungkan, serta mentransformasikan informasi dari domain visual ke dalam bentuk linguistik. Salah satu aplikasi penting dari integrasi tersebut adalah sistem penerjemahan bahasa isyarat atau *Sign Language Translation* [2].

Sign Language Translation (SLT) merupakan penelitian yang bertujuan untuk membangun sistem otomatis yang mampu menerjemahkan bahasa isyarat ke dalam bentuk bahasa alami, seperti teks atau suara, sehingga informasi yang disampaikan oleh penutur dapat dipahami oleh pengguna yang tidak menguasai bahasa tersebut. Secara umum, sistem SLT terdiri atas tiga tahapan yang saling terintegrasi. Tahap pertama adalah pemetaan informasi visual dari video serta teks target ke dalam bentuk representasi fitur. Selanjutnya adalah tahap pemrosesan representasi tersebut menggunakan model *Machine Learning* untuk memahami pola bahasa. Terakhir adalah proses penerjemahan yang menghasilkan teks dalam bahasa yang sesuai dengan semantik [3].

Sebagai fondasi dari sistem SLT, *Sign Language Recognition* (SLR) memegang peran penting dalam proses pemahaman bahasa isyarat. SLR berfokus pada pengenalan serta klasifikasi gerakan isyarat dari video ke dalam bentuk label atau *gloss* tertentu. Dalam banyak penelitian, model SLR biasanya dilatih menggunakan video bahasa isyarat yang telah diberi anotasi label yang sesuai karena setiap gerakan isyarat diasumsikan merepresentasikan satu makna atau kata tertentu. Meskipun demikian, kemampuan pengenalan ini memiliki cakupan aplikasi yang luas, antara lain untuk mendeteksi kemunculan isyarat dalam sebuah

video, pencarian video bahasa isyarat berbasis kata kunci, serta pengenalan bahasa isyarat kontinu atau *Continuous Sign Language Recognition* [4].

Untuk memungkinkan proses tersebut, digunakan suatu alur pemrosesan multimodal yang disebut *pipeline Vision-and-Language* untuk menggabungkan informasi visual dengan representasi linguistik dalam satu sistem terpadu [5]. *Pipeline* ini memungkinkan data visual seperti video bahasa isyarat diproses secara bertahap hingga menghasilkan keluaran berupa teks bahasa alami yang dapat dipahami oleh non-pengguna bahasa isyarat. Pendekatan *Vision-and-Language* menjadi fondasi penting dalam pengembangan sistem SLT karena berfungsi sebagai jembatan antara representasi visual yang bersifat kontinu dengan struktur linguistik yang bersifat diskrit dan terstruktur [6].

Pipeline Vision-and-Language biasanya melibatkan beberapa komponen. Tahap awal fokus kepada ekstraksi fitur spasio-temporal menggunakan *Convolutional Neural Network* (CNN), *Recurrent Neural Network* (RNN) atau arsitektur berbasis Transformer untuk menangkap informasi penting dalam video. Berikutnya adalah pemrosesan urutan menggunakan model seperti *Long Short-Term Memory* dan *Gated Recurrent Unit* (GRU) untuk memahami hubungan antar gerakan. Terakhir adalah proses *decoding* semantik menggunakan model bahasa yang telah dilatih secara luas untuk menghasilkan kalimat berdasarkan representasi fitur sebelumnya [7]. Namun, pendekatan ini masih memerlukan penyesuaian yang kompleks antara arsitektur model, *loss function*, serta parameter lainnya agar mampu menghasilkan representasi semantik yang stabil dan akurat [8].

Bahasa isyarat sendiri memiliki peran sosial yang sangat penting sebagai sarana komunikasi utama bagi komunitas Tunarungu, dan memiliki struktur linguistik yang kompleks serta bervariasi antar negara dan budaya [9]. World Health Organization (WHO) melaporkan lebih dari 1,5 miliar orang di seluruh dunia, atau hampir 20% dari total populasi global, hidup dengan gangguan pendengaran. Sekitar 430 juta orang diantaranya mengalami gangguan pendengaran berat yang memerlukan rehabilitasi, dan jumlah tersebut diproyeksikan meningkat hingga lebih dari 700 juta orang pada tahun 2050 [10]. Kondisi di Indonesia juga menunjukkan urgensi yang serupa. Berdasarkan Survei Kesehatan Indonesia (SKI)

2023 oleh Kementerian Kesehatan, prevalensi disabilitas pendengaran pada populasi usia ≥ 1 tahun tercatat sebesar 0,4% [11].

Indonesia memiliki dua jenis bahasa isyarat utama, yaitu Sistem Isyarat Bahasa Indonesia (SIBI) yang distandardisasi oleh pemerintah dan banyak digunakan dalam konteks pendidikan formal, serta Bahasa Isyarat Indonesia (BISINDO) yang berkembang secara alami dalam komunitas Tunarungu dan memiliki struktur gramatikal yang berbeda dari bahasa Indonesia lisan [12]. Meskipun BISINDO lebih luas digunakan dalam komunikasi sehari-hari oleh komunitas Tunarungu, penelitian serta pengembangan teknologi penerjemahannya masih tertinggal dibandingkan bahasa isyarat dari negara lain seperti *American Sign Language* (ASL) dan *German Sign Language* (DGS) [13].

Salah satu faktor penghambat perkembangan teknologi penerjemahan BISINDO adalah keterbatasan dokumentasi dan ketersediaan dataset digital. Jika dibandingkan dengan dataset *benchmark* seperti WLASL dan RWTH-PHOENIX-Weather, dataset BISINDO masih sangat terbatas baik dari sisi jumlah data maupun variasi linguistik [14]. Selain itu, kompleksitas visual dari bahasa isyarat yang melibatkan koordinasi gerakan tangan, orientasi tubuh, serta ekspresi lainnya menjadi tantangan tersendiri dalam ranah *Computer Vision*. Banyak penelitian terdahulu masih berfokus pada rekognisi semata, sehingga belum mampu menghasilkan terjemahan dalam bentuk kalimat bahasa alami yang koheren secara semantik dan gramatikal [15].

Dengan mempertimbangkan meningkatnya kebutuhan aksesibilitas komunikasi serta keterbatasan penelitian dan dataset yang tersedia, dibutuhkan suatu pendekatan *Vision-and-Language* yang dirancang untuk konteks Bahasa Isyarat Indonesia. Pengembangan *pipeline* yang mampu mengekstraksi fitur gerakan secara efektif, memodelkan urutan isyarat, serta menghasilkan keluaran dalam bentuk bahasa alami menjadi langkah penting untuk menjembatani kesenjangan teknologi tersebut. Oleh karena itu, penelitian ini berupaya merancang sistem penerjemahan bahasa isyarat BISINDO yang dapat menjadi fondasi bagi pengembangan yang lebih luas di masa mendatang.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang *pipeline Vision-and-Language* yang terintegrasi dengan mekanisme *Attention* untuk menerjemahkan video Bahasa Isyarat Indonesia (BISINDO) menjadi teks bahasa alami secara efektif?
2. Bagaimana mengukur performa sistem penerjemahan yang dibangun berdasarkan metrik evaluasi terjemahan untuk menilai akurasi dan kesesuaian semantik hasil terjemahan?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan sebuah *pipeline Vision-and-Language* yang mampu menerjemahkan video Bahasa Isyarat Indonesia (BISINDO) menjadi teks bahasa alami secara efektif.
2. Menilai kualitas keluaran sistem secara objektif menggunakan metrik evaluasi terjemahan guna mengetahui tingkat akurasi dan kesesuaian semantik hasil terjemahan.

1.4 Batasan Masalah

Penelitian ini memiliki sejumlah batasan agar ruang lingkup penelitian tetap terarah sesuai dengan tujuan, yaitu:

1. Dataset yang digunakan dalam penelitian ini terbatas pada 30 kalimat yang mewakili percakapan sehari-hari, dengan 30 kali pengulangan untuk setiap kalimat. Partisipan dipilih dari Sekolah Luar Biasa Negeri (SLBN) untuk mendapatkan penutur yang kompeten.
2. Evaluasi performa model dibatasi pada metrik *Bilingual Evaluation Understudy* (BLEU) dan *Word Error Rate* (WER) untuk mengukur kualitas semantik dan akurasi teks hasil terjemahan.
3. Penelitian ini mengadaptasi kerangka kerja CRISP-DM sampai dengan tahap *evaluation*. Tahap *deployment* tidak dilakukan karena penelitian berfokus pada perancangan *pipeline* dan evaluasi performa model dalam lingkungan eksperimen.

1.5 Manfaat

Penelitian tugas akhir ini diharapkan dapat memberikan beberapa manfaat terhadap literatur dan operasional. Manfaat tersebut yaitu:

1.5.1 Bagi Literatur

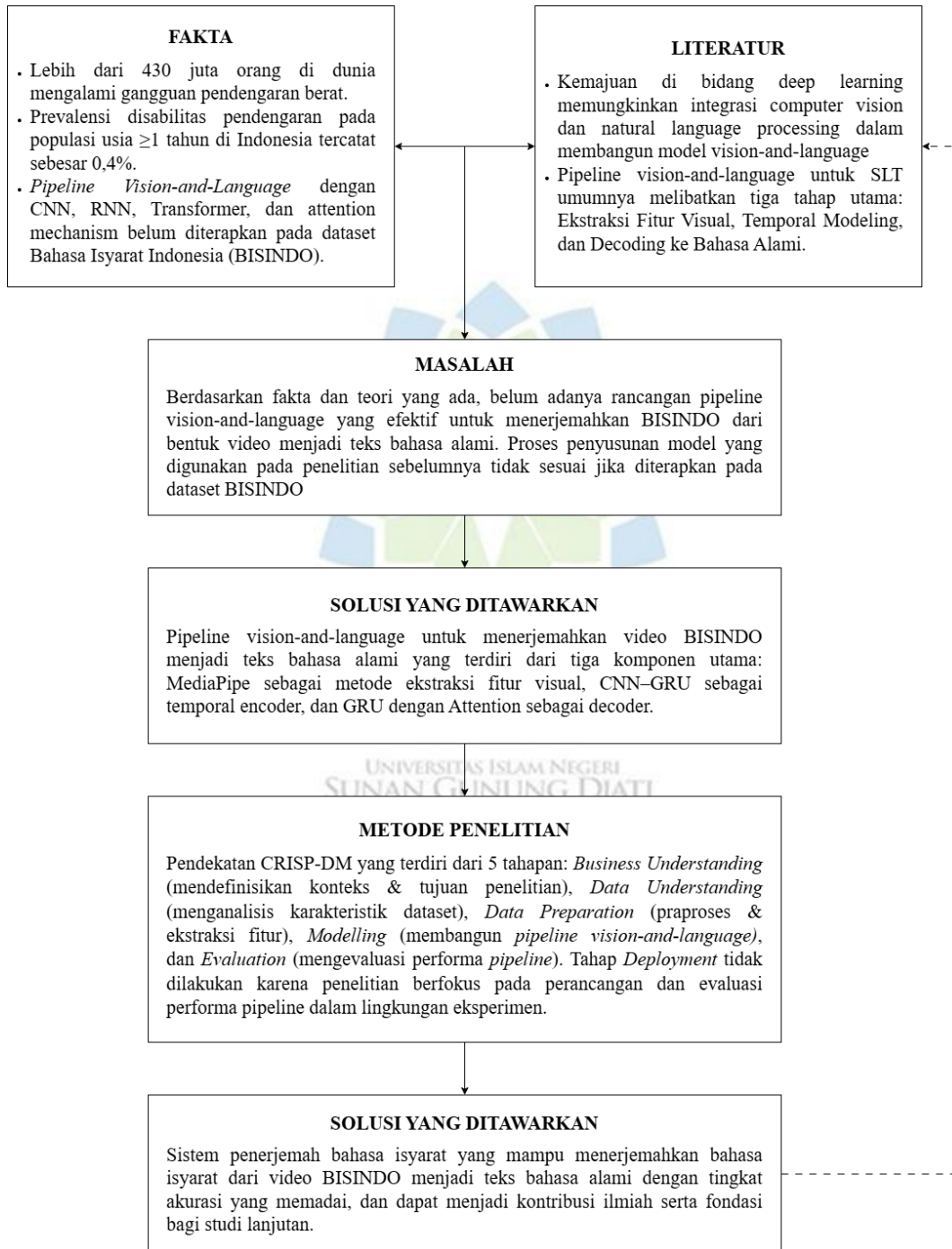
1. Meningkatkan pemahaman dalam penerapan *Machine Learning*, khususnya dalam penggabungan bidang *Computer Vision* dan *Natural Language Processing*.
2. Memberikan kontribusi ilmiah untuk pengembangan sistem penerjemahan Bahasa Isyarat Indonesia (BISINDO) berbasis *Deep Learning*.
3. Menjadi landasan bagi penelitian lanjutan seperti penerjemahan dua arah dan pengembangan model yang lebih adaptif terhadap variasi dan kompleksitas bahasa isyarat lokal.

1.5.2 Bagi Operasional

1. Menyediakan rancangan *pipeline Vision-and-Language* yang dapat dijadikan acuan untuk pengembangan sistem penerjemah bahasa isyarat atau sistem terkait lainnya.
2. Menjadi tahap awal dalam membantu masyarakat non-penutur bahasa isyarat untuk memahami dan berkomunikasi dengan penyandang tunarungu di Indonesia.

1.6 Kerangka Pemikiran

Kerangka pemikiran ini menggambarkan alur penelitian berdasarkan fakta, literatur, masalah, solusi, metodologi yang digunakan, hingga hasil yang diharapkan dalam penelitian ini.



Gambar 1.1 Kerangka Pikiran Penelitian

Seperti yang telah ditampilkan di gambar 1.1, penelitian ini berangkat dari fakta bahwa lebih dari 430 juta orang di dunia mengalami gangguan pendengaran berat [10]. Di Indonesia sendiri, prevalensi disabilitas pendengaran pada populasi usia satu tahun ke atas tercatat sebesar 0,4% [11]. Di sisi lain, penerapan *pipeline Vision-and-Language* yang memanfaatkan metode *Deep Learning* seperti CNN, RNN, *Transformer*, maupun mekanisme *attention* masih sangat terbatas untuk bahasa isyarat lokal seperti Bahasa Isyarat Indonesia (BISINDO) [14].

Dalam konteks penerjemahan bahasa isyarat, penelitian-penelitian sebelumnya umumnya membuat sistem yang terdiri dari tiga tahapan utama, yaitu ekstraksi fitur visual dari video bahasa isyarat, pemodelan temporal untuk memahami urutan gerakan, serta proses *decoding* untuk mengubah menjadi kalimat dalam bahasa alami [7]. Namun berdasarkan fakta dan kajian literatur yang ada, *pipeline* yang digunakan pada penelitian sebelumnya umumnya dikembangkan untuk dataset bahasa isyarat lain yang memiliki karakteristik berbeda, sehingga penerapannya pada BISINDO tidak dapat memberikan hasil yang optimal [15].

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan sebuah *pipeline Vision-and-Language* yang dirancang untuk menerjemahkan video BISINDO menjadi teks bahasa alami. *Pipeline* ini terdiri dari tiga komponen utama. *MediaPipe* sebagai alat ekstraksi fitur visual, CNN-GRU sebagai *encoder*, dan GRU-*Attention* sebagai *decoder*. Penelitian ini mengadaptasi pendekatan *Cross-Industry Standard Process for Data Mining (CRISP-DM)* yang dibatasi pada lima tahapan, yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, dan *evaluation*.

Penelitian ini diharapkan dapat membantu meningkatkan aksesibilitas komunikasi bagi penyandang disabilitas pendengaran. Selain itu, *pipeline* yang dihasilkan juga dapat menjadi dasar bagi penelitian lanjutan dalam pengembangan sistem penerjemah bahasa isyarat yang lebih kompleks dan aplikatif.

1.7 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, rumusan masalah, tujuan penelitian, batasan masalah, kerangka pemikiran, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas literatur atau penelitian terdahulu, konsep dan teori, model yang digunakan, dan rumus terkait yang menjadi landasan dalam proses analisis permasalahan dan kebutuhan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini 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, yaitu temuan atau hasil penelitian berdasarkan langkah-langkah penelitian yang telah dilakukan, dan pembahasan hasil atau temuan penelitian sebagai jawaban terhadap rumusan masalah penelitian.

BAB V SIMPULAN DAN SARAN

Bab ini berfokus pada penarikan simpulan dari hasil penelitian yang diperoleh serta menjawab pertanyaan penelitian, serta memberikan saran yang dapat dilakukan penelitian selanjutnya untuk meningkatkan kualitas penelitian.