



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**RANCANG BANGUN APLIKASI BELAJAR BAHASA
ISYARAT SEBAGAI MEDIA PEMBELAJARAN
BERBASIS WEBSITE**

**IMPLEMENTASI MODEL PENGENALAN BAHASA
ISYARAT LANJUTAN MENGGUNAKAN
MEDIAPIPE DAN CTC-ATTENTION
SKRIPSI**

MUHAMMAD RIZIQ RAMADHAN AL ARAFAH 2107411035

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN
KOMPUTER
POLITEKNIK NEGERI JAKARTA
TAHUN 2024/2025**



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**RANCANG BANGUN APLIKASI BELAJAR BAHASA
ISYARAT SEBAGAI MEDIA PEMBELAJARAN
BERBASIS WEBSITE**

**IMPLEMENTASI MODEL PENGENALAN BAHASA
ISYARAT LANJUTAN MENGGUNAKAN
MEDIAPIPE DAN CTC-ATTENTION
SKRIPSI**

**Dibuat untuk Melengkapi Syarat-Syarat yang Diperlukan untuk
Memperoleh Diploma Empat Politeknik**

MUHAMMAD RIZIQ RAMADHAN AL ARAFAH 2107411035

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER
POLITEKNIK NEGERI JAKARTA
TAHUN 2024/2025**



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

SURAT PERNYATAAN BEBAS PLAGIARISME

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Riziq Ramadhan Al Arafah
NIM : 2107411035
Jurusan/Program Studi : T.Informatika dan Komputer / Teknik Informatika
Judul skripsi : IMPLEMENTASI MODEL PENGENALAN
BAHASA ISYARAT LANJUTAN MENGGUNAKAN MEDIA PIPE DAN CTC-
ATTENTION

Menyatakan dengan sebenarnya bahwa skripsi ini benar-benar merupakan hasil karya saya sendiri, bebas dari peniruan terhadap karya dari orang lain. Kutipan pendapat dan tulisan orang lain ditunjuk sesuai dengan cara-cara penulisan karya ilmiah yang berlaku.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa dalam skripsi ini terkandung ciri-ciri plagiat dan bentuk-bentuk peniruan lain yang dianggap melanggar peraturan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Depok, 17 Juni 2025

Yang membuat pernyataan

Muhammad Riziq Ramadhan Al Arafah

NIM 2107411035



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta:

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

LEMBAR PENGESAHAN

Skripsi diajukan oleh:

Nama : Muhammad Riziq Ramadhan Al Arafah
NIM : 2107411035
Program Studi : Teknik Informatika
Judul Skripsi : IMPLEMENTASI MODEL PENGENALAN BAHASA
ISYARAT MENGGUNAKAN MEDIAPIPE DAN CTC-ATTENTION

Telah diuji oleh tim penguji dalam sidang skripsi pada hari Jumat, tanggal 20, bulan Juni, tahun 2025 dan dinyatakan **LULUS**.

Disahkan oleh:

Pembimbing I : Chandra Wirawan S.Kom., M.Kom.

Penguji I : Mera Kartika Delimayanti., S.Si., M.T., Ph.D

Penguji II : Rizki Elisa Nalawati, S.T., M.T.

Penguji III : Bambang Warsuta S.Kom., M.T.I.

Mengetahui:

Jurusan Teknik Informatika dan Komputer
Ketua Jurusan



Dr. Anita Hidayati, S.Kom., M.Kom.

NIP 197908032003122003



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “IMPLEMENTASI MODEL PENGENALAN BAHASA ISYARAT LANJUTAN MENGGUNAKAN MEDIAPIPE DAN CTC-ATTENTION”. Skripsi ini disusun sebagai salah satu syarat kelulusan pada program Teknik Informatika di Politeknik Negeri Jakarta.

Dalam proses penyusunan ini, penulis menerima banyak dukungan, arahan, dan bantuan dari berbagai pihak. Untuk itu, penulis mengucapkan terima kasih sebesar-besarnya kepada:

1. Chandra Wirawan, S.Kom., M.Kom., selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta masukan berharga dalam penyusunan skripsi ini.
2. Seluruh Tim Pelaksanaan Skripsi Jurusan Teknik Informatika dan Komputer Politeknik Negeri Jakarta, atas dukungan dan fasilitas yang diberikan selama proses penyusunan.
3. Kedua orang tua saya, Nawawie Arafat dan Riny Holilah, yang senantiasa memberikan doa serta dukungan tanpa henti dalam setiap tahap perjalanan ini.
4. Shieria Intan Anggari, yang telah memberikan pengalaman studi yang berkesan selama berkuliah di Politeknik Negeri Jakarta, serta membuat semuanya menjadi mungkin.
5. Muhammad Fahmi Zuhdi dan Billa Fathia, rekan dalam pengembangan skripsi ini, atas dukungan, semangat, dan bantuan yang telah diberikan hingga terselesaikannya penelitian ini.

Semoga skripsi ini bermanfaat dan dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan di bidang teknologi. Kritik dan saran sangat penulis harapkan untuk perbaikan ke depan.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

**SURAT PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI
UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Politeknik Negeri Jakarta, saya bertanda tangan dibawah ini :

Nama : Muhammad Riziq Ramadhan Al Arafah

NIM : 2107411035

Jurusan/ProgramStudi : T.Informatika dan Komputer / Teknik Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Politeknik Negeri Jakarta Hak Bebas Royalti Non-Eksklusif atas karya ilmiah saya yang berjudul :

**IMPLEMENTASI MODEL PENGENALAN BAHASA ISYARAT
LANJUTAN MENGGUNAKAN MEDIAPIPE DAN CTC-ATTENTION**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksklusif ini Politeknik Negeri Jakarta Berhak menyimpan, mengalihmediakan/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya tanpa meminta izin darisaya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Depok, 17 Juni 2025

Yang Menyatakan

**POLITEKNIK
NEGERI
JAKARTA**

Muhammad Riziq Ramadhan Al Arafah
NIM. 21074110



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Implementasi Model Pengenalan Bahasa Isyarat Lanjutan Menggunakan Mediapipe dan CTC-ATTENTION

Abstrak

Pengenalan bahasa isyarat secara lanjutan atau *Continuous Sign Language Recognition* (CSLR) merupakan tantangan dalam bidang teknologi asistif yang bertujuan menerjemahkan gerakan isyarat secara real-time ke dalam teks natural. Penelitian ini bertujuan mengembangkan model CSLR dengan pendekatan hybrid Connectionist Temporal Classification (CTC) dan Attention Mechanism yang memanfaatkan ekstraksi fitur *keypoints* dari video menggunakan MediaPipe Holistic. Dataset yang digunakan terdiri dari kalimat dalam bahasa isyarat Sistem Isyarat Bahasa Indonesia (SIBI) yang mencakup kategori dasar anggota keluarga, benda sekitar, dan binatang. Model dikembangkan menggunakan dua arsitektur utama pada tahap encoder, yaitu Pyramid BiLSTM dan Pyramid BiGRU, yang kemudian dibandingkan performanya berdasarkan metrik Word Error Rate (WER), *CTC Loss*, serta *Attention Loss*. Model dilatih menggunakan data dengan dan tanpa augmentasi untuk mengevaluasi dampak teknik augmentasi terhadap performa model. Hasil evaluasi menunjukkan bahwa model berbasis BiGRU dengan augmentasi dataset mencapai performa terbaik dengan WER sebesar 26,20%, yang secara signifikan lebih baik dibandingkan BiLSTM yang mencatatkan WER sebesar 38,13%. Model kemudian diimplementasikan dalam sistem berbasis web yang mendukung proses inferensi baik secara batch maupun real-time melalui integrasi REST API dan WebSocket dengan *back-end* Flask pada layanan Amazon Web Services (AWS).

Kata kunci: arsitektur bilstm bigru, mediapipe holistic, pengenalan bahasa isyarat lanjutan, sibi, sistem isyarat bahasa indonesia.

POLITEKNIK
NEGERI
JAKARTA



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta

2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

DAFTAR ISI

KATA PENGANTAR.....	i
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	ii
ABSTRAK	iii
DAFTAR ISI.....	iv
Daftar Gambar	vi
Daftar Tabel.....	vii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat Penelitian	3
1.5 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Tunarungu	6
2.2 Bahasa Isyarat	6
2.3 <i>Machine Learning</i>	7
2.4 <i>Deep Learning</i>	8
2.5 Python	8
2.6 Tensorflow	9
2.7 Mediapipe.....	9
2.8 <i>Connectionist Temporal Classification (CTC)</i>	10
2.9 <i>Attention Mechanism</i>	11
2.10 <i>Beam Search</i>	11
2.11 Kalman Filter	12
2.12 Penelitian Terdahulu	13
BAB III METODE PENELITIAN	15
3.1 Rancangan Penelitian	15
3.2 Tahapan Penelitian	15
3.3 Objek Penelitian	18
BAB IV HASIL DAN PEMBAHASAN.....	19
4.1 Analisis Kebutuhan	19



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta

2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

4.1.1	Kebutuhan Pengembangan Model	19
4.1.1.1	Kebutuhan Dataset	19
4.1.1.2	Kebutuhan Prapemrosesan Data	21
4.1.1.3	Kebutuhan Perangkat Lunak	22
4.1.1.4	Kebutuhan Perangkat Keras	23
4.1.2	Kebutuhan <i>Model Deployment</i>	24
4.1.2.1	Kebutuhan Perangkat Lunak	24
4.1.2.2	Kebutuhan Perangkat Keras	25
4.1.3	Kebutuhan Sistem	25
4.1.3.1	Kebutuhan Fungsional	25
4.1.3.2	Kebutuhan Non-Fungsional	27
4.2	Desain	29
4.3	Pengembangan Model	35
4.3.1	Ekstraksi Data	36
4.3.3	Pelatihan Model dengan BiLSTM	43
4.3.4	Pelatihan Model dengan BiGRU	47
4.4	Evaluasi Model	48
4.4.1	Evaluasi Model BiLSTM	51
4.4.2	Evaluasi Model BiGRU	56
4.4.3	Evaluasi Pakar	60
4.5	<i>Model Deployment</i>	61
BAB V PENUTUP		66
5.1	Kesimpulan	66
5.2	Saran	66
DAFTAR PUSTAKA		68
LAMPIRAN		71



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Daftar Gambar

Gambar 2. 1 Deteksi Obyek Menggunakan MediaPipe.....	10
Gambar 2. 2 Mekanisme <i>Beam Search</i>	12
Gambar 3. 3 Tahapan Penelitian	16
Gambar 4. 1 Diagram Arsitektur Model CSLR	30
Gambar 4. 2 Diagram Proses dari sisi klien	34
Gambar 4. 3 Visualisasi Plot Koordinat Pose	39
Gambar 4. 4 Program Tokenisasi Label	40
Gambar 4. 5 Program <i>Dataset Preparation</i>	41
Gambar 4. 6 Program <i>Encoder</i> BiLSTM	43
Gambar 4. 7 Arsitektur Model <i>Encoder</i>	44
Gambar 4. 8 Rangkuman Model <i>Encoder</i> Pyramid BiLSTM.....	44
Gambar 4. 9 Program Decoder <i>Attention</i>	45
Gambar 4. 10 Arsitektur Model Decoder <i>Attention</i>	45
Gambar 4. 11 Ringkasan Arsitektur Model Decoder <i>Attention</i>	45
Gambar 4. 12 Program Implementasi <i>Word Error Rate</i>	50
Gambar 4. 13 Output Hasil Training Model BiLSTM Dengan Augmentasi.....	52
Gambar 4. 14 Output Hasil Training Model BiLSTM Tanpa Augmentasi	53
Gambar 4. 15 Perbandingan Metrik BiLSTM per Epoch	53
Gambar 4. 16 Confussion Matrix LSTM Non Augmentasi.....	54
Gambar 4. 17 Confussion Matrix LSTM dengan Augmentasi	55
Gambar 4. 18 Output Hasil Training BiGRU dengan Augmentasi	57
Gambar 4. 19 Output Hasil Training BiGRU Tanpa Augmentasi	57
Gambar 4. 20 Perbandingan Grafik Metriks BiGRU per <i>Epoch</i>	57
Gambar 4. 21 Confussion Matrix GRU Tanpa Augmentasi.....	59
Gambar 4. 22 Confussion Matrix GRU dengan Augmentasi	60
Gambar 4. 23 Program Endpoint /upload	63
Gambar 4. 24 Program Endpoint /upload	64



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Daftar Tabel

Tabel 4. 1 Distribusi Kata yang muncul pada dataset.....	20
Tabel 4. 2 Kebutuhan Perangkat Lunak Pengembangan Model.....	22
Tabel 4. 3 Tabel Kebutuhan Perangkat Keras Pengembangan Model.....	23
Tabel 4. 4 Kebutuhan Perangkat Lunak <i>Model Deployment</i>	24
Tabel 4. 5 Kebutuhan Perangkat Keras <i>Model Deployment</i>	25
Tabel 4. 6 Kebutuhan Fungsional	25
Tabel 4. 7 Kebutuhan Non-Fungsional	28
Tabel 4. 8 Output Ekstraksi Mediapipe Holistic.....	36
Tabel 4. 9 Parameter Mediapipe Guna Kebutuhan Ekstraksi	37
Tabel 4. 10 Hasil Validasi Dataset Ekstraksi <i>Keypoints</i>	38
Tabel 4. 11 Contoh Representasi Token	40
Tabel 4. 12 Hyperparameter Pelatihan Model BiLSTM.....	46
Tabel 4. 13 Hasil Evaluasi Model BiLSTM.....	52
Tabel 4. 14 Hasil Evaluasi Model BiGRU.....	57

**POLITEKNIK
NEGERI
JAKARTA**



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Bahasa isyarat adalah sarana komunikasi utama bagi penyandang tunarungu, termasuk di Indonesia yang memiliki jumlah penyandang disabilitas cukup signifikan. Menurut data WHO tahun 2019, sekitar 466 juta orang di dunia (6,1% populasi) mengalami gangguan pendengaran, termasuk 34 juta anak-anak. Di Indonesia, data Kementerian Sosial pada tahun 2019 menunjukkan bahwa 7,03% penyandang disabilitas mengalami gangguan pendengaran (Pamungkas, et al., 2023). Disamping itu, fasilitas pendidikan bagi tunarungu di Indonesia saat ini telah mengalami peningkatan, dengan tercatat sebanyak 2.314 Sekolah Luar Biasa (SLB) pada tahun ajaran 2023/2024 yang tersebar di seluruh Indonesia (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2024). Meskipun demikian, kesadaran masyarakat terhadap bahasa isyarat sebagai bentuk komunikasi inklusif masih tergolong rendah, sehingga banyak individu tunarungu masih menghadapi hambatan dalam berkomunikasi. Akibat dari kesenjangan komunikasi ini, penyandang tunarungu kerap mengalami diskriminasi, termasuk dalam dunia kerja, di mana hanya 5% dari 300 ribu pekerja di Jakarta merupakan penyandang disabilitas (Huang, et al., 2023).

Memperkenalkan bahasa isyarat sejak usia dini memiliki dampak signifikan dalam mengurangi diskriminasi terhadap penyandang tunarungu. Anak-anak yang belajar bahasa isyarat sejak kecil akan lebih terbuka terhadap keberagaman dan memahami pentingnya komunikasi yang inklusif (Rahmah & Janah, 2024). Hal ini berkontribusi dalam membentuk generasi yang lebih peduli terhadap kebutuhan orang lain, termasuk komunitas tunarungu, sehingga tercipta interaksi sosial yang lebih setara di masa depan. Dengan pemahaman tersebut, anak-anak tidak hanya mampu berkomunikasi dengan penyandang tunarungu, tetapi juga menjadi agen perubahan dalam menghapus stigma negatif yang kerap dialami kelompok tersebut (Husna, et al., 2024).

Pengenalan Bahasa Isyarat menggunakan deep learning dapat dikelompokkan menjadi Static Sign Language Recognition (SSLR), Isolated Sign Language Recognition (ISLR), dan *Continuous Sign Language Recognition* (CSLR). SSLR menggunakan dataset berupa gambar statis untuk mengenali huruf dan angka dalam bahasa isyarat. ISLR memanfaatkan dataset video pendek, di mana setiap video biasanya berisi satu kata atau simbol dalam bahasa isyarat. Sementara itu, CSLR berfokus pada pengenalan bahasa isyarat tingkat kalimat menggunakan dataset berupa video yang mencakup satu kalimat lengkap. Meskipun SSLR telah mencapai akurasi hingga 98% dalam penelitian sebelumnya (Laines, et al., 2023). Pendekatan ini memiliki keterbatasan dalam aplikasi dunia nyata. Bahasa isyarat pada praktiknya lebih kompleks, petunjuk gerakan tubuh, tangan, dan ekspresi wajah cenderung melengkapi pesan pengucapannya (Saraswati, et al., 2022). Ketiga gestur tersebut sulit diidentifikasi hanya melalui gambar statis. Oleh karena itu, untuk menangkap dinamika gerakan yang membentuk rangkaian isyarat, SLR harus dapat mengenali gestur sebagai rangkaian *frame* gambar yang merepresentasikan urutan gerakan (Pramanto & Suharjito, 2023).

Dalam penelitian ini, dataset bahasa isyarat yang digunakan dibatasi pada kosa kata dasar. Batasan ini merujuk pada materi dalam buku "Belajar Bahasa Isyarat untuk Anak Tunarungu (Dasar)" yang ditulis oleh Nattaya Lakshita, yang menyediakan panduan untuk pembelajaran bahasa isyarat pada tahap awal. Buku ini dipilih karena isinya mencakup isyarat yang sering digunakan dalam kehidupan sehari-hari, sehingga sesuai dengan tujuan penelitian untuk mengembangkan model *Continuous Sign Language Recognition* (CSLR).

Penelitian ini berfokus pada implementasi *Continuous Sign Language Recognition* (CSLR) dengan memanfaatkan teknologi canggih seperti MediaPipe dan pendekatan Connectionist Temporal Classification (CTC) dan *Attention Mechanism*. Melalui penelitian ini, penulis bertujuan untuk memberikan kontribusi nyata dalam memfasilitasi komunikasi antara pengguna bahasa isyarat dan masyarakat umum, khususnya dalam konteks bahasa isyarat Sistem Isyarat Bahasa Indonesia (SIBI). Selain itu, penelitian ini diharapkan dapat menjadi langkah maju

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

dalam pengembangan teknologi pengenalan bahasa isyarat yang lebih adaptif dan relevan untuk mendukung kebutuhan pengguna bahasa Indonesia secara luas.

1.2 Rumusan Masalah

Pemaparan latar belakang di atas merumuskan masalah secara singkat menjadi sebagai berikut “Bagaimana proses implementasi *end-to-end Continuous Sign Language Recognition* (CSLR) menggunakan Teknik ekstraksi *keypoints* Mediapipe dan Algoritma *Hybrid CTC-Attention*?”

1.3 Batasan Masalah

Dari batasan masalah yang telah diuraikan di atas, maka batasan masalah yang ditetapkan adalah:

- 1) Topik dataset kata yang dikumpulkan hanya mengacu pada 3 topik yang terdapat pada buku “Belajar Bahasa Isyarat Untuk Anak Tunarungu”, yaitu Anggota Keluarga, Benda Sekitar, dan Binatang.
- 2) Penelitian ini hanya berfokus pada terjemahan satu arah, yaitu gestur ke teks.
- 3) Pengambilan Dataset untuk pelatihan difokuskan pada konteks bahasa isyarat SIBI.
- 4) Model dikembangkan menggunakan dataset hasil ekstraksi *keypoints* melalui Mediapipe Holistic.
- 5) Model CSLR akan diaplikasikan pada platform website yang digabungkan dengan subtopik anggota kelompok lain.
- 6) Evaluasi yang menjadi indikator performa adalah *Word Error Rate* (WER), *CTC Loss*, dan *Attention Loss*.

1.4 Tujuan dan Manfaat Penelitian

Tujuan dan manfaat dari pengembangan model CSLR adalah sebagai berikut:

a. Tujuan

Tujuan dari penelitian ini adalah untuk mengembangkan proses ekstraksi fitur gestur bahasa isyarat dari video berdasarkan konteks Sistem Isyarat Bahasa Indonesia (SIBI) dengan memanfaatkan MediaPipe Holistic sebagai alat untuk mendeteksi dan merepresentasikan titik-titik penting tubuh, tangan, dan wajah secara terstruktur. Selain itu, penelitian ini juga bertujuan untuk

Jurusan Teknik Informatika dan Komputer – Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta

2. Dilarang mengumuk dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

mengimplementasikan pendekatan Connectionist Temporal Classification (CTC) dan *Attention Mechanism* dalam pelatihan model.

b. Manfaat

Berikut adalah manfaat yang bisa diperoleh dari penelitian ini:

1. Bagi masyarakat Indonesia, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknologi pengenalan bahasa isyarat yang dapat digunakan untuk memfasilitasi komunikasi antara penyandang disabilitas dan masyarakat umum, khususnya dalam konteks Sistem Isyarat Bahasa Indonesia (SIBI).
2. Bagi peneliti, penelitian ini diharapkan dapat menjadi referensi untuk pengembangan model *Continuous Sign Language Recognition* (CSLR) yang lebih akurat dan efisien, serta memberikan wawasan dalam penerapan teknologi MediaPipe, CTC, dan *Attention Mechanism* untuk pengenalan bahasa isyarat.

1.5 Sistematika Penulisan

Untuk memudahkan penelusuran pembahasan yang ada pada penulisan ini, maka adapun penulisan laporan skripsi menggunakan sistematika sebagai berikut:

1. Bagian awal skripsi
Ini memuat halaman awal seperti halaman sampul, halaman judul, pernyataan orisinalitas, pengesahan, kata pengantar, abstrak, daftar isi, daftar tabel, daftar gambar dan daftar lampiran.
2. Bagian utama skripsi

Bagian ini terdiri dari bab-bab antara lain sebagai berikut:

- 1) Bab 1 Pendahuluan, yang berisi latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat serta sistematika penulisan.
- 2) Bab 2 Tinjauan Pustaka, yang meliputi hasil telaah penulis berdasarkan penelitian terdahulu yang memiliki topik berkaitan dengan penelitian yang dilakukan. Dilanjutkan dengan landasan teori yang berisi tentang pembahasan yang dilakukan pada penelitian.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

- 3) Bab 3 Metode Penelitian, untuk menjabarkan tentang metode-metode penelitian yang digunakan dalam pengembangan yang meliputi berbagai aspek, seperti pemilihan lokasi, waktu penelitian, analisis kebutuhan serta alur penelitian.
 - 4) Bab 4 Hasil dan Pembahasan, terdiri dari hasil penelitian dan analisa yang dilakukan baik secara kualitatif, kuantitatif maupun kombinasi serta hasil penelitian.
 - 5) Bab 5 Penutup, bab yang berisi kesimpulan dan saran berdasarkan penelitian. Disini menjabarkan kesimpulan yang dapat dikemukakan, kelemahan yang ada dan bagaimana keselarasan dengan tujuan dan manfaat penelitian. Dilanjutkan dengan saran yang dapat dilakukan untuk mengatasi masalah dan kelemahan yang ada.
3. Bagian Akhir Skripsi
- Bagian akhir ini berisi tentang daftar pustaka serta daftar lampiran pendukung selama proses penelitian.

**POLITEKNIK
NEGERI
JAKARTA**

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

BAB II TINJAUAN PUSTAKA

2.1 Tunarungu

Tunarungu adalah kondisi gangguan pada sistem pendengaran yang menyebabkan ketidakmampuan mendengar, mulai dari tingkat ringan hingga berat. Kondisi ini terbagi menjadi dua kategori, yaitu tuli (ketidakmampuan mendengar sepenuhnya) dan kurang dengar (kehilangan pendengaran sebagian). Penderitanya mengalami kesulitan dalam memahami informasi bahasa melalui indra pendengaran, baik dengan maupun tanpa alat bantu dengar. Orang yang termasuk dalam kategori kurang dengar umumnya masih memiliki sisa pendengaran yang memungkinkan mereka memproses informasi verbal, terutama jika menggunakan alat bantu. Dengan alat tersebut, mereka masih dapat menangkap dan memahami percakapan melalui indra pendengarannya yang tersisa (Salsabila, 2022).

2.2 Bahasa Isyarat

Konversi huruf, angka, frasa, dan kalimat dari bahasa alami ke dalam gerakan tangan dikenal sebagai bahasa isyarat, dan ini membantu orang dengan gangguan suara berinteraksi dengan orang lain. Sebagai pengganti suara, bahasa isyarat berkomunikasi dengan gerakan tangan dan ekspresi wajah (Athira, et al., 2019). Bahasa isyarat memiliki variasi yang berbeda di berbagai negara, dengan American Sign Language (ASL) dari Amerika Serikat sebagai salah satu yang paling umum digunakan. Di Indonesia, terdapat dua sistem utama, yaitu Bahasa Isyarat Indonesia (BISINDO) dan Sistem Isyarat Bahasa Indonesia (SIBI). SIBI, yang saat ini lebih dominan digunakan, dirancang secara sistematis melalui gerakan jari, tangan, dan tubuh untuk merepresentasikan kosakata bahasa Indonesia, serta dipromosikan melalui sekolah luar biasa sebagai media pembelajaran. Sementara itu, BISINDO dikembangkan oleh Gerakan Kesejahteraan Tunarungu Indonesia (GERKATIN) dan lebih banyak dipakai dalam interaksi sehari-hari oleh komunitas tunarungu. Sistem ini menyerap keragaman budaya dan bahasa daerah, sehingga memiliki ciri khas berbeda di tiap wilayah. Dengan demikian, kedua metode ini mencerminkan

adaptasi kebutuhan komunikasi formal dan informal dalam masyarakat (Rahma, 2024).

2.3 *Machine Learning*

Machine learning (ML) adalah cabang ilmu komputer yang mengkaji algoritme dan pendekatan untuk mengotomatisasi solusi bagi masalah rumit yang sulit diprogram dengan metode pemrograman konvensional (Rebala, et al., 2019). Untuk membuat dan mengembangkan algoritme yang dapat belajar dari data, mendapatkan pengetahuan dari pengalaman, dan secara bertahap meningkatkan perilaku pembelajarannya, *Machine Learning* berfokus pada pemahaman kecerdasan. Menemukan pola struktural dan/atau temporal yang relevan (atau “pengetahuan”) dalam data merupakan hal yang sulit karena sering kali tersembunyi di wilayah dengan dimensi yang sangat tinggi sehingga tidak dapat diakses oleh manusia (Holzinger, 2019).

Menurut (Choi, et al., 2020), terdapat Empat teknik pembelajaran yang populer dalam *machine learning*, yaitu *supervised*, *unsupervised*, *semi-supervised*, dan *reinforcement learning*. *Supervised learning* memetakan fitur ke target menggunakan pola yang ditemukan dalam kumpulan data pelatihan. Metode ini diawasi karena target menginformasikan model tentang apakah ia membuat prediksi yang benar setelah model menyimpulkan algoritme menggunakan pasangan fitur-target. Berbeda dengan *supervised learning*, *unsupervised learning* bertujuan untuk mendeteksi pola dalam dataset dan mengkategorikan masing-masing contoh dalam dataset ke dalam kategori tersebut. Algoritme ini bersifat tidak terawasi karena pola yang mungkin ada atau tidak ada dalam dataset tidak diinformasikan oleh target dan dibiarkan ditentukan oleh algoritme. Selain itu, *semi-supervised learning* dapat dianggap sebagai "jalan tengah" antara *supervised* dan *unsupervised learning*, serta sangat berguna untuk dataset yang mengandung data berlabel dan tidak berlabel. Terakhir, *reinforcement learning* adalah teknik melatih algoritma untuk tugas tertentu di mana tidak ada satu jawaban yang benar, tetapi hasil keseluruhan yang diinginkan. Teknik ini bisa dibilang merupakan upaya yang paling mendekati dalam memodelkan pengalaman belajar manusia karena ia juga belajar dari proses *trial* dan *error*, dan bukan hanya dari data semata.

Jurusan Teknik Informatika dan Komputer – Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

2.4 Deep Learning

Deep learning (DL) adalah cabang khusus dari machine learning (ML). Awalnya terinspirasi oleh model biologis komputasi dan kognisi di otak manusia, salah satu kekuatan utama DL adalah potensinya untuk mengekstrak fitur tingkat tinggi dari data input mentah. Sistem konvensional di berbagai aspek kehidupan sehari-hari dengan cepat tergantikan oleh aplikasi deep learning (DL), seperti pada pengenalan gambar dan suara, pencarian di web, deteksi penipuan, penyaringan email atau spam, hingga pemodelan risiko keuangan. Teknik DL telah membuktikan kemampuannya yang luar biasa di berbagai bidang, termasuk bermain Go, kendaraan otonom, navigasi, desain chip, fisika partikel, ilmu protein, penemuan obat, astrofisika, dan pengenalan objek. Belakangan ini, metode DL telah melampaui teknik pembelajaran mesin lainnya di sejumlah disiplin ilmu, seperti kimia, fisika, biologi, dan ilmu material. Dalam ilmu dan rekayasa material (MSE), penerapan DL masih dalam tahap awal, sehingga potensi, dampak, dan keterbatasannya belum sepenuhnya terungkap. Pendekatan DL membuka cara baru untuk memahami fenomena material dan mendorong para ilmuwan material memperluas alat-alat tradisional mereka. Sebagai pelengkap metode berbasis fisika dalam desain material, DL telah terbukti efektif. Meskipun sering dianggap bahwa dataset besar diperlukan untuk keberhasilan aplikasi DL, teknik seperti transfer learning, pemodelan multi-fidelity, dan active learning memungkinkan DL tetap relevan meskipun menggunakan dataset kecil (Choudhary, et al., 2022).

2.5 Python

Menurut (Dhruv, et al., 2022), Python adalah bahasa pemrograman berorientasi objek yang dapat ditafsirkan, interaktif, dan mudah digunakan. Struktur data tingkat tinggi seperti daftar, tuple, set, array asosiatif (juga dikenal sebagai kamus), pengetikan dan pengikatan dinamis, modul, kelas, pengecualian, dan manajemen memori otomatis adalah semua tersedia dalam bahasa ini. Selain digunakan dalam sistem komputasi paralel, bahasa pemrograman ini memiliki sintaks yang cukup sederhana dan mudah dikodekan, tetapi masih merupakan bahasa pemrograman yang kuat. Python memiliki penerjemah untuk Java yang disebut JPython, yang mirip dengan penerjemah untuk bahasa C. Ada banyak keunggulan Python

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

dibandingkan bahasa lain, seperti ketersediaan berbagai pustaka yang memungkinkan programmer mengurangi jumlah kode hingga sepertiga. Oleh karena itu, Python dianggap sebagai bahasa terbaik dalam hal pembelajaran mesin. Sulit bagi banyak orang untuk memecahkan masalah.

2.6 Tensorflow

TensorFlow adalah pustaka yang dapat diskalakan dan beradaptasi untuk komputasi numerik berbasis grafik aliran data. Pengguna dapat memprogram, melatih, dan menggunakan jaringan saraf dan model *machine learning* lainnya secara efektif untuk produksi dengan bantuan pustaka ini dan alat terkait. Pustaka ini berisi ide-ide dasar termasuk alat bantu tampilan TensorBoard, alat bantu eksekusi grafik, dan fungsi pembuatan grafik (Pang, et al., 2020).

2.7 Mediapipe

Banyak *framework* dan pustaka *machine learning* untuk deteksi gerakan tangan telah dikembangkan untuk memfasilitasi pengembangan aplikasi berbasis AI (Artificial Intelligence). MediaPipe adalah salah satunya. Deteksi Wajah, Face Mesh, Iris, Tangan, Pose, Holistik, Segmentasi Rambut, Deteksi Objek, dan Pelacakan Gerakan Instan adalah beberapa teknologi pembelajaran mesin yang ditawarkan Google melalui kerangka kerja MediaPipe. Salah satu manfaat mengadopsi karakteristik *framework* MediaPipe adalah memungkinkan pengembang untuk berkonsentrasi pada pengembangan algoritma dan model aplikasi sekaligus mendukung aplikasi lingkungan dengan menghasilkan hasil yang dapat direproduksi di banyak platform dan perangkat (Indriani, et al., 2021).

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

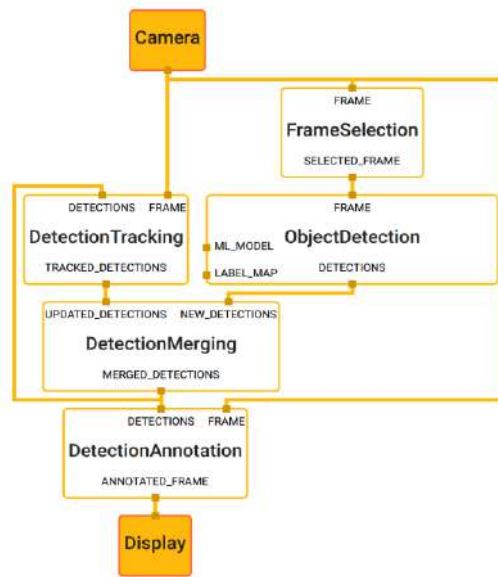
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta

2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 2. 1 Deteksi Obyek Menggunakan MediaPipe
(Sumber: (Lugaresi, et al., 2019))

Dengan MediaPipe, pipeline perception dapat dibangun sebagai grafik komponen modular, termasuk, misalnya, model inferensi dan fungsi pemrosesan media. Grafik menerima data sensorik, seperti umpan audio dan video, dan menghasilkan deskripsi yang dipahami, seperti anotasi tengara wajah atau temuan pendeteksian objek. Sebuah ilustrasi ditampilkan dalam Gambar (Lugaresi, et al., 2019).

2.8 *Connectionist Temporal Classification (CTC)*

Teknik pembelajaran *sequence* yang sangat disukai untuk menyelaraskan urutan input dan target adalah *Connectionist Temporal Classification (CTC)*. CTC mengintegrasikan semua potensi penyelarasan selama pelatihan model daripada bergantung pada penyelarasan sebelumnya antara urutan input dan output (Li, et al., 2020). Pendekatan CTC mengandalkan pemrograman dinamis, algoritma Forward-Backward, untuk menghitung *loss function* tertentu berdasarkan probabilitas jalur yang mungkin yang dapat menghasilkan *output label sequence*. Salah satu cara terbaik untuk menghilangkan pra-segmentasi sampel pelatihan dan pasca-pemrosesan untuk mengubah output dari jaringan saraf tiruan menjadi urutan label adalah dengan menggunakan CTC. Satu unit per label ada di lapisan akhir layer,

atau layer softmax, yang menghasilkan kemungkinan memiliki label yang cocok pada langkah waktu tertentu (Soullard, et al., 2019).

2.9 *Attention Mechanism*

Kapasitas sistem visual manusia untuk berkonsentrasi pada area utama yang mendukung identifikasi objek memberikan dasar bagi operasi *Attention Mechanism*. Awalnya ditujukan untuk tugas-tugas yang melibatkan pemrosesan bahasa alami, jaringan berbasis *Attention* sekarang umumnya digunakan dalam visi computer (Kumari & Anand, 2024). Menurut penelitian psikologi kognitif, manusia hanya mendeteksi sebagian dari informasi yang terlihat karena adanya hambatan pemrosesan informasi. Termotivasi oleh mekanisme *Attention* visual ini, para peneliti telah berusaha mengembangkan model *Attention* selektif visual yang meniru persepsi visual manusia untuk menggambarkan bagaimana perhatian manusia didistribusikan ketika melihat gambar dan video dan untuk memperluas penggunaannya. Kemajuan yang signifikan dalam proses *Attention* telah dibuat dalam domain bahasa alami dan pemrosesan gambar dalam beberapa tahun terakhir. Selain sesuai dengan mekanisme persepsi otak dan mata manusia, strategi mekanisme *Attention* telah terbukti meningkatkan kinerja model. Sebagai contoh, sebagian besar penelitian yang menggabungkan deep learning dan *Attention Mechanism* visual berfokus pada penggunaan mask di bidang *computer vision*. Ide di balik mask adalah lapisan tambahan dengan bobot baru yang mengidentifikasi elemen-elemen penting dalam data gambar. Jaringan syaraf tiruan dapat membentuk perhatian dengan belajar dan berlatih untuk mengidentifikasi bagian-bagian dari setiap gambar baru yang membutuhkan perhatian (Yang, 2020).

2.10 *Beam Search*

Beam Search merupakan algoritma pencarian heuristik yang umum digunakan sebagai metode decoding dalam model sequence-to-sequence. Algoritma ini bekerja dengan mempertahankan sejumlah kandidat output terbaik (sesuai nilai probabilitasnya) pada setiap langkah generasi, sehingga mampu menelusuri beberapa kemungkinan hasil secara paralel daripada memilih satu keluaran secara greedy. Pendekatan tersebut terbukti efektif meningkatkan kualitas hasil pada

Hak Cipta :

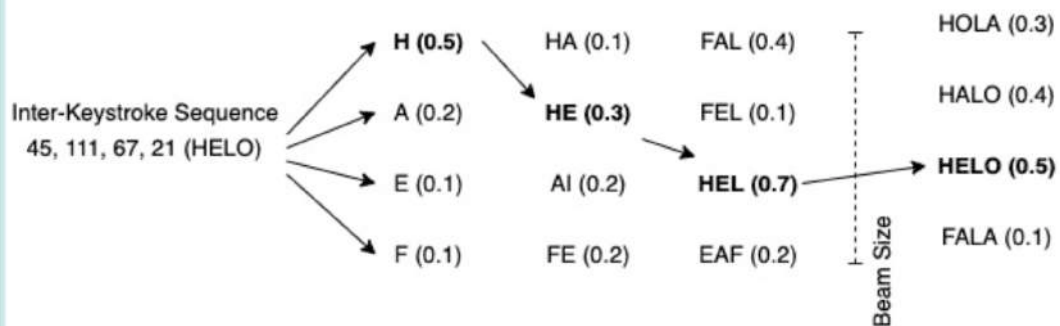
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta

2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

berbagai tugas sequence-to-sequence, termasuk penerjemahan mesin dan pengenalan ucapan otomatis, di mana Beam Search membantu menemukan urutan kata terjemahan atau transkripsi dengan probabilitas tertinggi dibandingkan metode naif. Dalam konteks tersebut, Beam Search telah menjadi teknik decoding standar untuk menghasilkan keluaran yang lebih akurat pada model *neural generative* (Kasai, et al., 2024). Algoritma ini bekerja dengan cara mempertahankan beberapa (n) kandidat sekuens terbaik—disebut juga hipotesis—secara paralel pada setiap langkah. Kandidat-kandidat ini, yang dikenal sebagai beam, diperluas dengan memprediksi kata atau token berikutnya. Dari semua kemungkinan yang ada, hanya n hipotesis dengan probabilitas tertinggi yang akan dilanjutkan ke tahap berikutnya. Proses ini dapat disempurnakan dengan constrained decoding, yaitu sebuah metode untuk menyaring hipotesis yang ditambahkan ke dalam beam, sehingga menjamin hanya kandidat yang valid yang akan diproses pada setiap langkah (Shapiro, et al., 2023)



Gambar 2. 2 Mekanisme *Beam Search*
(Sumber: (Siahaan, et al., 2025))

2.11 Kalman Filter

Filter Kalman adalah serangkaian persamaan matematis yang menyediakan solusi komputasi yang efisien dan rekursif dari metode least-squares. Filter ini sangat andal karena mampu melakukan estimasi terhadap keadaan di masa lalu, masa kini, dan bahkan masa depan, bahkan ketika sifat dari sistem yang dimodelkan tidak diketahui secara pasti. Filter ini secara efektif dapat mengurangi *error* (galat) dan meningkatkan akurasi, terutama jika terdapat *noise* (derau). Keunggulannya mencakup penyediaan estimasi yang optimal, dukungan untuk pemrosesan real-

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

time, kemampuan menangani ketidakpastian, serta fleksibilitas untuk beradaptasi dengan berbagai model dan fusi data sensor. Dalam bidang visi komputer, Filter Kalman banyak digunakan untuk pelacakan objek atau gestur karena membantu memperhalus lintasan, mengurangi noise pada deteksi, memprediksi pergerakan objek, serta meningkatkan stabilitas dan akurasi pelacakan (Fan, 2024).

Kombinasi antara Filter Kalman dan MediaPipe menawarkan solusi yang efektif untuk berbagai tantangan umum pada sistem pelacakan pandangan mata (gaze tracking). Hal ini didukung oleh temuan (Patel & Singh, 2022) yang membuktikan bahwa penggabungan kedua teknologi pada aplikasi AR berhasil meningkatkan akurasi sekaligus mengurangi noise. Riset mereka juga menyoroti bahwa pendekatan gabungan ini mampu beradaptasi di beragam situasi dengan kalibrasi yang minim. Sejalan dengan itu, studi oleh (Sharma, 2023) pada produk VR juga menunjukkan manfaat nyata dari integrasi ini, terutama dalam menyempurnakan presisi pelacakan pandangan dan meningkatkan kenyamanan pengguna (Ramesh, et al., 2025).

2.12 Penelitian Terdahulu

Sampai saat ini, Pengenalan bahasa isyarat (SLR) telah dikembangkan di beberapa penelitian. Berbagai metode deep learning telah diimplementasikan untuk mengembangkan model SLR ini. Para peneliti telah mengembangkan teknologi SLR yang sama dengan berbagai pendekatan yang berbeda. Seperti penelitian yang dilakukan oleh (Li, et al., 2020) menggunakan pendekatan Pose RNN dan TCGN untuk pengembangan model SLR. Disisi lain, peneliti (Laines, et al., 2023) mengembangkan model yang sama namun dengan metode *spatio-temporal skeleton representation*. Kedua penelitian tersebut menghasilkan akurasi yang baik, yaitu sebesar 89.92% dan 98%, namun model yang dihasilkan hanya terbatas pada *Isolated Sign Language Recognition* (ISLR).

Dari studi literatur yang penulis lakukan, penulis belum menemukan adanya penelitian terkait *Continuous Sign Language Recognition* (CSLR) yang dilatih dengan Bahasa Indonesia. Namun, terdapat beberapa penulisan yang berfokus pada CSLR, seperti yang ditulis oleh (Zhu, et al., 2023) menggunakan *Temporal Super-*

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta

2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Resolution Network (TSRNet) dan (Zhou, et al., 2022) dengan metode *Spatial-Temporal Multi-Cue Network* (STMC). Performa kedua penelitian tersebut diukur menggunakan *Word Error Rate* (WER).

Di sisi lain, Penulis menemukan penelitian yang dapat mendukung penelitian CSLR penulis. Ditulis oleh (Li, et al., 2020), pengembangan model dilakukan menggunakan *CTC-Attention* dan menggunakan dataset Bahasa Mandarin. Penelitian ini menggabungkan *Connectionist Temporal Classification* (CTC) dan *Attention Mechanism*, dengan hasil WER sebesar 9.1%.





Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

BAB III METODE PENELITIAN

3.1 Rancangan Penelitian

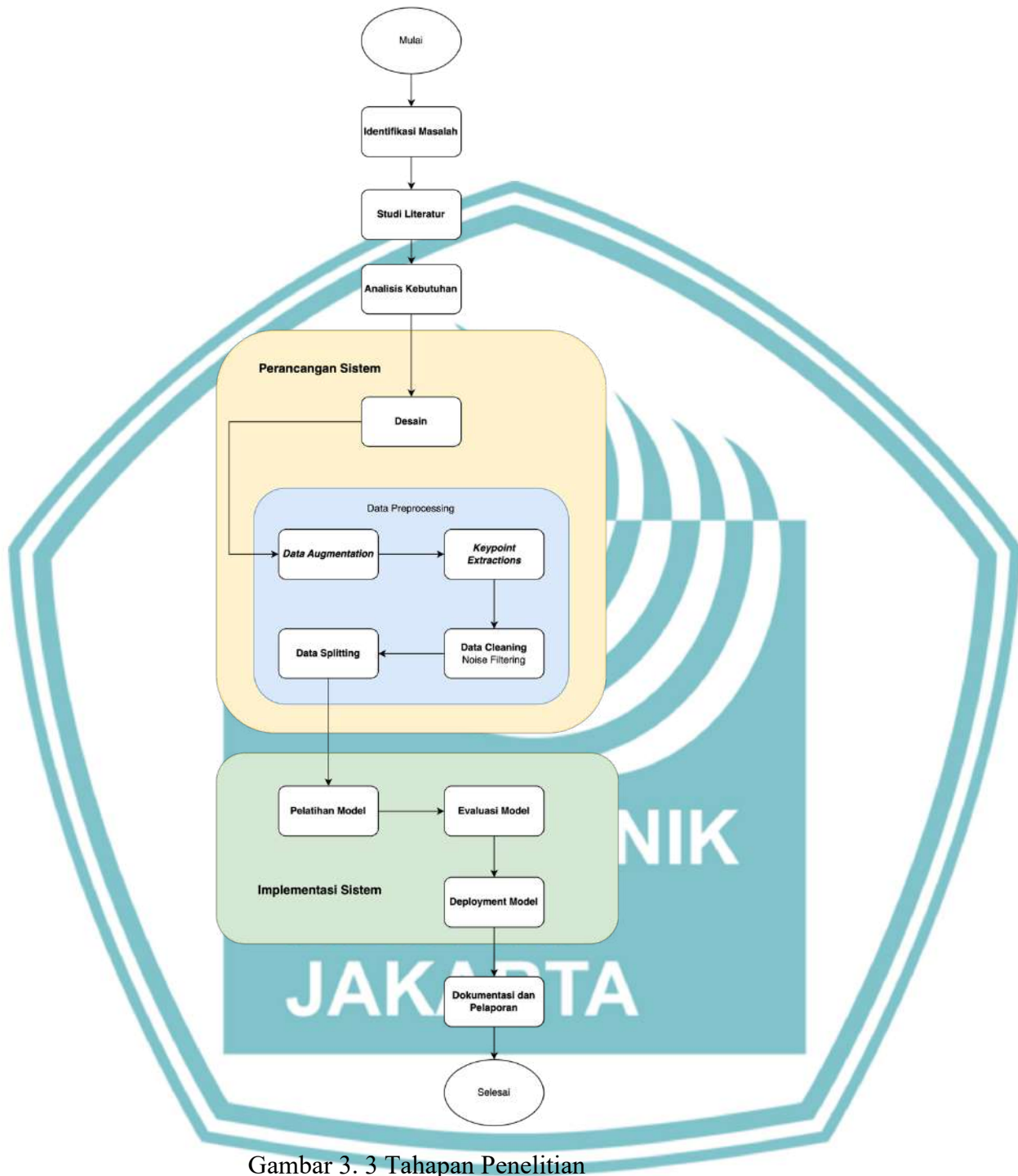
Dengan menggunakan *keypoints* yang diambil menggunakan MediaPipe dan dilatih menggunakan kombinasi teknik *Connectionist Temporal Classification* (CTC) dan *Attention Mechanism*, penelitian ini bertujuan untuk membangun model Pengenalan Bahasa Isyarat Kontinu (*Continuous Sign Language Recognition*, CSLR). Penelitian ini menggunakan metodologi kuantitatif, menganalisis data numerik tentang kinerja model pembelajaran mesin yang ditentukan oleh metrik Word Error Rate (WER). Penelitian ini juga termasuk dalam kategori penelitian eksperimental karena menggabungkan model yang sudah ada melalui penggunaan sistem berbasis web. Teknik uji eksperimental digunakan dalam prosedur pengumpulan data, yang melibatkan penggunaan MediaPipe untuk mengubah kumpulan data gerakan bahasa isyarat menjadi kumpulan titik-titik kunci spasial-temporal. Pelatihan dan pengujian model serta menilai kinerja sistem dalam keadaan yang sebenarnya adalah bagian dari prosedur eksperimental. Survei pengguna juga dilakukan untuk mengukur seberapa baik sistem berbasis web bekerja, dengan mempertimbangkan faktor-faktor seperti akurasi, kegunaan, dan fungsionalitas. Tujuan utama dari penelitian ini adalah untuk menciptakan model CSLR yang dapat secara konsisten dan akurat menginterpretasikan gerakan bahasa isyarat. Selain itu, penelitian ini berusaha untuk menunjukkan kesulitan dalam mengintegrasikan model ke dalam aplikasi berbasis web dan menawarkan saran untuk kemajuan di masa depan dalam model pembelajaran mesin dan integrasi sistem secara umum.

3.2 Tahapan Penelitian

Tahapan penelitian dirancang untuk membimbing penulis dalam melakukan penelitian dengan sistematis dan terstruktur. Pada penelitian ini terdiri dari beberapa tahapan proses, yaitu studi literatur, pengumpulan data, preprocessing, pelatihan model, dan evaluasi hasil.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 3. 3 Tahapan Penelitian

1. Identifikasi Masalah

Penelitian dimulai dengan mengidentifikasi permasalahan utama, yaitu belum tersedianya sistem terotomatisasi yang mampu menerjemahkan gesture bahasa isyarat Indonesia secara kontinu menjadi teks, khususnya yang mendukung Sistem Isyarat Bahasa Indonesia (SIBI). Masalah ini berdampak pada kesenjangan komunikasi antara penyandang tunarungu dan



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

masyarakat umum, sehingga diperlukan solusi teknologi asistif berbasis pengenalan bahasa isyarat.

2. Studi Literatur

Setelah masalah dirumuskan, dilakukan penelusuran dan kajian pustaka untuk memahami perkembangan terkini dalam bidang Sign Language Recognition (SLR). Kajian mencakup berbagai pendekatan seperti SSLR, ISLR, dan CSLR, serta teknik-teknik yang digunakan dalam penelitian terdahulu seperti MediaPipe untuk ekstraksi fitur, algoritma CTC dan Attention, serta implementasi decoding menggunakan Beam Search. Studi ini menjadi dasar dalam memilih metode dan arsitektur model yang sesuai untuk penelitian.

3. Analisis Kebutuhan

Pada tahap ini dilakukan analisis terhadap kebutuhan sistem baik dari sisi fungsional, seperti proses inferensi gesture ke teks secara real-time, maupun non-fungsional, seperti kompatibilitas antarmuka web dan akurasi prediksi. Selain itu, ruang lingkup penelitian dibatasi pada gestur-gestur dasar dalam kategori anggota keluarga, benda sekitar, dan binatang.

4. Perancangan Sistem

Merancang alur sistem secara menyeluruh, mulai dari tahapan ekstraksi keypoints tiga dimensi menggunakan MediaPipe Holistic, preprocessing data termasuk normalisasi dan augmentasi, hingga rancangan arsitektur encoder-decoder berbasis BiLSTM dan BiGRU. Proses decoding dirancang menggunakan Beam Search untuk menghasilkan prediksi teks yang lebih akurat.

5. Implementasi Sistem

Pada tahap ini, model dilatih menggunakan data yang telah diproses dan dibagi sesuai kebutuhan training dan testing. Model kemudian diintegrasikan dalam platform web berbasis Flask REST API dan WebSocket, yang memungkinkan proses inferensi dilakukan baik dari video unggahan maupun input webcam secara langsung.

6. Pengujian dan Evaluasi Model



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Evaluasi dilakukan terhadap performa model menggunakan data unseen dengan metrik kuantitatif seperti Word Error Rate (WER), CTC Loss, dan Attention Loss. Pengujian ini juga mencakup visualisasi confusion matrix untuk menganalisis distribusi kesalahan antar prediksi kata, serta membandingkan hasil model dengan dan tanpa augmentasi.

7. Evaluasi dan Dokumentasi

Seluruh tahapan dan hasil penelitian didokumentasikan secara sistematis dalam bentuk laporan skripsi. Pada tahap ini juga disusun kesimpulan dan rekomendasi untuk pengembangan sistem CSLR di masa depan.

3.3 Objek Penelitian

Objek dalam penelitian ini adalah masyarakat umum yang memiliki keterbatasan atau minimnya pengetahuan tentang bahasa isyarat dan ingin mempelajarinya melalui aplikasi berbasis website. Penelitian ini berfokus pada bagaimana aplikasi dapat memfasilitasi pengguna dalam memahami serta mempraktikkan bahasa isyarat secara interaktif dengan dukungan model *Continuous Sign Language Recognition* (CSLR) yang diimplementasikan menggunakan MediaPipe dan CTC-Attention.

**POLITEKNIK
NEGERI
JAKARTA**

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisis Kebutuhan

Pada tahap ini dilakukan analisis kebutuhan sistem untuk mendukung pengembangan model *Continuous Sign Language Recognition* (CSLR) menggunakan kombinasi pendekatan MediaPipe Holistic sebagai alat ekstraksi fitur dan arsitektur Joint CTC-Attention untuk pemodelan sekuens. Analisis kebutuhan ini disusun berdasarkan kebutuhan teknis yang dibutuhkan selama proses pengembangan, pelatihan, dan evaluasi model secara end-to-end. Analisis ini terbagi menjadi kebutuhan fungsional dan non-fungsional untuk memastikan sistem yang dibangun dapat berfungsi secara efektif, efisien, dan stabil di lingkungan komputasi yang terbatas.

4.1.1 Kebutuhan Pengembangan Model

4.1.1.1 Kebutuhan Dataset

Dalam penelitian ini, pengembangan model *Continuous Sign Language Recognition* (CSLR) difokuskan pada bahasa isyarat SIBI dengan konteks dasar sesuai buku “Bahasa Isyarat untuk Anak Tunarungu (Dasar)” oleh Nattaya Lakshita. Oleh karena itu, dibutuhkan dataset video berisi kalimat dalam bahasa isyarat Indonesia dengan struktur gestur alami dan berkelanjutan. Kebutuhan dataset dirancang agar mampu mewakili kombinasi makna dari kata dasar yang dikelompokkan ke dalam tiga kategori utama: Anggota Keluarga, Benda Sekitar, dan Binatang.

Dalam pengumpulan dataset video untuk pelatihan model CSLR, terdapat sejumlah kriteria kualitas yang ditetapkan guna memastikan bahwa data yang digunakan sesuai dengan kebutuhan sistem dan menghasilkan model yang akurat. Pertama, seluruh video direkam menggunakan kamera dengan resolusi tinggi untuk memastikan ketajaman gambar dan kemudahan dalam proses ekstraksi *keypoints*. Pencahayaan yang stabil agar deteksi gestur tidak terganggu oleh bayangan atau



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

pencahayaan yang berlebihan. Selain itu, bahasa isyarat yang digunakan dalam setiap video harus sesuai dengan Sistem Isyarat Bahasa Indonesia (SIBI), bukan BISINDO, agar model yang dikembangkan tetap konsisten dengan standar formal yang digunakan dalam pendidikan inklusif di Indonesia. Gerakan isyarat dilakukan dengan tempo yang alami, tidak terlalu cepat maupun terlalu lambat, serta dilakukan secara konsisten di seluruh dataset untuk menghindari variasi yang berlebihan antar video. Durasi setiap video disesuaikan agar mencakup satu kalimat utuh, biasanya dalam rentang 2 hingga 5 detik, sehingga cukup menangkap konteks dan dinamika gerakan dalam satuan waktu yang representatif. Tabel 4.1 menampilkan jumlah distribusi kata di dalam seluruh kalimat yang muncul pada dataset.

Tabel 4. 1 Distribusi Kata yang muncul pada dataset

Kata	Frekuensi
dan	420
suka	340
aku	236
punya	136
susu	128
minum	124
tidak	120
makan	120
tomat	104
belajar	100
di	88
tidur	84
jeruk	80
rusa	80
air	80
kakak	76
adik	76
ikan	72
ibu	68
tempat	68
me	64
dengan	64
jerapah	64
singa	60
boneka	60
bola	60
brokoli	56
lihat	56



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Kata	Frekuensi
kucing	56
paman	52
kentang	52
telur	52
gajah	52
pisang	48
nenek	48
main	48
ayah	44
monyet	44
kakek	44
bebek	40
mau	40
payung	36
sapi	36
ada	32
bibi	32
ini	32
domba	28
es	28
ber	28
besar	28
langit	24
kebun	12
itu	12
untuk	12
buku	12

4.1.1.2 Kebutuhan Prapemrosesan Data

Pada tahap preprocessing data, dilakukan serangkaian proses untuk mengubah video bahasa isyarat menjadi data numerik yang dapat digunakan sebagai input model. Proses diawali dengan membaca metadata video yang berisi nama file dan transkrip kalimat. Setiap video kemudian diproses *frame per frame* menggunakan MediaPipe Holistic untuk mendeteksi titik-titik penting (*keypoints*) pada tubuh, wajah, dan tangan. Hasil deteksi tersebut mencakup koordinat tiga dimensi (x, y, z) serta skor visibilitas untuk 33 titik tubuh, 468 titik wajah, dan masing-masing 21 titik untuk tangan kiri dan kanan, yang kemudian dikonversi menjadi vektor satu dimensi dan disimpan dalam array NumPy.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Untuk meningkatkan keragaman data dan mengurangi risiko overfitting, diterapkan teknik augmentasi seperti flipping horizontal, penyesuaian tingkat kecerahan (brightness adjustment), serta *frame* skipping (mengambil setiap dua *frame* untuk mempercepat sekuen). Augmentasi ini menghasilkan beberapa variasi dari satu video asli, sehingga memperbanyak jumlah data pelatihan tanpa perlu merekam ulang. Setiap hasil ekstraksi disimpan sebagai file .npy yang mewakili satu sekuen gerakan isyarat dalam bentuk urutan *keypoints*. Proses ini memastikan bahwa data yang digunakan untuk pelatihan model tidak hanya bersih dan terstandarisasi, tetapi juga cukup beragam untuk meningkatkan kemampuan generalisasi model dalam mengenali berbagai variasi gestur.

4.1.1.3 Kebutuhan Perangkat Lunak

Dalam proses pengembangan model *Continuous Sign Language Recognition* (CSLR), digunakan sejumlah perangkat lunak yang berfungsi dalam proses ekstraksi fitur, pelatihan model, dan evaluasi performa. Bahasa pemrograman utama yang digunakan adalah Python, karena memiliki dukungan pustaka yang lengkap untuk kebutuhan pemrosesan data dan pengembangan model deep learning. Untuk proses ekstraksi *keypoints* dari video, digunakan pustaka MediaPipe, yang mampu mendeteksi landmark tubuh, tangan, dan wajah secara real-time. Proses pelatihan model dilakukan menggunakan TensorFlow, sebuah *framework* deep learning yang mendukung implementasi arsitektur Connectionist Temporal Classification (CTC) dan *Attention* Mechanism, serta dapat berjalan secara optimal di GPU. Selain itu, pustaka seperti NumPy, Pandas, dan OpenCV digunakan untuk manipulasi data dan pengolahan video, sementara scikit-learn digunakan untuk keperluan pembagian data dan preprocessing. Seluruh proses pelatihan model dilakukan pada platform Kaggle, yang menyediakan akses ke GPU NVIDIA T4 secara gratis terbatas untuk mempercepat proses training dalam lingkungan yang terintegrasi dan siap pakai.

Tabel 4. 2 Kebutuhan Perangkat Lunak Pengembangan Model

Nama	Spesifikasi
Python	Bahasa pemrograman utama untuk pengembangan model



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Nama	Spesifikasi
MediaPipe	Library untuk ekstraksi <i>keypoints</i> dari video
Tensorflow	<i>Framework</i> deep learning untuk pelatihan model CSLR
NumPy, Pandas	Pustaka manipulasi data numerik dan struktural
OpenCV	Untuk pembacaan dan pemrosesan video <i>frame</i>
scikit-learn	Untuk data splitting dan preprocessing
Kaggle Notebook	Platform pelatihan model berbasis cloud dengan GPU T4
Matplotlib	Untuk visualisasi data dan hasil pelatihan model

4.1.1.4 Kebutuhan Perangkat Keras

Dalam proses penelitian ini, kebutuhan perangkat keras dibedakan menjadi dua tahap utama, yaitu tahap ekstraksi fitur dan tahap pelatihan model. Untuk tahap ekstraksi *keypoints* dari video bahasa isyarat, digunakan perangkat komputasi lokal dengan yang dapat dilihat pada Tabel 4.3 Tabel Kebutuhan Perangkat Keras Pengembangan Model.

Tabel 4. 3 Tabel Kebutuhan Perangkat Keras Pengembangan Model

Nama	Spesifikasi
Laptop	CPU berbasis arsitektur ARM 8-core dan RAM sebesar 8 GB

Spesifikasi ini dinilai cukup untuk menjalankan proses ekstraksi menggunakan MediaPipe Holistic secara efisien pada resolusi video standar. Sementara itu, untuk



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

tahap pelatihan model *Continuous Sign Language Recognition* (CSLR) yang membutuhkan sumber daya komputasi lebih tinggi, digunakan platform komputasi awan dengan dukungan dua unit GPU NVIDIA T4, masing-masing memiliki memori 16 GB dan kemampuan akselerasi inferensi yang optimal untuk deep learning. Pemanfaatan GPU ini memungkinkan proses pelatihan berlangsung lebih cepat dan efisien, terutama mengingat kompleksitas model berbasis *CTC-Attention* yang digunakan dalam penelitian ini.

4.1.2 Kebutuhan *Model Deployment*

4.1.2.1 Kebutuhan Perangkat Lunak

Dalam proses *Model Deployment*, perangkat lunak dibutuhkan untuk membangun REST API sekaligus WebSocket yang memungkinkan komunikasi secara real-time antara client dan server. *Framework* utama yang digunakan adalah Flask sebagai basis pengembangan REST API, serta Flask-SocketIO untuk menangani pengiriman data *frame* video secara berkelanjutan dari client ke server. Selain itu, editor kode seperti Visual Studio Code digunakan untuk pengembangan dan pengujian sistem secara lokal sebelum diunggah ke server produksi. Detail kebutuhan perangkat lunak ditampilkan pada Tabel 4.4 Tabel Kebutuhan Perangkat Lunak *Model Deployment*.

Tabel 4. 4 Kebutuhan Perangkat Lunak *Model Deployment*

Nama	Deskripsi
Flask	<i>Framework</i> guna pengembangan REST API
Flask-SocketIO	Library untuk komunikasi data real-time menggunakan WebSocket
Visual Studio Code	Source-code editor untuk pengembangan dan debugging



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

4.1.2.2 Kebutuhan Perangkat Keras

Perangkat keras yang digunakan dalam proses *Model Deployment* bertanggung jawab untuk menjalankan API dan menerima *frame* input secara terus-menerus dari client. Karena sistem berbasis WebSocket bekerja secara real-time, perangkat minimal harus mampu menjalankan server Flask secara stabil dan efisien. Oleh karena itu, dibutuhkan perangkat dengan CPU minimal dual-core dan RAM minimal 4 GB untuk menjamin performa dasar sistem selama proses komunikasi dan inferensi berlangsung. Detail perangkat keras dapat dilihat pada Tabel 4.5 Tabel Kebutuhan Perangkat Keras *Model Deployment*.

Tabel 4. 5 Kebutuhan Perangkat Keras *Model Deployment*

Nama	Deskripsi
Laptop	Dengan minimal RAM 4 GB dan CPU dual-core

4.1.3 Kebutuhan Sistem

4.1.3.1 Kebutuhan Fungsional

Kebutuhan fungsional merupakan fitur utama yang harus dimiliki sistem agar dapat berjalan sesuai tujuan. Kebutuhan-kebutuhan tersebut bisa dilihat di Tabel 4.6 Kebutuhan Fungsional.

Tabel 4. 6 Kebutuhan Fungsional

KODE	Kebutuhan Fungsional
F1	Sistem dapat menerima input dari pengguna berupa file video penuh atau kumpulan <i>frame</i>
F2	Sistem mengunggah input video/ <i>frame</i> ke server untuk diproses lebih lanjut



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

KODE	Kebutuhan Fungsional
F3	Sistem melakukan ekstraksi <i>keypoints</i> dari video/ <i>frame</i> di sisi server menggunakan MediaPipe Holistic
F4	Sistem menyimpan hasil ekstraksi <i>keypoints</i> dalam format numerik (.npy) untuk keperluan inference
F5	Sistem menjalankan inferensi model CSLR terhadap data <i>keypoints</i> dan menghasilkan output berupa kalimat teks
F6	Sistem menampilkan hasil terjemahan gesture dalam antarmuka pengguna berbasis web
F7	Sistem memberikan notifikasi jika video/ <i>frame</i> gagal diproses atau tidak valid

Tabel 4.6 merangkum kebutuhan fungsional yang menjadi fondasi dalam perancangan sistem *Continuous Sign Language Recognition* (CSLR) agar dapat beroperasi secara menyeluruh dan mendukung berbagai skenario penggunaan. Kebutuhan F1, misalnya, diperlukan agar sistem mampu menerima input dari pengguna dalam bentuk file video penuh maupun kumpulan *frame*, sehingga sistem dapat menangani baik pemrosesan secara batch maupun real-time. Kebutuhan F2 memastikan sistem dapat mengunggah input tersebut ke server, sehingga proses lebih lanjut seperti ekstraksi dan inferensi dapat dilakukan secara terpusat dan efisien. Pada tahap berikutnya, sistem harus memenuhi kebutuhan F3, yaitu menjalankan proses ekstraksi *keypoints* menggunakan MediaPipe Holistic di sisi server. Ekstraksi ini penting untuk memperoleh representasi numerik dari gesture



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

pengguna melalui koordinat titik-titik penting pada tubuh, wajah, dan tangan. Selanjutnya, kebutuhan F4 mendukung penyimpanan hasil ekstraksi dalam format file numerik (.npy), sehingga data dapat dengan mudah diolah pada tahap inferensi. Sistem kemudian wajib memenuhi kebutuhan F5, yang menuntut kemampuan untuk menjalankan model CSLR terhadap data *keypoints* tersebut dan menghasilkan kalimat prediksi dalam bentuk teks. Untuk memastikan hasil ini dapat diakses dengan mudah oleh pengguna, kebutuhan F6 mengharuskan sistem menampilkan terjemahan gesture secara langsung di antarmuka web yang responsif. Terakhir, kebutuhan F7 dirancang agar sistem dapat memberikan notifikasi kepada pengguna apabila terjadi kegagalan pemrosesan, baik pada tahap unggah, ekstraksi, maupun inferensi, sehingga proses troubleshooting dapat dilakukan dengan cepat.

Selanjutnya, sistem diharapkan mampu menjalankan inferensi secara optimal menggunakan model *Continuous Sign Language Recognition* (CSLR) yang dibangun dengan arsitektur hybrid *CTC-Attention*. Hal ini bertujuan agar sistem dapat menghasilkan kalimat teks yang merepresentasikan terjemahan gesture bahasa isyarat secara akurat, sesuai dengan yang diharapkan dalam kebutuhan F5. Kalimat hasil prediksi ini kemudian ditampilkan melalui antarmuka web yang telah dirancang agar sederhana, intuitif, dan mudah dioperasikan oleh pengguna dari berbagai latar belakang, sebagaimana dimaksud dalam kebutuhan F6. Sistem juga dilengkapi dengan mekanisme notifikasi sebagai bagian dari kebutuhan F7. Fitur ini memungkinkan sistem memberikan peringatan apabila terjadi masalah selama pemrosesan, misalnya ketika file video atau *frame* yang diunggah tidak valid, data mengalami kerusakan, atau tidak terdapat gesture yang dapat dikenali.

4.1.3.2 Kebutuhan Non-Fungsional

Kebutuhan non-fungsional menggambarkan karakteristik atau kualitas sistem yang harus dipenuhi agar sistem dapat digunakan dengan nyaman, cepat, dan andal. Kebutuhan-kebutuhan tersebut dipaparkan di Tabel 4.7 Tabel Kebutuhan Non-Fungsional.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Tabel 4. 7 Kebutuhan Non-Fungsional

KODE	Kebutuhan Fungsional
NF2	Sistem mampu berjalan pada browser modern seperti Chrome, Firefox, atau Safari
NF3	Sistem memiliki akurasi prediksi dengan Word Error Rate (WER) di bawah 30%
NF4	Sistem tetap stabil meskipun pengguna memberikan input dengan variasi pencahayaan berbeda

Tabel 4.7 memuat kebutuhan non-fungsional yang dirumuskan untuk mendukung kualitas, keandalan, dan kenyamanan sistem *Continuous Sign Language Recognition* (CSLR) yang dikembangkan dalam penelitian ini. Kebutuhan non-fungsional ini berfokus pada aspek kompatibilitas, performa, dan stabilitas agar sistem dapat dioperasikan dengan lancar dalam berbagai kondisi penggunaan. Salah satu kebutuhan utama adalah kemampuan sistem untuk berjalan secara optimal pada browser modern yang banyak digunakan, seperti Google Chrome, Mozilla Firefox, maupun Safari, sebagaimana tercantum dalam kebutuhan NF2. Tujuan dari persyaratan ini adalah memastikan bahwa sistem dapat diakses secara luas tanpa perlu instalasi tambahan ataupun konfigurasi perangkat lunak khusus, sehingga memudahkan pengguna dari berbagai kalangan. Selanjutnya, untuk menjamin performa model dalam menginterpretasikan gestur bahasa isyarat, sistem dirancang agar mampu mencapai tingkat akurasi dengan Word Error Rate (WER) di bawah 30 persen, sesuai dengan kebutuhan NF3. Batas WER ini dipilih agar hasil kalimat prediksi cukup mendekati kalimat referensi, sehingga dapat mendukung tujuan sistem dalam konteks pembelajaran, komunikasi dasar, maupun aplikasi praktis lainnya. Selain performa, aspek stabilitas juga menjadi fokus penting. Sistem harus tetap dapat berfungsi secara konsisten meskipun pengguna merekam gestur dalam kondisi pencahayaan yang bervariasi. Hal ini dituangkan dalam kebutuhan NF4,



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

yang mengharuskan sistem memiliki toleransi terhadap variasi cahaya agar tetap mampu mendeteksi dan memproses gesture dengan akurat.

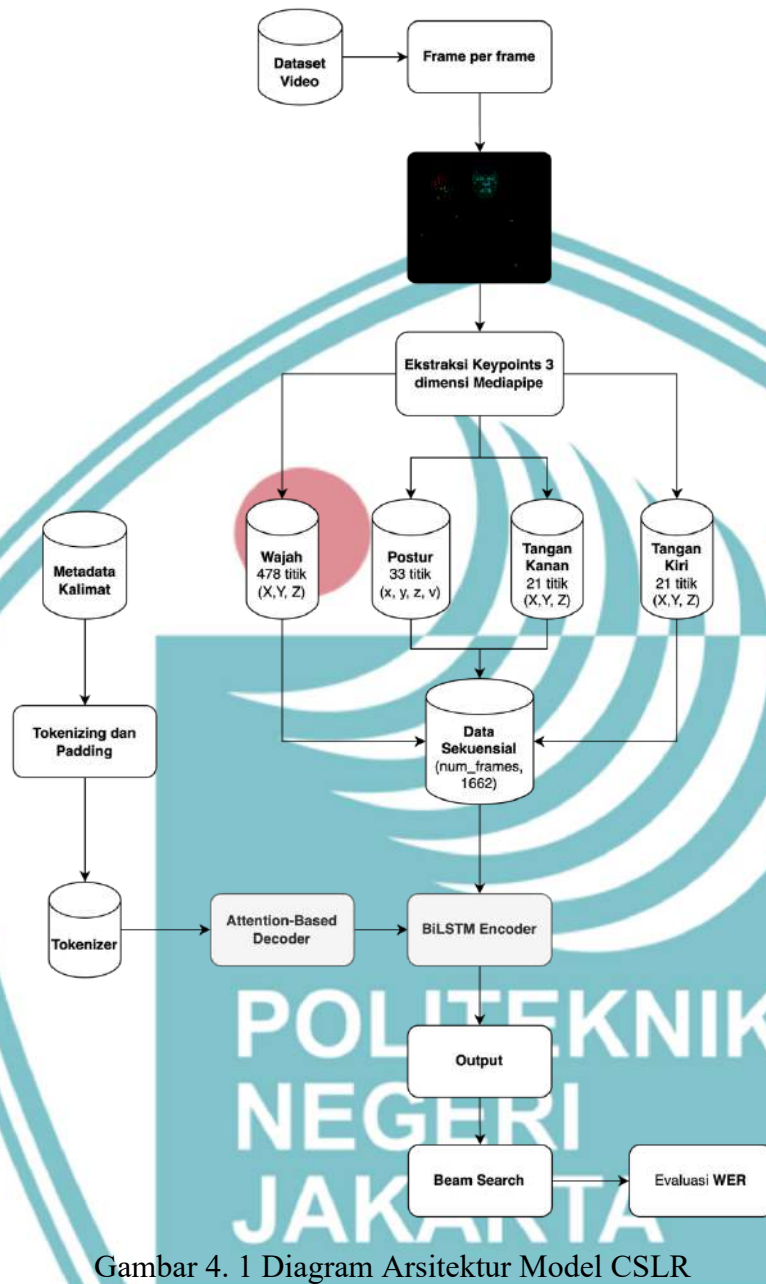
4.2 Desain

Tahap desain dalam penelitian ini mencakup rancangan sistem secara detail mulai dari proses ekstraksi fitur video hingga pelatihan dan evaluasi model *Continuous Sign Language Recognition* (CSLR) menggunakan pendekatan *Joint CTC-Attention*. Diagram yang disajikan pada Gambar 4.1 menggambarkan secara lengkap bagaimana data video bahasa isyarat diproses dari input hingga menghasilkan keluaran berupa kalimat prediksi.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4. 1 Diagram Arsitektur Model CSLR

Pada tahap awal, dataset video diproses dengan mengambil setiap *frame* secara terurut. Setiap *frame* tersebut kemudian diproses menggunakan MediaPipe Holistic untuk mengekstraksi *keypoints* berupa koordinat tiga dimensi (X, Y, Z) yang mewakili titik-titik penting pada gerakan tangan, wajah, dan postur tubuh. MediaPipe menghasilkan output koordinat untuk 33 titik postur tubuh, 468 titik wajah, serta masing-masing tangan sebanyak 21 titik. Selain koordinat tiga dimensi, khusus untuk titik-titik tubuh juga disertai nilai visibilitas yang menunjukkan tingkat kepercayaan deteksi per titik. Secara keseluruhan, setiap *frame* video



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

menghasilkan sebuah vektor berdimensi 1662 setelah semua titik dari wajah, tangan, dan postur tubuh digabungkan dalam satu array. Proses ini menghasilkan matriks berdimensi (jumlah *frame*, 1662) yang kemudian digunakan sebagai input model. Setelah tahap ekstraksi selesai, data sekuensial tersebut selanjutnya diolah oleh model CSLR yang berbasis deep learning dengan dua komponen utama, yaitu *Encoder* dan decoder. Pada tahap *Encoder*, digunakan layer Bi-directional Long Short-Term Memory (BiLSTM) yang berfungsi untuk mengolah informasi temporal dari gerakan isyarat yang terdapat dalam data sekuensial. *Encoder* BiLSTM bekerja dengan cara membaca data sekuensial secara dua arah, baik dari awal ke akhir (forward) maupun sebaliknya (backward), sehingga dapat menangkap pola temporal secara lebih baik. Hasil dari *Encoder* ini adalah representasi laten yang menyimpan informasi konteks temporal dari seluruh *frame* input dalam bentuk vektor berdimensi lebih kecil yang sangat representatif terhadap gerakan isyarat secara keseluruhan.

Tahap berikutnya adalah proses decoding menggunakan *Attention*-Based Decoder yang dilengkapi dengan mekanisme *Attention*. Pada saat proses training, decoder ini menerima dua masukan utama: representasi laten dari *Encoder* serta target token yang sudah diproses menggunakan tokenizer TensorFlow. Mekanisme *Attention* memungkinkan decoder untuk secara dinamis menimbang setiap representasi *Encoder* berdasarkan relevansi terhadap prediksi token yang sedang dilakukan. Dengan demikian, decoder tidak hanya mengandalkan representasi akhir saja, namun dapat berfokus pada bagian-bagian tertentu dari input yang relevan dengan keluaran token yang sedang diprediksi pada setiap langkah waktu decoding.

Proses training sendiri dilakukan dengan pendekatan Joint CTC-*Attention*, yaitu menggabungkan dua jenis loss function, yakni Connectionist Temporal Classification (CTC) dan *Attention*. Loss CTC membantu model belajar untuk melakukan alignment antara urutan input *frame* dengan output token tanpa perlu segmentasi manual, sehingga sangat efektif untuk mengatasi perbedaan panjang antara input dan output. Di sisi lain, loss *Attention* memungkinkan model belajar secara eksplisit tentang konteks urutan bahasa, sehingga menghasilkan prediksi token yang lebih konsisten secara linguistik.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Setelah prediksi token dilakukan oleh decoder, hasil token tersebut kemudian dikonversi kembali menjadi kalimat utuh menggunakan tokenizer yang telah dilatih dengan vocabulary khusus bahasa Indonesia sesuai dataset pelatihan. Evaluasi akhir dari prediksi kalimat dilakukan menggunakan metrik Word Error Rate (WER) yang mengukur akurasi kalimat hasil prediksi dibandingkan dengan kalimat referensi dari metadata asli. Dengan cara ini, performa model CSLR dapat diukur secara kuantitatif untuk memastikan kualitas terjemahan gestur bahasa isyarat menjadi kalimat teks yang relevan.

Evaluasi performa model dilakukan menggunakan metrik Word Error Rate (WER), yang mengukur jumlah kesalahan dalam kalimat hasil prediksi dibandingkan dengan kalimat referensi yang sebenarnya. Untuk mendapatkan hasil terbaik selama evaluasi, model memanfaatkan algoritma decoding Beam Search. Beam Search digunakan karena mampu menghasilkan prediksi dengan mempertimbangkan beberapa kandidat terbaik pada setiap langkah decoding, bukan hanya satu kandidat paling optimal seperti pada greedy decoding. Strategi ini secara signifikan meningkatkan kualitas prediksi karena mampu mempertahankan kandidat yang memiliki peluang tinggi dalam menghasilkan kalimat yang benar secara keseluruhan.

Beam Search bekerja dengan mempertahankan sejumlah kandidat prediksi terbaik dalam setiap tahap decoding berdasarkan skor probabilitas yang diberikan oleh decoder. Kandidat dengan skor tertinggi akan dipilih sebagai prediksi akhir. Penggunaan Beam Search ini terbukti efektif dalam menurunkan WER secara signifikan, memastikan hasil kalimat prediksi memiliki akurasi linguistik yang tinggi serta konsistensi kontekstual.

Dalam proses pelatihan dan inferensi model CSLR, pendekatan tokenisasi digunakan untuk mengubah kalimat bahasa Indonesia menjadi bentuk representasi numerik yang lebih mudah dipahami oleh jaringan saraf tiruan. Setiap kata dalam kalimat diubah menjadi token unik yang merepresentasikan posisi kata tersebut dalam vocabulary. Penggunaan tokenisasi memungkinkan model secara efektif menangkap hubungan spasial-temporal dalam gesture dengan representasi



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

linguistik yang jelas dan konsisten. Selain itu, tokenisasi membantu dalam mereduksi kompleksitas ruang output model, sehingga meningkatkan efisiensi proses training dan inferensi.

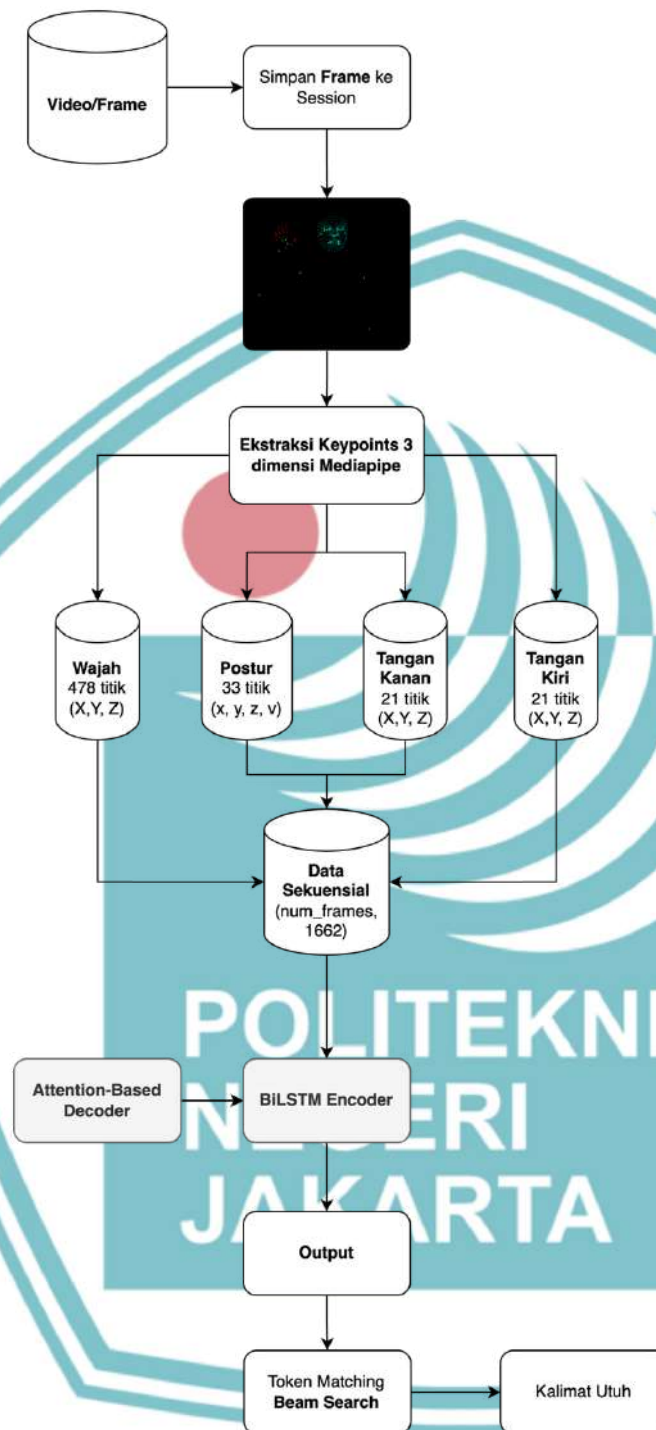
Padding digunakan pada tahap preprocessing baik untuk label kalimat maupun untuk data input *keypoints*. Tujuan utama dari padding adalah menyelaraskan panjang sequence antara berbagai data, karena model deep learning, khususnya yang berbasis sequence-to-sequence, mengharuskan input dalam satu batch memiliki panjang yang sama. Dalam konteks data input, padding memastikan bahwa setiap sequence input memiliki panjang yang seragam, sehingga dapat diolah secara paralel oleh *encoder BiLSTM*. Sementara untuk label kalimat, padding digunakan untuk memastikan bahwa setiap kalimat yang diubah menjadi token memiliki panjang yang konsisten. Padding ini juga disertai penggunaan token khusus (biasanya <PAD>) yang diabaikan selama proses training melalui masking, sehingga tidak mempengaruhi pembelajaran model.

Diagram pada Gambar 4.1 Diagram Arsitektur Model CSLR secara keseluruhan memberikan gambaran visual tentang alur proses data dan bagaimana tiap komponen berinteraksi dalam sistem ini, yang bertujuan tidak hanya untuk meningkatkan pemahaman teknis, tetapi juga memudahkan implementasi model pada aplikasi berbasis web secara real-time.

Selain alur pengembangan model, sistem juga dirancang agar dapat dikonsumsi oleh aplikasi berbasis web secara real-time. Proses ini secara spesifik mencakup interaksi antara klien dan server, mulai dari tahap unggah data hingga hasil prediksi kalimat yang dapat langsung ditampilkan kepada pengguna. Diagram pada Gambar 4.2 Diagram Proses dari sisi klien berikut menjelaskan proses konsumsi layanan model dari sisi klien secara lengkap.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4. 2 Diagram Proses dari sisi klien

Proses konsumsi ini diawali dengan pengguna yang mengirimkan video utuh atau kumpulan *frame* video ke server melalui aplikasi berbasis web. Setelah video atau *frame* diterima, sistem secara otomatis mengumpulkan dan menyimpan terlebih dahulu *frame-frame* tersebut ke dalam session khusus di server hingga mencapai jumlah tertentu, yakni sebanyak 30 *frame*. Proses penyimpanan ini dilakukan agar



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

data memiliki panjang sekuensial yang konsisten sebelum diteruskan ke tahap ekstraksi *keypoints* menggunakan MediaPipe Holistic. Setelah jumlah *frame* yang dikumpulkan mencapai 30, server melanjutkan proses ekstraksi *keypoints* dari setiap *frame* tersebut. Tahap ekstraksi ini menghasilkan koordinat tiga dimensi dari titik-titik wajah, tangan, serta postur tubuh. Seluruh koordinat ini kemudian digabungkan ke dalam bentuk array data sekuensial dengan dimensi tetap (30, 1662).

Data sekuensial tersebut selanjutnya dikirim ke model CSLR berbasis Joint CTC-*Attention* yang telah di-deploy pada layanan cloud engine untuk dilakukan proses inferensi. Proses inferensi diawali dengan *Encoder* BiLSTM yang memproses data sekuensial untuk mendapatkan representasi laten yang mengandung informasi pola temporal gesture. Selanjutnya, decoder berbasis *Attention* menggunakan representasi ini untuk menghasilkan token prediksi berdasarkan *vocabulary tokenizer* TensorFlow. Token prediksi ini kemudian dikonversi menjadi kalimat utuh dalam bahasa Indonesia melalui tokenizer yang telah dilatih sebelumnya. Hasil akhir berupa kalimat prediksi tersebut dikirimkan kembali ke klien untuk ditampilkan secara real-time pada antarmuka pengguna. Rancangan ini memastikan hasil inferensi dapat diberikan dengan cepat dan konsisten, sehingga mampu mendukung interaksi yang lancaya dan efektif pada aplikasi web pembelajaran bahasa isyarat.

4.3 Pengembangan Model

Pengembangan model pada penelitian ini bertujuan untuk membangun sistem *Continuous Sign Language Recognition* (CSLR) berbasis deep learning dengan pendekatan hybrid CTC-*Attention*. Model yang dikembangkan dirancang untuk mengonversi urutan *keypoints* dari wajah, tangan, dan pose tubuh menjadi kalimat natural yang merepresentasikan makna dari gerakan isyarat. Proses pelatihan model mengacu pada arsitektur *Encoder-decoder* yang menggabungkan dua jalur pembelajaran: CTC (Connectionist Temporal Classification) dan *Attention*-based sequence-to-sequence decoder, sebagaimana digunakan dalam studi (Li, et al. 2020).



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Model dikembangkan dalam dua varian utama untuk keperluan komparasi, yaitu versi pertama dengan *Encoder BiLSTM* dan versi kedua dengan *Encoder BiGRU*, sementara struktur decoder dan tokenizer tetap sama. Evaluasi dilakukan menggunakan kombinasi metrik *Word Error Rate (WER)*, *CTC loss*, *Attention loss*, serta hasil decoding menggunakan metode *beam search*.

4.3.1 Ekstraksi Data

Tahap ekstraksi data merupakan proses penting dalam sistem *Continuous Sign Language Recognition (CSLR)*, karena di sinilah gesture dalam video dikonversi menjadi representasi numerik yang dapat dipahami oleh model pembelajaran mesin. Dalam penelitian ini, proses ekstraksi dilakukan menggunakan *MediaPipe Holistic*, sebuah *framework* deteksi landmark manusia yang dikembangkan oleh Google, yang mampu mendeteksi *keypoints* dari tiga bagian utama secara simultan, yaitu wajah (468 titik), pose tubuh (33 titik), serta tangan kiri dan kanan (masing-masing 21 titik). Seluruh *keypoints* ini digabungkan menjadi satu vektor berdimensi 1662 per *frame* video, yang kemudian disusun menjadi array berdimensi (jumlah *frame*, 1662) untuk setiap input sekuens.

Tabel 4. 8 Output Ekstraksi Mediapipe Holistic

Nama	Deskripsi Output Koordinat
Wajah	(x, y, z) * 468 Titik
Pose Tubuh	(x, y, z, visibility) * 33 Titik
Tangan Kanan	(x, y, z) * 21 Titik
Tangan Kiri	(x, y, z) * 21 Titik

Implementasi *MediaPipe Holistic* pada penelitian ini menggunakan konfigurasi parameter sebagai berikut: *static_image_mode=False* untuk memungkinkan pelacakan lintas *frame* secara berkelanjutan, *model_complexity=2* untuk memaksimalkan akurasi deteksi dengan arsitektur model yang lebih kompleks, serta *min_detection_confidence=0.7* dan *min_tracking_confidence=0.7* untuk menyaring prediksi dengan tingkat keyakinan rendah. Pengaturan ini bertujuan menghasilkan deteksi *keypoints* yang stabil dan akurat untuk kebutuhan pelatihan



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

model deep learning. Semua nilai parameter yang digunakan untuk ekstraksi bisa dilihat pada Tabel 4.9 Parameter Mediapipe Guna Kebutuhan Ekstraksi berikut.

Tabel 4. 9 Parameter Mediapipe Guna Kebutuhan Ekstraksi

Parameter	Nilai
static_image_mode	False
model_complexity	2 (Full Model)
min_detection_confidence	0.7
min_tracking_confidence	0.7

Data *keypoints* tiga dimensi yang dihasilkan oleh MediaPipe tidak langsung digunakan dalam bentuk aslinya (*frame*, titik, 3), melainkan diolah terlebih dahulu menjadi vektor satu dimensi (*frame*, 1662). Pendekatan ini dipilih untuk menyederhanakan struktur input agar sesuai dengan kebutuhan model berbasis Recurrent Neural Network (RNN), khususnya BiLSTM, yang mengharuskan input berupa vektor fitur per langkah waktu. Dengan menggunakan representasi 1D per *frame*, input menjadi lebih efisien untuk diproses dan dapat langsung dimasukkan ke dalam model tanpa memerlukan layer tambahan untuk memanipulasi struktur data spasial (Samaan, et al., 2022).

Skrip ekstraksi ini juga menyertakan augmentasi data untuk meningkatkan variasi dataset, mencakup flipping horizontal, penyesuaian tingkat kecerahan (bright-up dan bright-down), serta *frame* skipping. Variasi augmentasi ini menghasilkan beberapa versi dari satu video asli, sehingga dapat meningkatkan kemampuan generalisasi model terhadap variasi pencahayaan, arah gerakan, dan kecepatan isyarat.

Secara implementasi, proses ekstraksi dilakukan berdasarkan file metadata yang berisi nama video dan label kalimat. Video diproses *frame* demi *frame* menggunakan OpenCV, dikonversi ke format RGB, dan diekstraksi *keypoints*-nya melalui pipeline MediaPipe. Hasil ekstraksi kemudian dikonversi ke array NumPy dan disimpan dalam format .npy untuk digunakan sebagai input dalam proses pelatihan model. Validasi dilakukan untuk memastikan tidak ada data kosong yang disimpan, serta log akan menandai jika file video tidak ditemukan. Dengan



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

pendekatan ini, sistem ekstraksi tidak hanya menghasilkan representasi spasial-temporal gesture yang konsisten dan informatif, tetapi juga disusun dalam format efisien yang mendukung performa tinggi dalam pelatihan dan inferensi model berbasis deep learning.

Sebagai langkah untuk memastikan integritas dan kualitas data hasil ekstraksi, dilakukan validasi terhadap beberapa sampel dataset. Validasi ini mencakup empat aspek utama yang diperiksa secara numerik dan satu aspek tambahan yang diperiksa secara visual. Empat indikator numerik meliputi: jumlah *frame* per file, dimensi fitur setiap *frame*, keberadaan nilai kosong (NaN atau inf), serta kesesuaian ukuran array dengan metadata.

Tabel 4.10 Hasil Validasi Dataset Ekstraksi *Keypoints* menampilkan hasil validasi terhadap beberapa file hasil ekstraksi *keypoints* dari video yang telah diproses menggunakan MediaPipe Holistic. Validasi ini meliputi pengecekan jumlah *frame* pada setiap file, dimensi fitur yang dihasilkan, ada tidaknya nilai kosong (missing values), serta ukuran array hasil ekstraksi.

Pada tabel ini, setiap baris mewakili satu file hasil ekstraksi (Nama File). Kolom Jumlah *Frame* menunjukkan jumlah *frame* video yang berhasil diekstrak pada file tersebut. Kolom Dimensi Fitur menampilkan jumlah fitur pada setiap *frame*, yang konsisten sebanyak 1662 dimensi sesuai dengan jumlah *keypoints* yang diekstrak. Kolom Nilai Kosong menginformasikan ada tidaknya data kosong atau missing values pada array, dan pada ketiga file yang diuji tidak ditemukan nilai kosong (“Tidak Ada”). Kolom Ukuran Array menggambarkan dimensi keseluruhan array, dengan format (jumlah *frame*, jumlah fitur per *frame*).

Tabel 4. 10 Hasil Validasi Dataset Ekstraksi *Keypoints*

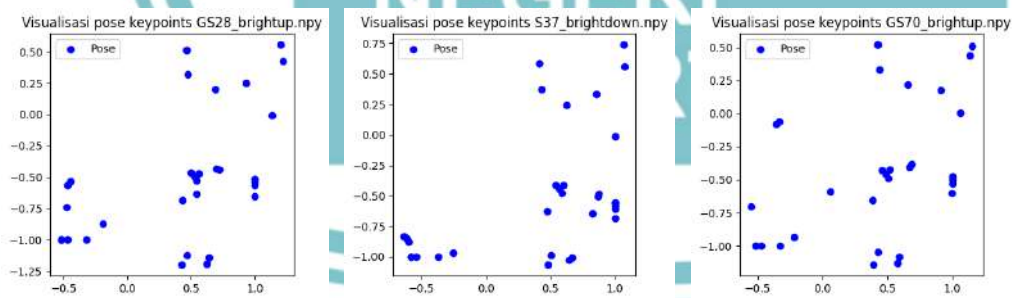
Nama File	Jumlah <i>Frame</i>	Dimensi Fitur	Nilai Kosong	Ukuran Array
GS28_brightup.npy	144	1662	Tidak Ada	(144, 1662)
S37_brightdown.npy	137	1662	Tidak Ada	(137, 1662)
GS70_brightup.npy	140	1662	Tidak Ada	(140, 1662)

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Sebagai tambahan untuk meningkatkan kualitas data dan mereduksi gangguan visual yang dapat memengaruhi hasil deteksi *keypoints*, dilakukan penerapan Kalman Filter pada setiap *frame* hasil ekstraksi. Kalman Filter digunakan sebagai metode noise filtering numerik yang efektif dalam melacak pergerakan objek dinamis berdasarkan observasi sebelumnya. Setiap vektor *keypoints* dari *frame* yang valid akan difilter agar memiliki transisi spasial yang lebih halus dan mengurangi jitter atau fluktuasi koordinat akibat kesalahan deteksi sesaat. Filter ini hanya diterapkan pada data yang valid (tidak nol), dan tidak dipaksakan jika seluruh titik tidak terdeteksi dalam satu *frame*.

Validasi terakhir dilakukan secara visual dengan menampilkan hasil deteksi *keypoints* pada satu *frame* dari setiap file. Visualisasi ini memplot titik-titik pose (33 titik utama tubuh) pada bidang 2D, dan hasil pengamatan menunjukkan bahwa semua landmark terdeteksi secara konsisten dan berada pada posisi anatomi yang sesuai dengan postur pengguna di dalam video. Hal ini menunjukkan bahwa MediaPipe Holistic berhasil mengekstraksi informasi gerakan tubuh secara konsisten dan sistematis. Dengan demikian, keseluruhan proses ekstraksi data dalam penelitian ini telah menghasilkan dataset yang layak untuk digunakan dalam pelatihan dan evaluasi model CSLR. Gambar 4.3 Visualisasi Plot Koordinat Pose menunjukkan konsistensi posisi koordinat dari tiga contoh *frame* dari 3 video yang diambil untuk sampel.



Gambar 4. 3 Visualisasi Plot Koordinat Pose

4.3.2 Data Preparation

Sebelum model dapat dilatih, kalimat target dari masing-masing video harus dikonversi terlebih dahulu menjadi format token numerik yang bisa diproses oleh decoder. Tokenisasi dilakukan menggunakan Tokenizer dari TensorFlow Text

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

dengan pendekatan berbasis word-level. Setiap kalimat dalam metadata diproses dengan penambahan token khusus <sos> di awal dan <eos> di akhir, yang digunakan untuk menandai awal dan akhir sekuens. Selain itu, token <pad> digunakan untuk memastikan seluruh sekuens memiliki panjang yang seragam, dan <oov> untuk menangani kata di luar kosakata (out-of-vocabulary). Gambar 4. 4 Program Tokenisasi Label menampilkan skrip *Tokenizing dan Padding* untuk seluruh kalimat di Metadata.

```
def load_metadata_and_tokenizer(metadata_path, feature_dir, tokenizer_path=None):
    df = pd.read_csv(metadata_path, delimiter=";")

    df = df[df['SENTENCE_NAME'].isin(
        [f.split('.')[0] for f in os.listdir(feature_dir) if f.endswith('.npy')]
    )]

    df['SENTENCE'] = df['SENTENCE'].astype(str)
    df['SENTENCE'] = df['SENTENCE'].apply(lambda x: '<sos> ' + x.strip() + ' <eos>')
    if tokenizer_path and os.path.exists(tokenizer_path):
        with open(tokenizer_path, 'rb') as f:
            tokenizer = pickle.load(f)
    else:
        tokenizer = Tokenizer(oov_token='<oov>', split=' ', filters='!"#$%&()*+,-./:;=?@[\\"')
        tokenizer.fit_on_texts(df['SENTENCE'].tolist())

        tokenizer.word_index['<pad>'] = 0
        tokenizer.index_word[0] = '<pad>'

        with open(os.path.join('/kaggle/working', "tokenizer.pkl"), 'wb') as f:
            pickle.dump(tokenizer, f)
            print("✅ Tokenizer disimpan ke tokenizer.pkl")

    vocab_size = len(tokenizer.word_index) + 1 # Termasuk semua token
    return df, tokenizer, vocab_size
```

Gambar 4. 4 Program Tokenisasi Label

Ukuran vocabulary bergantung pada jumlah kata unik di dalam transkrip pelatihan, ditambah token-token spesial tersebut. Contoh representasi token untuk kalimat “aku makan pisang” bisa dilihat di Tabel 4. 10 Contoh Representasi Token.

Tabel 4. 11 Contoh Representasi Token

Tokenized Transcript	Hasil Tokenisasi
<sos> aku makan pisang <eos> <pad>	[1, 12, 19, 21, 2, 0, 0]
<pad>	

Adapun format input ke model adalah berupa array berdimensi (T, 1662), di mana T merupakan jumlah *frame* hasil ekstraksi, dan 1662 adalah hasil konkatenasi dari seluruh koordinat (x, y, z) pada 33 *keypoints* pose, 21 *keypoints* tangan kiri, 21

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

tangan kanan, dan 468 wajah. Alih-alih menggunakan data mentah dalam bentuk 3 dimensi (x, y, z), seluruh *keypoints* digabung dan di-flatten menjadi bentuk 1 dimensi per *frame*. Pendekatan ini dilakukan agar model dapat mengolah input sebagai sequence of vectors yang bersifat seragam dan dapat langsung diteruskan ke RNN *Encoder*.

```
def create_tf_dataset(df, tokenizer, feature_dir, batch_size=8):
    paths = []
    labels_ctc = []
    labels_decoder = []

    for _, row in df.iterrows():
        sentence_name = row['SENTENCE_NAME']
        npy_path = os.path.join(feature_dir, sentence_name + ".npy")
        if os.path.exists(npy_path):
            sentence = row['SENTENCE']
            token_seq = tokenizer.texts_to_sequences([sentence])[0]
            # Hybrid: simpan dua versi label
            token_seq_ctc = filter_special_tokens(token_seq, tokenizer) # Untuk CTC (tanpa token spesial)
            token_seq_decoder = token_seq # Untuk decoder (dengan <sos>, <eos>)
            paths.append(npy_path.encode()) # TensorFlow wants byte strings
            labels_ctc.append(token_seq_ctc)
            labels_decoder.append(token_seq_decoder)

    def load_npy_and_pad(path, label_ctc, label_decoder):
        # path: tf.string (bytes), label: tf.Tensor([int]), label: tf.Tensor([int])
        def _load_and_pad(path_str, label_ctc_np, label_decoder_np):
            path_str = path_str.decode()
            x = np.load(path_str, mmap_mode='r').astype(np.float32)
            input_len = x.shape[0]
            ctc_len = len(label_ctc_np)
            decoder_len = len(label_decoder_np)
            return x, label_ctc_np.astype(np.int32), label_decoder_np.astype(np.int32), np.int32(input_len), np.int32(ctc_len), np.int32(decoder_len)
            x, y_ctc, y_decoder, x_len, y_ctc_len, y_decoder_len = tf.numpy_function(
                _load_and_pad,
                inp=[path, label_ctc, label_decoder],
                Tout=[tf.float32, tf.int32, tf.int32, tf.int32, tf.int32, tf.int32]
            )
            x.set_shape([None, 1662])
            y_ctc.set_shape([None])
            y_decoder.set_shape([None])
            return x, y_ctc, y_decoder, x_len, y_ctc_len, y_decoder_len

        ragged_labels_ctc = tf.ragged.constant(labels_ctc)
        ragged_labels_decoder = tf.ragged.constant(labels_decoder)
        dataset = tf.data.Dataset.from_tensor_slices((paths, ragged_labels_ctc, ragged_labels_decoder))
        dataset = dataset.map(load_npy_and_pad, num_parallel_calls=tf.data.AUTOTUNE)
        dataset = dataset.padded_batch(
            batch_size,
            padded_shapes=([None, 1662], [None], [None], [], [], []),
            drop_remainder=False
        )
        dataset = dataset.prefetch(tf.data.AUTOTUNE)
        return dataset
```

Gambar 4. 5 Program *Dataset Preparation*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Program pada Gambar 4.5 dirancang untuk mengonversi data sekuensial hasil ekstraksi *keypoints* (file .npy) menjadi dataset TensorFlow yang siap dikonsumsi oleh model selama proses training dan evaluasi. Proses diawali dengan iterasi pada baris-baris metadata untuk mencocokkan file fitur (.npy) dengan transkrip kalimat aslinya. Setiap kalimat transkrip diubah menjadi urutan token numerik menggunakan tokenizer, menghasilkan dua jenis label berbeda, yaitu label untuk Connectionist Temporal Classification (CTC) dan label untuk *Attention* Decoder. Label CTC disusun tanpa token khusus seperti <sos> atau <eos>, sedangkan label decoder mempertahankan token-token tersebut untuk mendukung arsitektur hybrid CTC-*Attention*. Pendekatan ini memungkinkan model untuk belajar dua tujuan secara simultan, yakni alignment temporal (CTC) dan pemodelan sekuens bahasa (*decoder*).

Setelah daftar path file fitur dan pasangan label disusun, dataset TensorFlow dibangun menggunakan `tf.data.Dataset` yang efisien. Fungsi `load_npy_and_pad` digunakan sebagai pemetaan untuk memuat data .npy dari disk secara dinamis saat training. Data yang dimuat mencakup array sekuens input dengan bentuk (T, 1662), di mana T adalah panjang sekuen (jumlah *frame*) dan 1662 adalah dimensi fitur (gabungan *keypoints* wajah, tangan, dan postur). Selain input sekuens, fungsi ini juga mengembalikan panjang input, panjang label CTC, dan panjang label decoder, yang seluruhnya diperlukan untuk menghitung loss saat pelatihan. Dataset kemudian dibungkus dalam `padded_batch` agar seluruh batch memiliki panjang sekuens dan label yang seragam secara dinamis, namun tetap efisien berkat masking otomatis pada padding. Langkah ini penting untuk mendukung pelatihan model berbasis sequence-to-sequence dengan arsitektur hybrid CTC-*Attention*.

Seluruh proses diparalelkan menggunakan AUTOTUNE dari TensorFlow untuk memaksimalkan efisiensi pada saat training dan meminimalkan latensi loading data. Hasil akhir dari skrip ini adalah dataset TensorFlow yang telah teroptimasi untuk streaming data secara efisien ke model, mendukung eksekusi GPU yang stabil, dan menjaga integritas label untuk kedua cabang pelatihan (CTC dan *Attention* Decoder) dalam arsitektur hybrid. Dengan desain seperti ini, pipeline data tidak hanya kuat

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

dalam skenario pelatihan standar, tetapi juga fleksibel untuk berbagai konfigurasi panjang sekuens input dan output.

4.3.3 Pelatihan Model dengan BiLSTM

Pada tahap ini dilakukan pelatihan model *Continuous Sign Language Recognition* (CSLR) dengan pendekatan arsitektur hybrid antara Connectionist Temporal Classification (CTC) dan *Attention*-based Sequence-to-Sequence. Model dilatih menggunakan input *keypoints* sekuens berdimensi (T, 1662), hasil dari ekstraksi dan penggabungan fitur wajah, tangan, dan pose tubuh yang diperoleh dari MediaPipe Holistic. Arsitektur utama model terdiri atas dua komponen: *Encoder* PyramidBiLSTM dan decoder LSTM dengan *Attention*.

```
class PyramidBiLSTM(tf.keras.Model):
    def __init__(self, output_vocab_size, hidden_size=HIDDEN_SIZE, pool_sizes=[N1, N2]):
        super().__init__()
        self.hidden_size = hidden_size
        self.pool_sizes = pool_sizes

        self.bilstm1 = tf.keras.layers.Bidirectional(
            tf.keras.layers.LSTM(256, return_sequences=True))
        self.bilstm2 = tf.keras.layers.Bidirectional(
            tf.keras.layers.LSTM(128, return_sequences=True, dropout=0.3, recurrent_dropout=
0.3))
        self.bilstm3 = tf.keras.layers.Bidirectional(
            tf.keras.layers.LSTM(128, return_sequences=True, dropout=0.3, recurrent_dropout=
0.3))
        self.proj1 = tf.keras.layers.Dense(HIDDEN_SIZE)
        self.proj2 = tf.keras.layers.Dense(HIDDEN_SIZE)
        self.classifier = tf.keras.layers.Dense(output_vocab_size + 1)
```

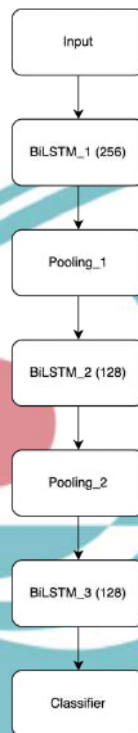
Gambar 4. 6 Program *Encoder* BiLSTM

Pada Gambar 4.6 Program *Encoder* BiLSTM, mengimplementasikan tiga lapis bidirectional LSTM, masing-masing terdiri atas 256 unit pada lapis pertama, dan dua lapis berikutnya masing-masing 128 unit. Untuk mengurangi dimensi temporal secara bertahap dan menstabilkan informasi temporal jangka panjang, diterapkan dua tahap sliding max pooling dengan ukuran *window* N1=6 dan N2=3. Setiap lapis BiLSTM juga dilengkapi dengan dropout sebesar 0.3 guna mencegah *overfitting*. Setelah melalui proses *encoding*, output fitur diklasifikasikan menggunakan layer dense untuk menghasilkan logits bagi cabang CTC dan juga dikirimkan ke *decoder* untuk pemodelan dependensi temporal lanjutan. Gambar 4.7 dan Gambar 4.8 juga menunjukkan arsitektur model *Encoder* yang digunakan dalam penelitian ini serta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

struktur detail dari model *Encoder* Pyramid BiLSTM yang digunakan, sekaligus memperlihatkan estimasi jumlah parameter pada masing-masing komponennya



Gambar 4. 7 Arsitektur Model *Encoder*

Model: "pyramid_bi_lstm_5"

Layer (type)	Output Shape	Param #
BiLSTM (Bidirectional)	(1, 30, 512)	3,930,112
BiLSTM Dropout 0.3 (Bidirectional)	(1, 9, 256)	263,168
BiLSTM Dropout 0.3 (Bidirectional)	(1, 7, 256)	263,168
dense_15 (Dense)	(1, 30, 128)	65,664
dense_16 (Dense)	(1, 9, 128)	32,896
Classifier (Dense)	(1, 7, 86)	22,102

Total params: 4,577,110 (17.46 MB)

Trainable params: 4,577,110 (17.46 MB)

Non-trainable params: 0 (0.00 B)

Gambar 4. 8 Rangkuman Model *Encoder* Pyramid BiLSTM

Pada program di Gambar 4.9, Decoder LSTM menggunakan embedding dengan ukuran 256, *Attention* scaled dot-product, serta satu lapisan LSTM dengan 128 unit. *Attention* digunakan untuk menyelaraskan antara urutan *Encoder* dan target secara dinamis, menghasilkan representasi kontekstual sebelum klasifikasi token. Output decoder berupa distribusi token dari vocabulary hasil tokenisasi teks kalimat

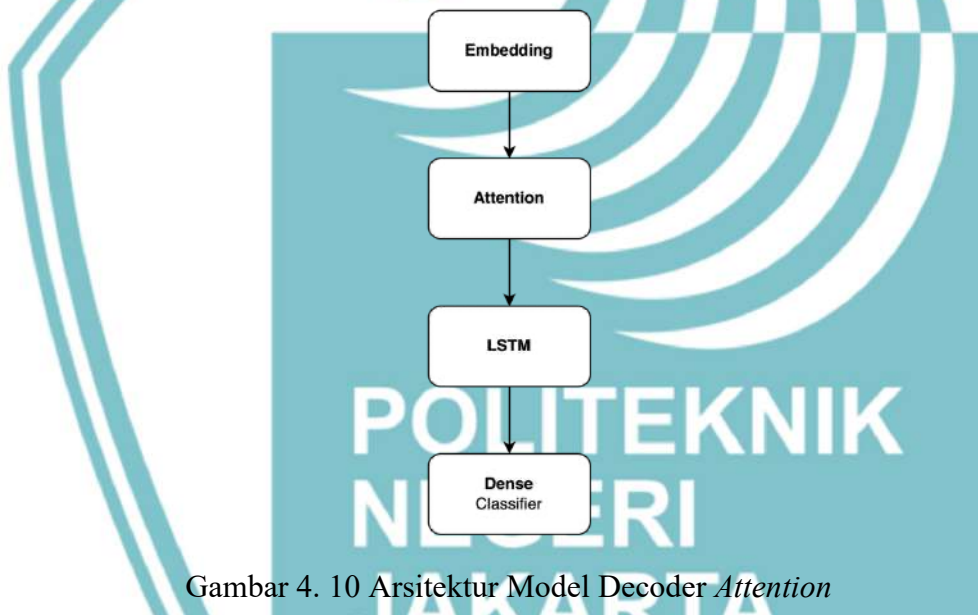
Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

menggunakan TensorFlow Text dengan token khusus seperti < sos>, < eos>, < pad>, dan < oov>. Target label yang digunakan selama pelatihan di-padding hingga panjang maksimum, dengan masking yang diterapkan untuk menghitung *Attention* loss.

```
class LSTMDecoder(tf.keras.Model):
    def __init__(self, vocab_size, embedding_dim=256, hidden_size=HIDDEN_SIZE):
        super().__init__()
        self.embedding = tf.keras.layers.Embedding(vocab_size, embedding_dim, mask_zero=True)
        self.encoder_projection = tf.keras.layers.Dense(embedding_dim) # Proyeksi encoder ke dimensi embedding
        self.lstm = tf.keras.layers.LSTM(hidden_size, return_sequences=True)
        self.attention = tf.keras.layers.Attention(use_scale=True)
        self.classifier = tf.keras.layers.Dense(vocab_size)
```

Gambar 4. 9 Program Decoder *Attention*



Gambar 4. 10 Arsitektur Model Decoder *Attention*

Model: "lstm_decoder_3"

Layer (type)	Output Shape	Param #
embedding_3 (Embedding)	(1, 1, 256)	21,760
dense_17 (Dense)	(1, 7, 256)	65,792
lstm_16 (LSTM)	(1, 1, 128)	328,192
attention_3 (Attention)	(1, 1, 256)	1
dense_18 (Dense)	(1, 1, 85)	10,965

Total params: 426,710 (1.63 MB)

Trainable params: 426,710 (1.63 MB)

Non-trainable params: 0 (0.00 B)

Gambar 4. 11 Ringkasan Arsitektur Model Decoder *Attention*



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Model dilatih dengan skema gabungan loss, yaitu CTC loss untuk output logits dari *Encoder* dan cross-entropy loss dari decoder dengan *Attention*. Nilai λ sebesar 0.7 digunakan untuk menyeimbangkan kontribusi masing-masing loss terhadap total loss. Optimizer yang digunakan adalah AdamW dengan learning rate sebesar $5e-4$ dan weight decay $1e-4$. Model dilatih selama 100 epoch dengan ukuran batch 16. Selama pelatihan, panjang sekuens logit dihitung secara dinamis berdasarkan ukuran input dan efek dua tahap pooling.

Tabel 4. 12 Hyperparameter Pelatihan Model BiLSTM

Nama Hyperparameter	Nilai	Deskripsi
Input Dimensi	1662	Jumlah fitur per <i>frame</i> hasil ekstraksi <i>keypoints</i> (wajah, tangan, pose)
Hidden Size	128	Ukuran hidden state untuk decoder LSTM dan proyeksi <i>Encoder</i>
Pool Size 1 (N1)	6	Ukuran jendela max pooling pertama pada <i>Encoder</i>
Pool Size 2 (N2)	3	Ukuran jendela max pooling kedua pada <i>Encoder</i>
Dropout	0.3	Tingkat dropout pada layer BiLSTM untuk mencegah overfitting
Learning Rate	5×10^{-4}	Kecepatan pembelajaran awal untuk optimizer AdamW
Weight Decay	1×10^{-4}	Nilai regularisasi L2 untuk mencegah bobot membesar selama pelatihan
Epochs	100	Jumlah putaran penuh melalui seluruh dataset pelatihan
Batch Size	16	Jumlah sampel data dalam satu batch update bobot
Lambda CTC (λ)	0.7	Bobot pembagi antara loss CTC dan loss <i>Attention</i>



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Nama Hyperparameter	Nilai	Deskripsi
Beam Width	5	Jumlah jalur yang dievaluasi dalam decoding beam search

Untuk keperluan inferensi, model menggunakan beam search decoding berbasis *Attention* untuk memprediksi urutan token yang paling memungkinkan, dengan beam width sebanyak 5. Hal ini dilakukan karena decoder sequence-to-sequence dengan *Attention* cenderung menghasilkan hasil lebih akurat dibandingkan decoding greedy, terutama dalam pengurutan kalimat panjang yang kompleks.

Evaluasi model dilakukan menggunakan metrik Word Error Rate (WER), CTC Loss, *Attention* Loss, serta validasi terhadap keduanya. Penggunaan pendekatan hybrid ini memanfaatkan keunggulan CTC dalam alignment yang fleksibel serta kekuatan *Attention* dalam pemodelan konteks antar token, sehingga meningkatkan performa transkripsi isyarat secara menyeluruh. Arsitektur ini juga disiapkan untuk dibandingkan dengan model serupa yang hanya berbeda di bagian *Encoder*, yaitu model BiGRU, yang akan dibahas pada subjudul berikutnya.

4.3.4 Pelatihan Model dengan BiGRU

Sebagai pendekatan alternatif dari arsitektur berbasis BiLSTM, penelitian ini juga mengembangkan model *Continuous Sign Language Recognition* (CSLR) dengan menggunakan Bidirectional Gated Recurrent Unit (BiGRU). Penggunaan BiGRU didasarkan pada karakteristiknya yang lebih ringan secara komputasi dibandingkan LSTM, namun tetap mampu mempertahankan performa kompetitif dalam pengolahan data sekuensial. GRU memiliki struktur yang lebih sederhana karena menggabungkan forget gate dan input gate menjadi satu komponen yang disebut update gate, sehingga menghasilkan lebih sedikit parameter dan proses komputasi yang lebih cepat. Pendekatan ini dinilai potensial dalam konteks pelatihan skala besar dan inferensi real-time, terutama ketika sumber daya perangkat keras terbatas. Struktur *Encoder* pada model BiGRU secara keseluruhan identik dengan model sebelumnya yang berbasis BiLSTM, dengan tiga lapisan utama. Lapisan pertama terdiri dari Bidirectional GRU dengan 256 unit, sedangkan lapisan kedua dan ketiga



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

masing-masing terdiri dari Bidirectional GRU dengan 128 unit. Setiap lapisan dilengkapi dengan mekanisme regularisasi berupa dropout sebesar 0.3 untuk mencegah overfitting, serta dua tahap sliding max pooling setelah lapisan pertama dan kedua, masing-masing menggunakan jendela sebesar $N1 = 6$ dan $N2 = 3$ dengan 50% overlap. Pooling ini berfungsi untuk mereduksi panjang sekuens input secara bertahap sambil mempertahankan informasi spasial yang relevan.

Arsitektur decoder yang digunakan tetap mempertahankan struktur sebelumnya seperti pada Gambar 4.6, yaitu menggunakan embedding layer dengan dimensi 256 yang kemudian diproses oleh *Attention* layer berbasis scaled dot-product. Output dari *Attention* digabungkan dengan embedding dan diproses lebih lanjut oleh layer LSTM berisi 128 unit. Hasil akhir dari LSTM diteruskan ke dense classifier yang bertugas memetakan ke token dalam vocabulary. *Decoder* ini dirancang untuk melakukan prediksi sekuens token secara *autoregressive* dengan mempertimbangkan konteks temporal dan semantik dari sekuens masukan.

Pelatihan model dilakukan dengan konfigurasi hyperparameter yang sama dengan model BiLSTM pada Tabel 4. 11 Hyperparameter Pelatihan Model BiLSTM agar evaluasi performa kedua model dapat dibandingkan secara adil. Optimizer yang digunakan adalah AdamW dengan learning rate sebesar 5×10^{-4} dan weight decay sebesar 1×10^{-4} . Proses pelatihan dilakukan selama 100 epoch dengan ukuran batch sebanyak 16. Fungsi loss yang digunakan merupakan kombinasi dari CTC loss dan cross-entropy loss dari decoder, dengan skema pembobotan hybrid sebesar $\lambda = 0.7$ untuk CTC dan $(1 - \lambda) = 0.3$ untuk decoder loss. Model dievaluasi tidak hanya berdasarkan loss, tetapi juga menggunakan skema decoding berbasis beam search dengan beam width sebanyak 5 untuk meningkatkan akurasi prediksi.

4.4 Evaluasi Model

Evaluasi dilakukan pada dataset uji yang berbeda jumlahnya sesuai skenario. Model dengan augmentasi diuji pada 254 sampel uji, sedangkan model tanpa augmentasi diuji pada 43 sampel uji (hasil split dataset 70:30). Semua model menggunakan pendekatan hybrid CTC-*Attention*, dan decoding dilakukan dengan beam search (beam width = 5) agar prediksi kalimat mempertimbangkan beberapa kandidat jalur



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

token pada setiap langkah waktu, bukan hanya jalur terbaik tunggal seperti greedy decoding.

Metrik utama yang digunakan dalam evaluasi ini adalah Word Error Rate (WER), CTC Loss, dan *Attention* Loss, serta nilai validasinya (Val CTC Loss dan Val *Attention* Loss). Metrik WER digunakan untuk mengukur sejauh mana kalimat hasil prediksi mendekati kalimat referensi, dengan menghitung jumlah kesalahan berupa substitusi, insersi, dan delesi. Sementara itu, CTC Loss menilai kemampuan model dalam memetakan urutan *frame* gesture menjadi urutan token tanpa memerlukan segmentasi eksplisit. *Attention* Loss mengevaluasi efektivitas decoder dalam memahami konteks kalimat secara linguistik. Nilai validasi untuk kedua loss ini digunakan sebagai indikator generalisasi model pada data uji.

Implementasi fungsi WER dengan Beam Search menggunakan program pada gambar 4.12.



POLITEKNIK
NEGERI
JAKARTA

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

```
def predict_attention(self, x_batch, tokenizer, beam_width=5, max_len=50):
    sos_token = tokenizer.word_index['<sos>']
    eos_token = tokenizer.word_index['<eos>']

    batch_size = tf.shape(x_batch)[0]
    encoder_features = self.encoder(x_batch, return_features=True)
    final_predictions = []

    for i in range(batch_size):
        single_encoder_feature = tf.expand_dims(encoder_features[i], 0)
        beams = [(0.0, [sos_token])]

        for _ in range(max_len):
            new_beams = []
            all_done = True
            for score, seq in beams:
                if seq[-1] == eos_token:
                    new_beams.append((score, seq))
                    continue

            all_done = False
            target_seq_input = tf.constant([seq], dtype=tf.int32)
            decoder_output = self.decoder(single_encoder_feature, target_seq_input)
            next_word_logits = decoder_output[:, -1, :]
            next_word_log_probs = tf.nn.log_softmax(next_word_logits)

            top_k_probs, top_k_indices = tf.math.top_k(next_word_log_probs, k=beam_width)

            for k in range(beam_width):
                new_seq = seq + [top_k_indices[0][k].numpy()]
                new_score = score + top_k_probs[0][k].numpy()
                new_beams.append((new_score, new_seq))

            if all_done: break
            beams = sorted(new_beams, key=lambda x: x[0], reverse=True)[:beam_width]

        # Normalisasi skor dengan panjang (opsional, tapi bisa membantu)
        best_score, best_seq = max(beams, key=lambda x: x[0] / len(x[1]) if len(x[1]) > 1 else -1e9)
        final_seq = [t for t in best_seq if t not in (sos_token, eos_token)]
        final_predictions.append(final_seq)

    return final_predictions
```

Gambar 4. 12 Program Implementasi *Word Error Rate*

Fungsi `predict_Attention` menerima masukan berupa batch data `x_batch`, objek `tokenizer`, nilai `beam_width`, dan panjang maksimum prediksi `max_len`. Langkah pertama dalam fungsi ini adalah mengambil indeks token khusus `<sos>` (start of sequence) dan `<eos>` (end of sequence) dari `tokenizer` untuk digunakan selama proses decoding. Seluruh batch input diproses satu per satu; untuk setiap contoh dalam batch, fitur laten dari *Encoder* diekstrak dengan opsi `return_features=True`, kemudian dimensi batch disesuaikan dengan menggunakan `tf.expand_dims` sehingga bentuk input kompatibel untuk *Decoder*.

Pada setiap langkah decoding, algoritma mempertahankan daftar beam, di mana setiap elemen terdiri atas pasangan skor kumulatif log-probabilitas dan urutan token yang dihasilkan hingga langkah tersebut. Inisialisasi dilakukan dengan beam yang



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

hanya berisi token <sos> dan skor awal nol. Untuk setiap langkah hingga max_len, fungsi mengevaluasi kandidat beam saat ini. Bila kandidat sudah berakhir dengan <eos>, kandidat ini diteruskan tanpa perubahan. Jika tidak, fungsi membentuk input sekuens untuk *Decoder* menggunakan tf.constant, kemudian menghitung keluaran *Decoder* pada langkah terakhir. Nilai logits dari *Decoder* diubah ke skala log-probabilitas menggunakan tf.nn.log_softmax.

Selanjutnya, fungsi mengambil beam_width token dengan log-probabilitas tertinggi dari distribusi keluaran menggunakan tf.math.top_k. Untuk setiap token hasil top_k, fungsi membentuk urutan baru dengan menambahkan token tersebut ke urutan sebelumnya dan memperbarui skor kumulatifnya. Daftar beam diperbarui dengan kandidat baru yang telah disortir berdasarkan skor kumulatif secara menurun, lalu dipangkas untuk mempertahankan hanya beam_width kandidat terbaik. Proses ini berlanjut hingga seluruh beam berakhir dengan token <eos> atau langkah mencapai max_len. Setelah proses decoding selesai, fungsi memilih urutan token dengan skor terbaik. Untuk meminimalisasi bias terhadap urutan pendek, skor dinormalisasi dengan membagi skor kumulatif dengan panjang urutan. Fungsi kemudian membuang token khusus <sos> dan <eos> dari urutan hasil sebelum mengembalikannya sebagai prediksi final.

4.4.1 Evaluasi Model BiLSTM

Pada skenario dengan augmentasi data, BiLSTM menunjukkan performa yang cukup baik. Model ini mampu menghasilkan WER sebesar **0.3813**, yang berarti sekitar **38%** kata dalam kalimat hasil prediksi berbeda dari kalimat referensi. CTC Loss pada model ini tercatat sebesar **0.5062**, sedangkan *Attention Loss* mencapai **0.0462**. Nilai validasi yang diperoleh juga cukup stabil, dengan Val CTC Loss sebesar **3.2355** dan Val *Attention Loss* sebesar **0.1801**. Hasil ini menunjukkan bahwa augmentasi data berhasil membantu model dalam mengenali variasi gestur dan mengurangi risiko overfitting. Augmentasi sederhana seperti horizontal flipping, brightness adjustment, dan *frame* skipping terbukti mampu mensimulasikan berbagai kondisi nyata, sehingga model lebih siap menghadapi data uji yang berbeda distribusinya dari data latih.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Namun, performa BiLSTM menurun drastis ketika augmentasi data tidak digunakan. Tanpa augmentasi, model mencatat WER sebesar **0.9765**, yang hampir menyamai performa acak, dengan CTC Loss sebesar **16.2762**, *Attention* Loss sebesar **0.8396**, serta nilai validasi yang sangat tinggi (Val CTC Loss sebesar **17.8062**, Val *Attention* Loss sebesar **2.0969**). Nilai-nilai ini menunjukkan bahwa BiLSTM sangat bergantung pada keberagaman data latih untuk menghasilkan prediksi yang akurat. Model gagal mengeneralisasi pada data uji yang memiliki pola sedikit berbeda dari data latih, sehingga prediksi kalimat menjadi tidak sesuai dengan kalimat referensi. Hasil akhir evaluasi untuk setiap *metrics* hasil training kedua model dipaparkan dalam Tabel 4.13 Hasil Evaluasi Model Berdasarkan Arsitektur dan Augmentasi BiLSTM. Performa model arsitektur ini juga ditampilkan dalam bentuk grafik untuk setiap iterasi pada Gambar 4.15 Perbandingan Grafik CTC Loss, *Attention* Loss, Val CTC Loss dan Val *Attention* Loss BiLSTM per Epoch.

Tabel 4. 13 Hasil Evaluasi Model BiLSTM

Model	WER	CTC Loss	Val CTC Loss	<i>Attention</i> Loss	Val <i>Attention</i> Loss
BiLSTM + Augmentasi	0.3813	0.5062	3.2355	0.0462	0.1801
BiLSTM tanpa Augmentasi	0.9765	16.2762	17.8062	0.8396	2.0969

Epoch 100 - Evaluating WER with Attention Beam Search (beam_width=5):
Eval WER (Beam Search): 0.3813
37/37 — 177s 5s/step - ctc_loss: 2.1580 - loss: 1.5458 - lstm_loss: 0.1
174 - val_ctc_loss: 3.7805 - val_loss: 2.7206 - val_lstm_loss: 0.2474 - wer: 0.3813

Gambar 4. 13 Output Hasil Training Model BiLSTM Dengan Augmentasi

Hak Cipta :

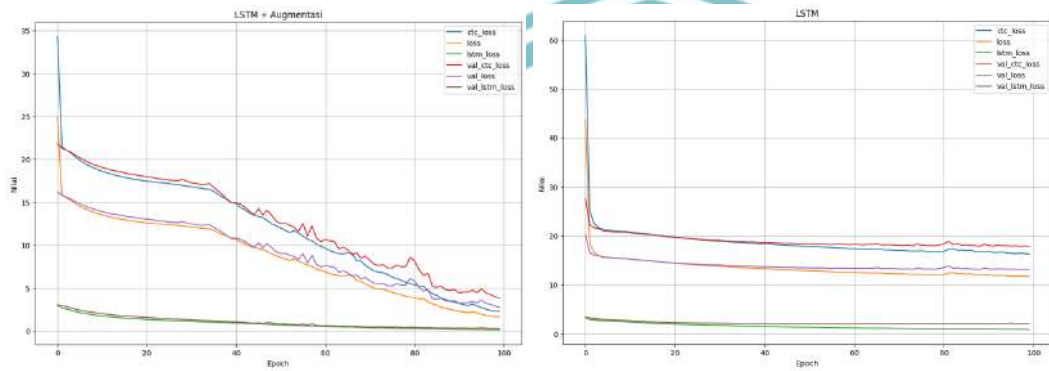
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Epoch 100 - Evaluating WER with Attention Beam Search (beam_width=5):

Eval WER (Beam Search): 0.9765

11/11 34s 3s/step - ctc_loss: 16.1891 - loss: 11.5909 - lstm_loss: 0.8618 - val_ctc_loss: 17.8062 - val_loss: 13.0934 - val_lstm_loss: 2.0969 - wer: 0.9765

Gambar 4. 14 Output Hasil Training Model BiLSTM Tanpa Augmentasi



Gambar 4. 15 Perbandingan Metrik BiLSTM per Epoch

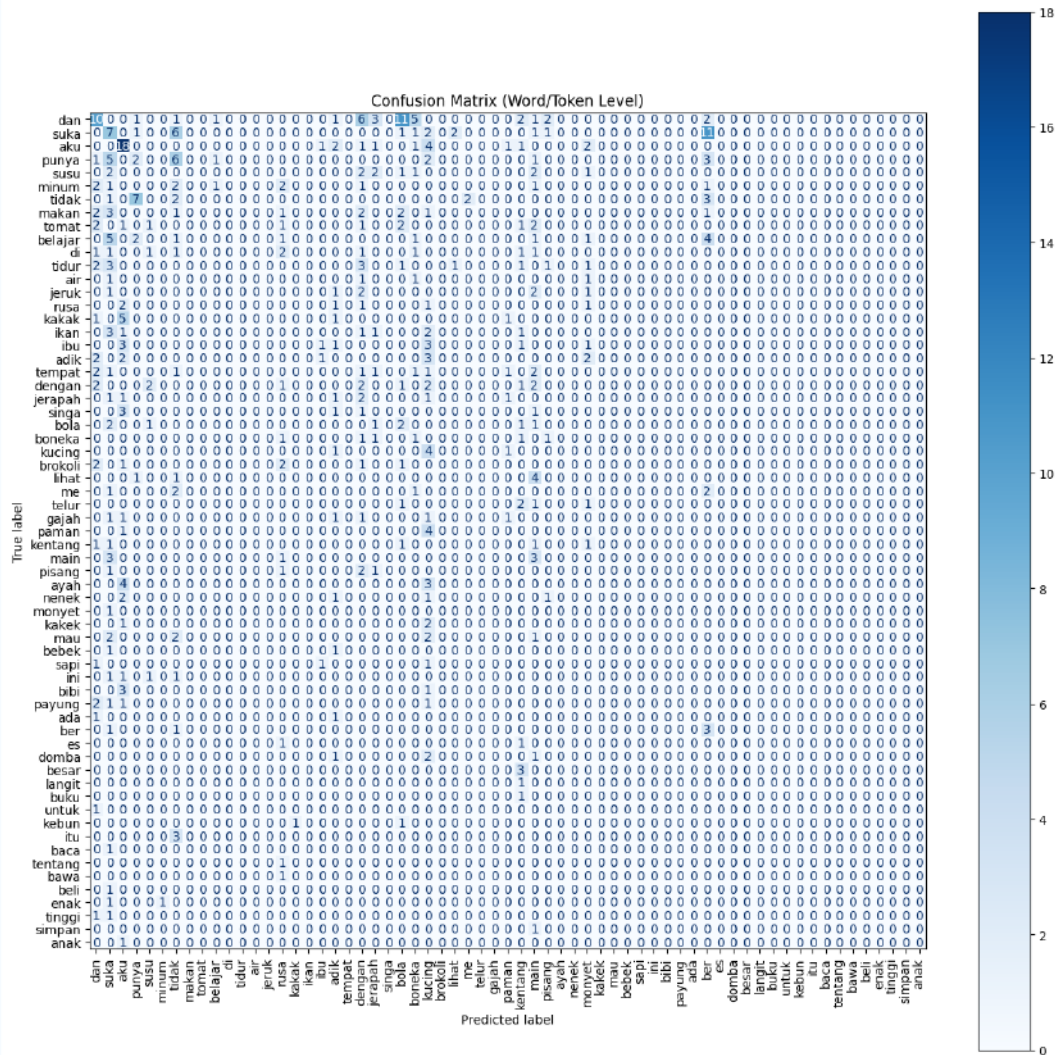
Pada Gambar 4.16, menunjukkan confusion matrix pada level token (kata) sebagai hasil evaluasi model CSLR berbasis LSTM tanpa augmentasi, yang diuji menggunakan data uji (unseen data). Matriks ini memvisualisasikan bagaimana prediksi kata oleh model dibandingkan dengan label sebenarnya untuk setiap token hasil pengenalan gesture bahasa isyarat.

Pada *confusion matrix* ini, sumbu vertikal menunjukkan label kata yang sebenarnya (true labels), sedangkan sumbu horizontal menunjukkan kata yang diprediksi oleh model (predicted labels). Nilai diagonal merepresentasikan prediksi yang benar, sementara nilai di luar diagonal mencerminkan kesalahan klasifikasi antar kata. Dari visualisasi tersebut, terlihat bahwa model LSTM tanpa augmentasi mengalami kesulitan dalam mengenali banyak kata dengan tepat. Hal ini ditunjukkan oleh banyaknya nilai non-nol di luar diagonal, yang menandakan tingginya tingkat kesalahan prediksi (*false positives* dan *false negatives*). Sebaliknya, nilai pada diagonal utama terlihat lemah dan tersebar, menunjukkan minimnya prediksi yang benar secara konsisten. Beberapa token seperti "dan", "suka", dan "aku" memang menunjukkan frekuensi prediksi yang cukup tinggi dan akurat, tetapi secara

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

keseluruhan, model tampak tidak stabil dalam menggeneralisasi gesture pada data uji.



Gambar 4. 16 *Confusion Matrix* LSTM Non Augmentasi

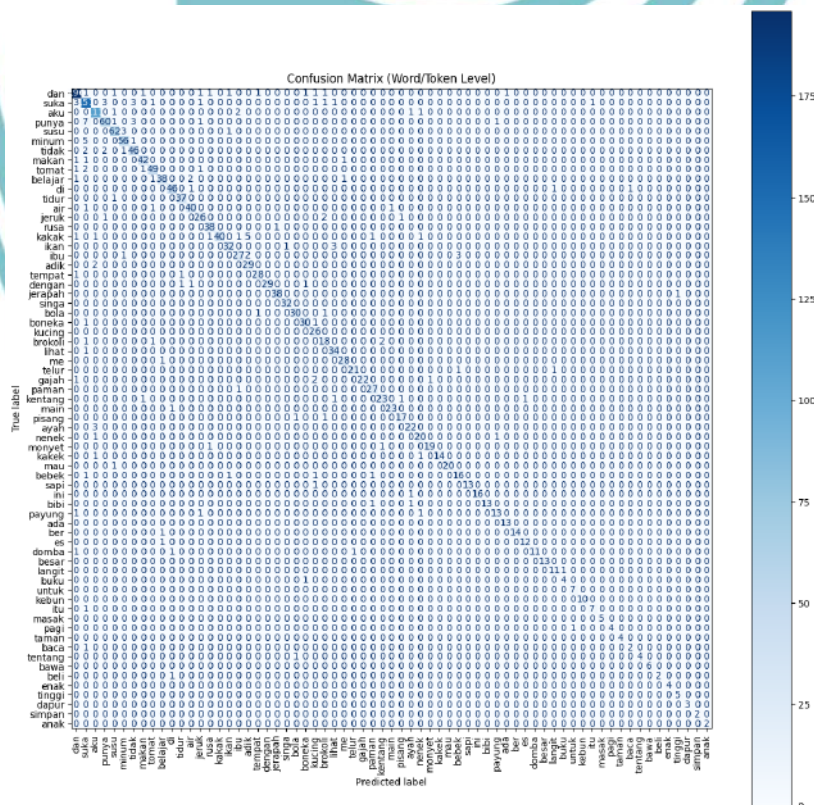
Sebagai kelanjutan dari analisis sebelumnya, Gambar 4.17 merupakan *Confusion Matrix* untuk model LSTM yang dilatih menggunakan dataset dengan augmentasi, dan dievaluasi pada unseen test data yang belum pernah digunakan selama pelatihan. Jika dibandingkan dengan model LSTM tanpa augmentasi, terlihat adanya peningkatan signifikan dalam ketepatan prediksi model, yang tercermin dari peningkatan jumlah prediksi benar pada diagonal utama.

Beberapa token seperti “dan”, “aku”, “punya”, “susu”, dan “tidak” menunjukkan prediksi yang jauh lebih stabil dan presisi. Bahkan token-token dengan kemunculan

Hak Cipta:

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

menengah seperti “kakak”, “adik”, atau “jeruk” juga memiliki jumlah prediksi benar yang konsisten lebih tinggi. Hal ini mengindikasikan bahwa augmentasi berhasil meningkatkan kemampuan generalisasi model terhadap variasi data, terutama dalam konteks gesture bahasa isyarat yang memiliki banyak nuansa gerakan. Berdasarkan evaluasi kuantitatif, model LSTM dengan augmentasi mencatatkan Word Error Rate (WER) sebesar 38,13%, yang jauh lebih baik dibandingkan dengan model serupa tanpa augmentasi yang memiliki WER mendekati 98%. Selain itu, nilai Attention Loss dan CTC Loss selama pelatihan juga menunjukkan penurunan yang stabil, memperkuat bukti bahwa data augmentasi berperan penting dalam stabilitas pelatihan dan performa inferensi. Dengan demikian, visualisasi Confusion Matrix ini mendukung kesimpulan bahwa augmentasi tidak hanya memperluas keragaman data pelatihan, tetapi juga memungkinkan model LSTM menangkap pola temporal dan spasial gesture dengan lebih baik.



Gambar 4. 17 *Confussion Matrix* LSTM dengan Augmentasi



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

4.4.2 Evaluasi Model BiGRU

Model BiGRU, yang memiliki arsitektur lebih ringkas dibanding BiLSTM, dirancang untuk memproses urutan gestur dengan efisien tanpa mengorbankan kemampuan dalam menangkap pola temporal. Evaluasi BiGRU dilakukan untuk mengetahui apakah arsitektur yang lebih ringan ini mampu bersaing atau bahkan mengungguli BiLSTM dalam tugas CSLR.

Hasil evaluasi menunjukkan bahwa BiGRU dengan augmentasi data berhasil mencapai performa terbaik di antara seluruh konfigurasi model. Model ini mencatat WER sebesar **0.2620**, yang menandakan kalimat hasil prediksi cukup mendekati kalimat referensi. CTC Loss tercatat sebesar **0.4674**, dan *Attention* Loss sebesar **0.0300**. Nilai validasi model juga relatif rendah, dengan Val CTC Loss sebesar **2.5055** dan Val *Attention* Loss sebesar **0.1619**, menunjukkan kemampuan generalisasi yang baik. Pencapaian ini menunjukkan bahwa arsitektur BiGRU, meskipun lebih sederhana, mampu memanfaatkan keberagaman data hasil augmentasi secara efektif dalam mempelajari pola gesture.

Pada skenario tanpa augmentasi, BiGRU mengalami penurunan performa, meskipun tidak sebesar penurunan yang terjadi pada BiLSTM. Model mencatat WER sebesar **0.7503**, CTC Loss sebesar **8.2045**, *Attention* Loss sebesar **0.6813**, Val CTC Loss sebesar **12.7841**, dan Val *Attention* Loss sebesar **2.0209**. Meskipun nilai-nilai ini cukup tinggi, BiGRU tanpa augmentasi tetap lebih baik dibandingkan BiLSTM tanpa augmentasi. Hal ini memperkuat dugaan bahwa BiGRU lebih stabil terhadap variasi pola gesture, meskipun pada akhirnya augmentasi tetap diperlukan untuk mencapai performa yang optimal.

Hasil akhir evaluasi untuk setiap *metrics* hasil training kedua model dipaparkan dalam Tabel 4.14 Hasil Evaluasi Model Berdasarkan Arsitektur dan Augmentasi BiGRU. Performa model arsitektur ini juga ditampilkan dalam bentuk grafik untuk setiap iterasi pada Gambar 4. 20.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Tabel 4. 14 Hasil Evaluasi Model BiGRU

Model	WER	CTC Loss	Val CTC Loss	Attention Loss	Val Attention Loss
BiGRU + Augmentasi	0.2620	0.4674	2.5055	0.0300	0.1619
BiGRU tanpa Augmentasi	0.7503	8.2045	12.7841	0.6813	2.0209

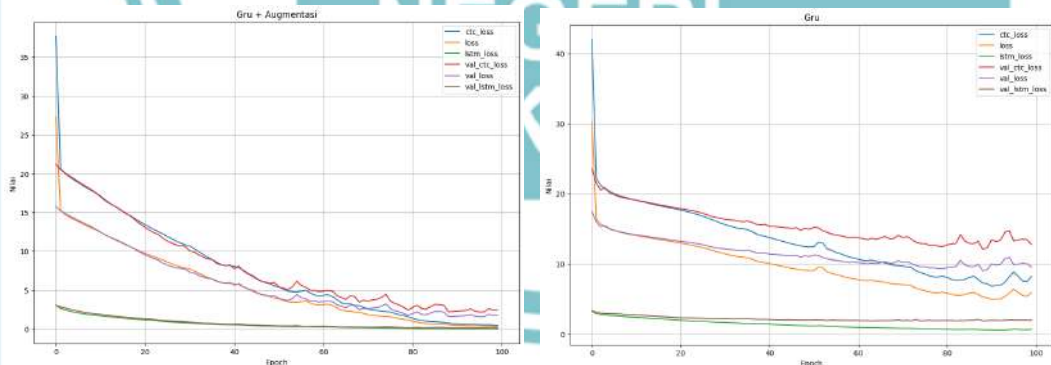
Epoch 100 - Evaluating WER with Attention Beam Search (beam_width=5):
 Eval WER (Beam Search): 0.2620
 37/37 ————— 208s 6s/step - ctc_loss: 0.4507 - loss: 0.3246 - lstm_loss: 0.0
 303 - val_ctc_loss: 2.5055 - val_loss: 1.8025 - val_lstm_loss: 0.1619 - wer: 0.2620

Gambar 4. 18 Output Hasil Training BiGRU dengan Augmentasi

Epoch 100: loss did not improve from 4.95922

Epoch 100 - Evaluating WER with Attention Beam Search (beam_width=5):
 Eval WER (Beam Search): 0.7503
 11/11 ————— 38s 4s/step - ctc_loss: 8.5266 - loss: 6.1833 - lstm_loss: 0.71
 56 - val_ctc_loss: 12.7841 - val_loss: 9.5551 - val_lstm_loss: 2.0209 - wer: 0.7503

Gambar 4. 19 Output Hasil Training BiGRU Tanpa Augmentasi



Gambar 4. 20 Perbandingan Grafik Metriks BiGRU per Epoch

Gambar 4.21 menampilkan *confusion matrix* pada level token (kata) sebagai hasil evaluasi performa model CSLR berbasis BiGRU tanpa augmentasi, yang diuji menggunakan data uji (unseen data). Matriks ini memberikan gambaran mengenai ketepatan model dalam memprediksi setiap kata dari gestur bahasa isyarat yang



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

ditransformasikan ke dalam bentuk urutan token. Terlihat bahwa prediksi yang benar (true positive) terkonsentrasi pada diagonal utama, yang menunjukkan bahwa sebagian kata berhasil diprediksi dengan tepat, seperti kata "aku", "punya", dan "tidur" yang memiliki jumlah prediksi benar lebih tinggi dibanding kata lainnya. Namun, dibandingkan dengan model yang dilatih dengan augmentasi, distribusi akurasi pada diagonal ini terlihat lebih menyebar dan lemah, yang menunjukkan bahwa akurasi model BiGRU tanpa augmentasi relatif rendah.

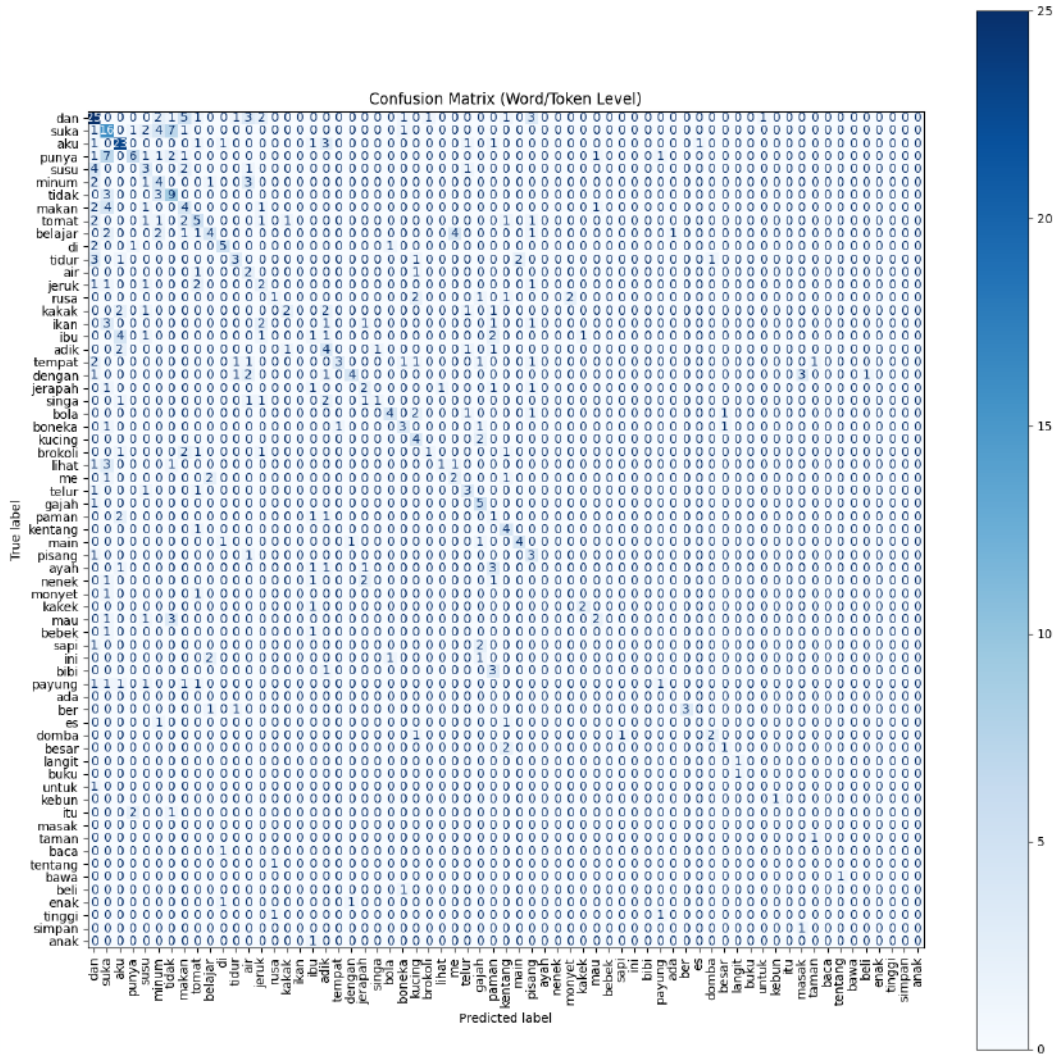
Di luar diagonal utama, terlihat banyak nilai non-nol yang tersebar pada baris dan kolom berbeda, mengindikasikan terjadinya kesalahan klasifikasi antar kata. Hal ini menandakan bahwa model kesulitan membedakan beberapa *gesture* yang memiliki kemiripan secara visual, terutama karena tidak cukupnya variasi data selama pelatihan.



POLITEKNIK
NEGERI
JAKARTA

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4. 21 Confussion Matrix GRU Tanpa Augmentasi

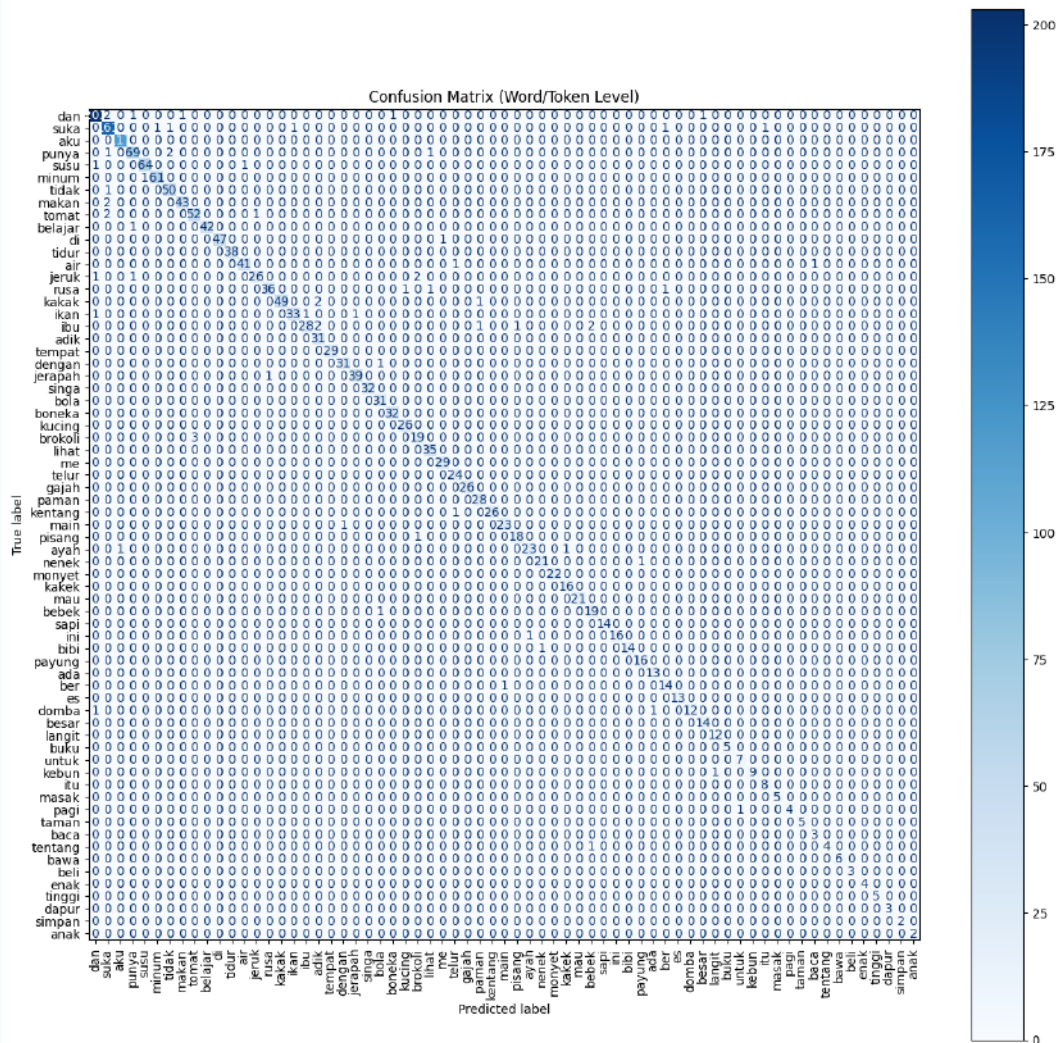
Gambar 4.22 adalah evaluasi *Confussion Matrix* untuk model BiGRU dengan augmentasi, yakni model yang terbukti memberikan performa terbaik dalam penelitian ini. Titik-titik bernilai tinggi pada diagonal utama menunjukkan bahwa model berhasil mengklasifikasikan banyak kata dengan benar. Beberapa kata seperti "aku", "punya", "susu", "tidak", dan "teman" memiliki akurasi pengenalan tinggi, terlihat dari konsistensi nilai diagonal yang besar.

Sebaliknya, sel-sel di luar diagonal menunjukkan kesalahan prediksi. Kesalahan ini relatif jarang dan tersebar tipis, menunjukkan bahwa model jarang melakukan prediksi yang meleset secara signifikan. Kesalahan yang muncul umumnya terjadi pada kata-kata yang secara visual memiliki gesture mirip, misalnya "dan" vs "suka" atau "punya" vs "tidak". *Confusion matrix* ini memperkuat hasil evaluasi metrik Jurusan Teknik Informatika dan Komputer – Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

lainnya, yaitu *Word Error Rate* (WER) sebesar **0.2620** serta CTC Loss dan Attention Loss yang rendah.



Gambar 4. 22 Confussion Matrix GRU dengan Augmentasi

4.4.3 Evaluasi Pakar

Sebagai pelengkap dari evaluasi kuantitatif berbasis metrik seperti *Word Error Rate* (WER), CTC Loss, dan Attention Loss, dilakukan pula evaluasi kualitatif terhadap sistem Continuous Sign Language Recognition (CSLR) yang dikembangkan. Evaluasi ini melibatkan seorang pakar pengguna bahasa isyarat dari SLBN 02 Lenteng Agung yang berpengalaman dalam pengajaran Sistem Isyarat Bahasa Indonesia (SIBI). Metode evaluasi dilakukan melalui survei kuesioner dengan skala



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Likert 1–5, yang mencakup berbagai aspek seperti akurasi terjemahan, kelancaran tata bahasa, kemudahan pemahaman, dan stabilitas sistem dalam kondisi real-time.

Hasil evaluasi menunjukkan bahwa sistem mendapatkan respons yang sangat positif pada aspek kelayakan dan kebermanfaatannya dalam lingkungan pendidikan inklusif, serta kemudahan penggunaan antarmuka oleh pengguna non-teknis. Beberapa indikator memperoleh skor maksimal, seperti pernyataan “Sistem ini layak untuk dikembangkan lebih lanjut dalam konteks pembelajaran bahasa isyarat” dan “Sistem tetap stabil saat input dilakukan secara real-time”, yang masing-masing mendapatkan skor 5. Sistem juga dinilai mampu menghasilkan prediksi yang konsisten dan bebas dari kata-kata yang tidak bermakna, yang tercermin dari skor sempurna pada aspek “Tidak ditemukan kata-kata aneh atau tidak bermakna dalam hasil prediksi” dan “Sistem menghasilkan hasil yang konsisten saat gesture yang sama diulang”.

Dari segi performa model, sistem menunjukkan kemampuan yang cukup baik dalam menghasilkan susunan kata yang sesuai dengan kaidah tata bahasa (skor 5), serta hasil prediksi yang relatif mudah dipahami oleh pengguna awam (skor 4). Sistem juga memperoleh skor positif (skor 3) dalam hal kesesuaian hasil prediksi dengan gestur yang diberikan serta akurasi dalam mengenali isyarat SIBI, yang menunjukkan bahwa model telah mampu menangkap pola-pola utama dalam data pelatihan dan mengaplikasikannya dengan cukup akurat dalam pengujian nyata. Meskipun demikian, skor sedang pada aspek “Terdapat keselarasan makna antara gestur dan hasil teks” (skor 3) menunjukkan masih adanya ruang untuk peningkatan pada pemodelan konteks semantik antar kata. Meski demikian, evaluasi juga mengungkapkan bahwa sistem masih memiliki keterbatasan dalam hal fleksibilitas terhadap variasi gaya isyarat pengguna. Hal ini tercermin dari skor yang lebih rendah (skor 2) pada pernyataan “Sistem tetap mampu menerjemahkan gesture meskipun terjadi sedikit variasi gaya pengguna”.

4.5 Model Deployment

Tahap *deployment* dalam penelitian ini dilakukan dengan mengintegrasikan model *Continuous Sign Language Recognition* (CSLR) ke dalam layanan *back-end*



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

berbasis cloud menggunakan *framework* Flask. Implementasi ini bertujuan agar model dapat diakses dan digunakan oleh aplikasi client secara fleksibel melalui dua jalur utama: REST API untuk pengolahan video utuh dan WebSocket untuk inference secara real-time berbasis *frame*.

Model dilatih menggunakan Platform Kaggle kemudian diunggah dan dijalankan pada server cloud yang menggunakan layanan Amazon Web Services (AWS), khususnya instance EC2 dengan tipe t3.medium. Pemilihan instansi ini mempertimbangkan kebutuhan komputasi yang relatif ringan namun tetap stabil untuk inference, mengingat proses ekstraksi *keypoints* dari *frame* video dan prediksi kalimat menggunakan model *deep learning* dilakukan di sisi server.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

```
def extract_keypoints_from_video(video_path):
    cap = cv2.VideoCapture(video_path)
    keypoints_seq = []
    while True:
        ret, frame = cap.read()
        if not ret:
            break
        rgb = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
        results = holistic.process(rgb)
        # Tambahkan hanya frame dengan gesture valid
        keypoints = extract_keypoints(results)
        keypoints_seq.append(keypoints)
    cap.release()
    return np.array(keypoints_seq)

@app.route('/upload', methods=['POST'])
def upload_video():
    if 'video' not in request.files:
        return jsonify({'error': 'No video file provided'}), 400

    file = request.files['video']
    if file.filename == '':
        return jsonify({'error': 'Empty filename'}), 400

    # Simpan file sementara
    video_id = str(uuid.uuid4())
    filename = f"{video_id}_{file.filename}"
    filepath = os.path.join(UPLOAD_FOLDER, filename)
    file.save(filepath)

    # Proses video
    sequence = extract_keypoints_from_video(filepath)
    if sequence is None or len(sequence) == 0:
        os.remove(filepath)
        return jsonify({'error': 'No valid keypoints detected'}), 400

    # Inference (menggunakan pipeline yang sama)
    features = encoder(np.expand_dims(sequence, axis=0), return_features=True)
    sos_token = tokenizer.word_index['<sos>']
    eos_token = tokenizer.word_index['<eos>']
    inp = [sos_token]
    out = []

    out = attention_beam_search(features, tokenizer, decoder, beam_width=5)

    sentence = ' '.join(tokenizer.index_word.get(t, '') for t in out)
    os.remove(filepath) # Hapus file setelah selesai

    return jsonify({'sentence': sentence})
```

Gambar 4. 23 Program Endpoint /upload

Seperti pada program di Gambar 4.23, Flask digunakan sebagai *back-end* utama karena kemudahan integrasinya dengan Python, yang merupakan bahasa

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

pengembangan utama penelitian ini. Untuk inference real-time, server menggunakan Flask-SocketIO untuk menerima *frame* dari klien melalui *websocket*. Sementara itu, untuk penggunaan non-realtime, seperti upload video penuh, server menyediakan endpoint REST API dengan route /upload yang akan memproses file video, mengekstrak *keypoints* menggunakan MediaPipe Holistic, kemudian melakukan inference dengan model CSLR.

```
@socketio.on('frame')
def handle_frame(data):
    """
    data['image']: base64 encoded image from client webcam
    """
    try:
        img_bytes = base64.b64decode(data['image'].split(',')[1])
        np_arr = np.frombuffer(img_bytes, np.uint8)
        frame = cv2.imdecode(np_arr, cv2.IMREAD_COLOR)
        rgb = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)

        results = holistic.process(rgb)
        keypoints = extract_keypoints(results)
        buffer.append(keypoints)

        if len(buffer) >= SEQUENCE_LENGTH:
            sequence = np.array(buffer)

            sos_token = tokenizer.word_index['<sos>']
            eos_token = tokenizer.word_index['<eos>']
            inp = [sos_token]
            out = []

            out = attention_beam_search(features, tokenizer, decoder, beam_width=5)

            sentence = ' '.join(tokenizer.index_word.get(t, '') for t in out)
            print(sentence)
            emit('prediction', {'sentence': sentence})
            buffer.clear()

    except Exception as e:
        emit('error', {'message': str(e)})
```

Gambar 4. 24 Program Endpoint /upload

Sementara itu, program pada gambar 4.24 menjelaskan skenario *real-time*. server menyediakan endpoint WebSocket yang menangani event *frame*. *Frame* dikirimkan oleh klien dalam format base64, kemudian didekode, diubah ke format RGB, dan diproses dengan MediaPipe Holistic untuk memperoleh *keypoints*. Setiap *keypoints* disimpan sementara dalam buffer hingga jumlah mencapai batas sekuens (misalnya 30 *frame*). Setelah cukup *frame* dikumpulkan, sistem memproses seluruh buffer untuk melakukan inferensi seperti pada jalur batch, dengan hasil prediksi



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

kalimat dikirim kembali ke klien melalui *event* websocket “prediction”. Sistem juga dilengkapi mekanisme penanganan *error* agar jika terjadi masalah, misalnya kegagalan ekstraksi *keypoints* atau *input frame* yang tidak valid, *server* dapat mengirimkan notifikasi kesalahan ke klien.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

BAB V PENUTUP

5.1 Kesimpulan

Penelitian ini dilakukan untuk mengembangkan dan mengevaluasi model *Continuous Sign Language Recognition* (CSLR) berbasis deep learning yang mampu menerjemahkan gesture bahasa isyarat ke dalam bentuk kalimat teks secara otomatis. Fokus utama penelitian ini adalah pada pembangunan model dengan pendekatan *Joint CTC-Attention*, yang menggabungkan keunggulan Connectionist Temporal Classification (CTC) dalam alignment sequence dan mekanisme *Attention* untuk memahami konteks linguistik. Model ini dilatih menggunakan dataset hasil ekstraksi *keypoints* tiga dimensi dari wajah, tangan, dan tubuh, yang diperoleh melalui *framework* MediaPipe Holistic.

Melalui implementasi *Beam Search*, Hasil evaluasi menunjukkan bahwa arsitektur BiGRU dengan augmentasi memberikan performa terbaik dengan WER sebesar **0.2620**, mengungguli model BiLSTM yang memiliki WER sebesar **0.3813** pada kondisi augmentasi yang sama. Sementara itu, model tanpa augmentasi menunjukkan performa yang jauh lebih rendah, dengan WER mendekati 1. Temuan ini menunjukkan bahwa keberadaan data augmentasi sangat berpengaruh terhadap kemampuan model dalam memahami variasi gesture bahasa isyarat.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan lebih lanjut agar model *Continuous Sign Language Recognition* (CSLR) dapat berfungsi secara lebih optimal dan adaptif, antara lain:

- a. Perluasan ukuran dan variasi dataset, tidak hanya mencakup kalimat dasar tetapi juga kalimat yang lebih kompleks, percakapan umum, serta melibatkan berbagai gaya gestur dari pengguna dengan latar belakang yang berbeda. Hal ini diharapkan dapat meningkatkan kemampuan generalisasi model terhadap ragam bahasa isyarat yang lebih luas.



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Perluasan ukuran ini juga berguna bagi model CSLR dalam mengenali Gerakan variatif dari berbagai peraga.

- b. melatih ulang model menggunakan dataset yang lebih besar dan representatif, serta mengeksplorasi arsitektur deep learning yang lebih optimal, misalnya model berbasis Transformer atau kombinasi CTC-*Attention* dengan Conformer yang lebih kuat dalam memahami konteks temporal gesture.





Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

DAFTAR PUSTAKA

- Dhruv, A. J., Patel, R. & Doshi, N., 2022. *Python: The Most Advanced Programming Language for Computer Science Applications*. Gandhinagar, SCITEPRESS.
- Indriani, Harris, M. & Agoes, A. S., 2021. *Applying Hand Gesture Recognition for User Guide Application Using MediaPipe*. Bandung, Atlantis Press International B.V.
- Pang, B., Nijkamp, E. & Wu, Y. N., 2020. Deep Learning With TensorFlow: A Review. *Journal of Educational and Behavioral Statistics*, 2(45), pp. 227-248.
- Choudhary, K. et al., 2022. Recent advances and applications of deep learning methods in materials science. *npj Computational Materials*, 8(59).
- Rebala, G., Ravi, A. & Churiwala, S., 2019. *An Introduction to Machine Learning*. Cham: Springer Nature Switzerland.
- Holzinger, A., 2019. Introduction to MACHINE Learning & Knowledge Extraction (MAKE). *Mach. Learn. Knowl*, Volume 1, pp. 1-20.
- Choi, R. Y. et al., 2020. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Translational Vision Science & Technology*, Volume 9.
- Athira, P. K., Sruthi, . C. J. & Lijiya, A., 2019. A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario. *Journal of King Saud University – Computer and Information Sciences*, Volume 34, pp. 771-781.
- Rahma, N. N., 2024. IMPLEMENTASI METODE MEMBACA AL-QUR'AN DENGAN BAHASA ISYARAT PADA PESERTA DIDIK TUNARUNGU DI SLB NEGERI 12 JAKARTA. pp. 40-41.
- Salsabila, A., 2022. POLA KOMUNIKASI GURU TERHADAP SISWA TUNARUNGU (Studi Kasus Siswa Sekolah Dasar Di SLB-B Don Bosco Wonosobo). *Jurnal Komunikasi dan Media*, Volume 1.
- Pamungkas, B., Wahab, R. & Suwarjo, S., 2023. Teaching of the Quran and Hadiths Using Sign Language to Islamic Boarding School Students with



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Hearing Impairment. *International Journal of Learning, Teaching and Educational Research*, 22(5), pp. 227-242.

Huang, E. G., Kusumawati, Y. A. & Gunawan, E. P., 2023. Deafvoir: Recognizing Sign Language Through Game. *Procedia Computer Science*, Volume 227, p. 614–622.

Rahmah, W. A. & Janah, S., 2024. EFEKTIVITAS PENGGUNAAN BAHASA ISYARAT BAGI PENYANDANG DISABILITAS DALAM PENDIDIKAN INKLUSIF DI SLB BINA MANDIRI. *Jurnal Ilmiah Multidisiplin Terpadu*, Volume 8.

Husna, S. F. D., Desintha, . S. & Mustikawan, A., 2024. PERANCANGAN BUKU "JARI BERBICARA" SEBAGAI PENGENALAN BAHASA ISYARAT BAGI KALANGAN GENERASI Z.

Laines, D., Gonzalez-Mendoza, M., Ochoa-Ruiz, G. & Bejarano, G., 2023. *Isolated sign language recognition based on tree structure skeleton images*. Vancouver, IEEE/CVF.

Pramanto, H. & Suharjo, 2023. *Continuous Sign Language Recognition Using Combination of Two Stream 3DCNN and SubUNet*. *JURNAL TEKNIK INFORMATIKA*, 16(2), pp. 170-184.

Lugaresi, C. et al., 2019. MediaPipe: A *Framework for Perceiving and Processing Reality*. In *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*.

Soullard, Y., Ruffino, C. & Thierry, P., 2019. *CTCModel: a Keras Model for Connectionist Temporal Classification*, Rouen: université de Rouen Normandie.

Kumari, D. & Anand, R. S., 2024. Isolated Video-Based Sign Language Recognition Using a Hybrid CNN-LSTM *Framework Based on Attention Mechanism*. *Electronics*, 13(1229).

Yang, X., 2020. An Overview of the *Attention Mechanisms in Computer Vision*. *Journal of Physics: Conference Series*, Volume 1693.

Li, D., Rodriguez, C., Yu, X. & LI, H., 2020. Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Comparison. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1459-1469.

Zhu, Q., Li, J., Yuan, F. & al, e., 2023. *Continuous Sign Language Recognition Via Temporal Super-Resolution Network*. *Arab J Sci Eng*, Volume 48.

Zhou, H., Zhou, W., Zhou, Y. & Li, H., 2022. Spatial-Temporal Multi-Cue Network for Sign Language Recognition and Translation. *IEEE Transactions on Multimedia*, Volume 24, pp. 768-779.

Li, H. et al., 2020. *KEY ACTION AND JOINT CTC-ATTENTION BASED SIGN LANGUAGE RECOGNITION*. Barcelona, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).



Hak Cipta :

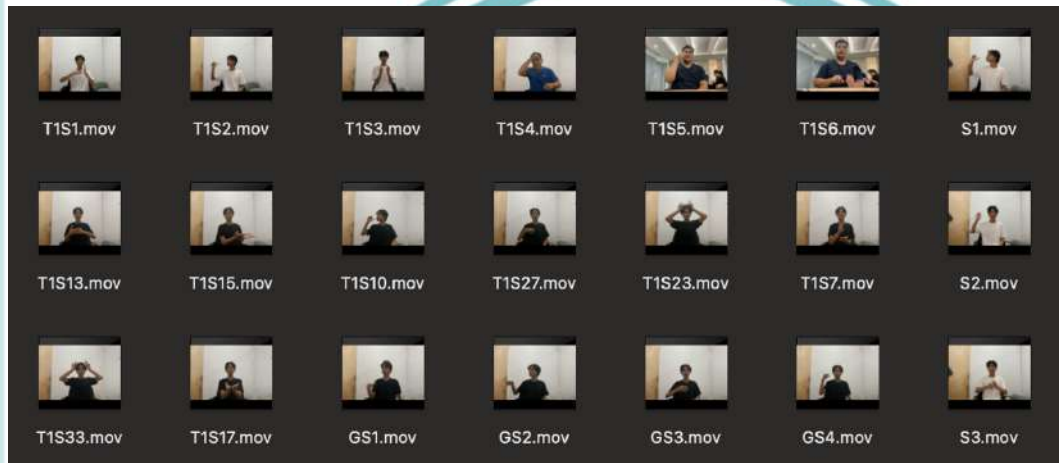
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

LAMPIRAN

Lampiran 1 – Buku Belajar Bahasa Isyarat Untuk Anak Tunarungu (Dasar)

<https://bintangpusnas.perpusnas.go.id/konten/BK8328/belajar-bahasa-isyarat-untuk-anak-tunarungu-dasar>

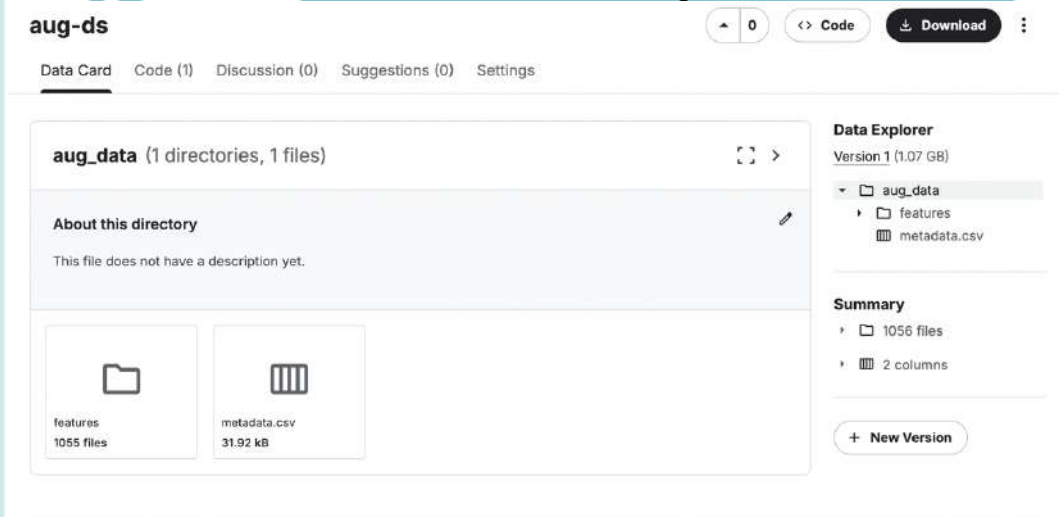
Lampiran 2 – Dataset Video Bahasa Isyarat SIBI



Dapat diakses melalui tautan berikut:

<https://drive.google.com/drive/folders/1zQAvukWJMxKor9SOM3PberSDwCu1rzt?usp=sharing>

Lampiran 3 – Dataset Sekuensial



Dapat diakses melalui tautan berikut:

Tanpa Augmentasi:

<https://www.kaggle.com/datasets/muhammadriziq54/non-aug>

Dengan Augmentasi:

<https://www.kaggle.com/datasets/muhammadriziq54/aug-ds>

Lampiran 4 – Program Training Beserta Log Evaluasi model

Our Silence Training

Notebook Input Output Logs Comments (0) Settings

```
In [1]:
import os
import pickle
import logging
import numpy as np
import pandas as pd
import tensorflow as tf
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.losses import SparseCategoricalCrossentropy
from tensorflow.keras.callbacks import ModelCheckpoint, CSVLogger, TensorBoard, LearningRateScheduler
```

Version 50 of 50

Runtime

56m 28s · GPU T4 x2

Input

DATASETS

aug-ds

cleaned-ds

ds-new

features

→ See All

Dapat diakses melalui tautan berikut:

<https://www.kaggle.com/code/muhammadriziq54/our-silence-training>

Lampiran 5 – Model Hasil Training

CSLR CTC-ATTENTION

muhammadriziq54/csrlr-ctc-attention

Model Card Code (0) Discussion (0) Competitions (0) Settings

Pending Actions

Add a subtitle

Stand out on the listings page with a snappy subtitle

Add tags

Make it easy for users to find your model in search

Fill in documentation

Add a model description, and for each model instance, add an overview and example usage

Specify provenance

Let others know how the data was collected and organized in the metadata tab

Model Details

This model does not have a description yet.

Edit

Downloads

0

0 in the last 30 days

Tags

Dapat diakses melalui tautan berikut:

<https://www.kaggle.com/models/muhammadriziq54/csrlr-ctc-attention>



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Lampiran 5 – Respon Kuesioner oleh Pakar

Kuesioner Sistem Pengenalan Bahasa Isyarat

Assalamualaikum Wr. Wb.
Selamat Pagi/Siang/Sore/Malam

Kepada Yth.
Responden penelitian
di Tempat

Perkenalkan, saya Muhammad Riziq Ramadhan mahasiswi tingkat akhir jurusan Teknik Informatika dan Komputer dari Politeknik Negeri Jakarta. Saat ini saya sedang melakukan penelitian guna menyelesaikan tugas akhir atau skripsi sebagai salah satu syarat kelulusan Diploma 4 (D4). Sehubungan dengan hal tersebut, saya memohon kesediaannya untuk berpartisipasi sebagai responden dalam penelitian ini. Silahkan klik [tautan ini](#) untuk melakukan uji coba sistem sebelum mengisi kuesioner di bawah

Semua informasi yang diberikan akan terjaga kerahasiaannya dan hanya akan digunakan untuk keperluan penelitian. Terima kasih atas perhatian dan kesediaannya.
Wassalamualaikum Wr. Wb.

Salam hormat,
Muhammad Riziq Ramadhan

Nama Lengkap *

Muhammad Yahya

Asal Sekolah *

SLBN 02 Lenteng agung

Nomor Telepon

087883844678



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Usia *

29

Bahasa Isyarat yang dipelajari di sekolah *

- ☐ SIBI
☐ BISINDO
☒ Keduanya

Penilaian Sistem

Petunjuk: Berikan penilaian untuk setiap pernyataan berikut dengan skala:

1 = Sangat Tidak Setuju | 2 = Tidak Setuju | 3 = Netral | 4 = Setuju | 5 = Sangat Setuju

Bagian A – Kesesuaian Terjemahan Bahasa Isyarat

Kalimat hasil prediksi sistem sesuai dengan gestur isyarat yang diberikan. *

- | | | | | |
|-----------------------|-----------------------|----------------------------------|-----------------------|-----------------------|
| 1 | 2 | 3 | 4 | 5 |
| ☐ | ☐ | ☒ | ☐ | ☐ |

Susunan kata pada kalimat prediksi sudah mengikuti tata bahasa yang benar.

- | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|----------------------------------|
| 1 | 2 | 3 | 4 | 5 |
| ☐ | ☐ | ☐ | ☐ | ☒ |

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Hasil terjemahan sistem mudah dipahami oleh pengguna awam. *

1	2	3	4	5
☐	☐	☐	☒	☐

Sistem mampu mengenali gestur bahasa isyarat SIBI dengan cukup akurat. *

1	2	3	4	5
☐	☐	☒	☐	☐

Proses inferensi berjalan lancar tanpa error *

1	2	3	4	5
☐	☐	☐	☐	☒

Bagian B – Kualitas Linguistik

Pilihan kata dalam kalimat prediksi sudah sesuai konteks. *

1	2	3	4	5
☐	☐	☐	☒	☐

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Sistem tetap stabil saat input dilakukan secara real-time. *

1	2	3	4	5
☐	☐	☐	☐	☒

Bagian D – Usability dan Potensi Implementasi

Sistem ini berpotensi digunakan dalam lingkungan pendidikan inklusif. *

1	2	3	4	5
☐	☐	☐	☐	☒

Antarmuka sistem mudah dipahami oleh pengguna non-teknis. *

1	2	3	4	5
☐	☐	☐	☐	☒

Sistem dapat membantu pengenalan bahasa isyarat pada anak atau pemula. *

1	2	3	4	5
☐	☐	☐	☒	☐

Bagian E – Penilaian Umum



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

77

Secara keseluruhan, saya puas terhadap performa sistem. *

1	2	3	4	5
☐	☐	☐	☒	☐

Sistem ini layak untuk dikembangkan lebih lanjut dalam konteks pembelajaran bahasa isyarat. *

1	2	3	4	5
☐	☐	☐	☐	☒

Konten ini tidak dibuat atau didukung oleh Google.

Google Formulir

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta

2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Tidak ditemukan kata-kata aneh atau tidak bermakna dalam hasil prediksi. *

1	2	3	4	5
☐	☐	☐	☐	☒

Terdapat keselarasan makna antara gestur dan hasil teks. *

1	2	3	4	5
☐	☐	☒	☐	☐

Bagian C – Konsistensi dan Stabilitas Judul Tanpa Judul

Sistem menghasilkan hasil yang konsisten saat gesture yang sama diulang. *

1	2	3	4	5
☐	☐	☐	☐	☒

Sistem tetap mampu menerjemahkan gesture meskipun terjadi sedikit variasi gaya pengguna. *

1	2	3	4	5
☐	☒	☐	☐	☐



Lampiran 6 – VALIDASI RESPONDEDN

SURAT VALIDASI EVALUASI RESPONDEN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Yahya

Profesi : Pengguna Bahasa Isyarat Aktif

Dengan ini menyatakan bahwa saya telah melakukan evaluasi terhadap produk skripsi mahasiswa atas nama:

Nama Mahasiswa : Muhammad Riziq Ramadhan Al Arafah

NIM : 2107411035

Judul Skripsi : Rancang Bangun Aplikasi Belajar Bahasa Isyarat sebagai Media Pembelajaran berbasis Website: Implementasi Model Pengenalan Bahasa Isyarat Lanjutan menggunakan Mediapipe dan CTC-Attention

Setelah melakukan evaluasi terhadap produk dan isi skripsi tersebut, saya menyatakan bahwa:

- 1) Produk telah layak digunakan sebagai media pembelajaran bahasa isyarat dasar bagi masyarakat umum.
- 2) Model pengenalan isyarat (CSLR) yang diimplementasikan menunjukkan performa yang cukup baik dalam mengenali urutan gerakan secara kontinu, ditinjau dari segi praktik bahasa isyarat.
- 3) Materi dan tampilan antarmuka aplikasi telah sesuai dan mudah dipahami oleh pengguna dengan tingkat pengetahuan rendah terhadap bahasa isyarat.
- 4) Secara keseluruhan, produk dapat divalidasi sebagai media pembelajaran yang relevan dan bermanfaat untuk digunakan dalam pembelajaran bahasa isyarat di tingkat dasar.

Demikian surat validasi ini dibuat dengan sebenarnya untuk dapat digunakan sebagaimana mestinya.

Jakarta, 20 Juli 2025

Validator,

Muhammad Yahya

Hak Cipta :
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta