BAB III
METODE PENELITIAN

3.1 Objek dan Subjek Penelitian

Objek penelitian ini adalah proses pengenalan gerakan bahasa isyarat berbasis *computer vision* dengan menerapkan algoritma *Bidirectional Long Short-Term Memory* (Bi-LSTM). Fokus kajian meliputi pengolahan data video bahasa isyarat, ekstraksi fitur visual berupa titik *landmark* menggunakan pustaka MediaPipe, serta pemodelan data sekuensial guna menghasilkan keluaran berupa klasifikasi gerakan isyarat.

Adapun subjek penelitian ini adalah sistem pengenalan gerakan bahasa isyarat yang diimplementasikan dalam bentuk aplikasi berbasis web. Sistem tersebut berfungsi sebagai media implementasi sekaligus wahana pengujian model Bi-LSTM yang telah dilatih menggunakan dataset video bahasa isyarat yang bersumber dari platform Kaggle. Melalui aplikasi web yang dikembangkan, sistem menerima masukan berupa video gerakan isyarat yang direkam secara langsung melalui kamera perangkat pengguna. Video masukan tersebut selanjutnya diproses melalui tahap ekstraksi fitur *landmark* menggunakan MediaPipe, kemudian dianalisis oleh model Bi-LSTM untuk mengenali pola gerakan isyarat yang bersangkutan.

Perlu ditegaskan bahwa dalam penelitian ini, pengguna aplikasi tidak berkedudukan sebagai subjek utama penelitian, melainkan sebagai pihak yang dilibatkan pada tahap pengujian sistem. Keterlibatan pengguna tersebut bertujuan untuk memverifikasi bahwa sistem dapat beroperasi sebagaimana mestinya serta mampu menghasilkan prediksi gerakan isyarat secara akurat dan andal.

3.2 Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian implementasi algoritma dengan pendekatan kuantitatif, yang difokuskan pada penerapan serta evaluasi kinerja algoritma *Bidirectional Long Short-Term Memory* (Bi-LSTM) dalam mengenali pola gerakan bahasa isyarat berbasis *computer vision*. Pendekatan kuantitatif digunakan untuk mengukur kinerja model secara objektif melalui pelatihan dan pengujian dengan dataset video bahasa isyarat Kaggle. Evaluasi kinerja model dilakukan berdasarkan

kemampuannya dalam mengklasifikasikan gerakan isyarat dengan tingkat akurasi yang terukur.

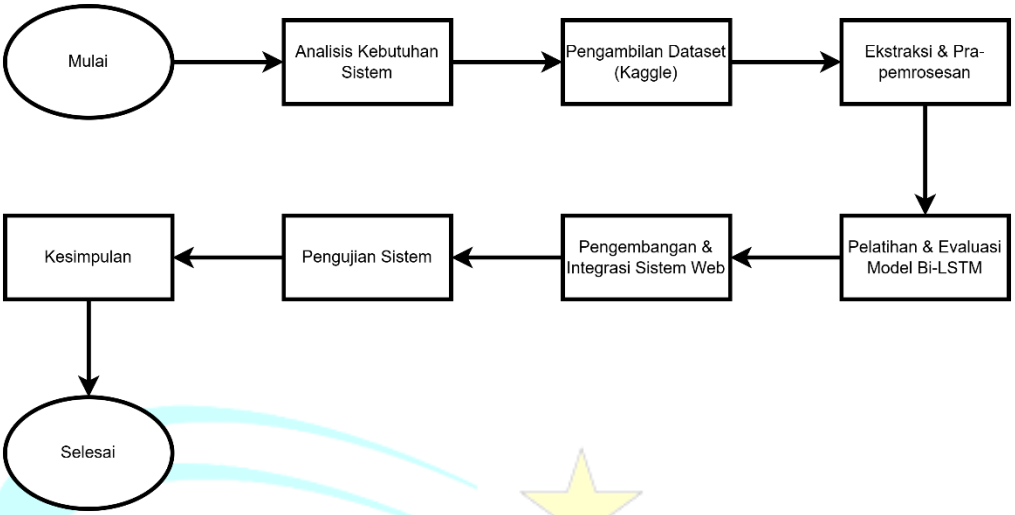
Dalam proses pengolahan data, video bahasa isyarat diproses melalui tahap ekstraksi fitur menggunakan pustaka MediaPipe guna memperoleh koordinat titik *landmark* pada tubuh dan tangan. Data hasil ekstraksi tersebut selanjutnya disusun dalam bentuk sekuensial sebagai masukan bagi model Bi-LSTM dalam menjalankan proses klasifikasi gerakan isyarat. Pendekatan berbasis data sekuensial ini dipilih karena mampu merepresentasikan dinamika temporal gerakan isyarat secara lebih komprehensif.

Model yang telah dilatih diintegrasikan ke dalam antarmuka web untuk pengujian dan implementasi sistem secara *real-time*. Studi ini mengevaluasi akurasi dan efisiensi pengenalan gestur bahasa isyarat dunia nyata dari algoritma Bi-LSTM.

3.3 Prosedur Penelitian

Prosedur penelitian ini menguraikan tahapan yang dilaksanakan dalam implementasi algoritma *Bidirectional Long Short-Term Memory* (Bi-LSTM) untuk pengenalan gerakan bahasa isyarat berbasis *computer vision* yang diintegrasikan ke dalam aplikasi berbasis web. Tahapan penelitian disusun secara sistematis dan berurutan, dimulai dari analisis kebutuhan sistem hingga penyusunan kesimpulan penelitian.

Secara umum, prosedur penelitian terdiri dari tujuh tahapan, yaitu: analisis kebutuhan sistem, pengambilan *dataset*, ekstraksi dan pra-pemrosesan data, pelatihan dan evaluasi model Bi-LSTM, pengembangan dan integrasi sistem web, pengujian sistem, serta penyusunan kesimpulan penelitian. Alur keseluruhan prosedur penelitian divisualisasikan dalam bentuk *flowchart* sebagaimana ditunjukkan pada Gambar 3.1.



Gambar 3.1 Flowchart Prosedur Penelitian

Berdasarkan *flowchart* pada Gambar 3.1, setiap tahapan penelitian dapat diuraikan sebagai berikut:

1. Analisis Kebutuhan Sistem

   Tahap ini dilaksanakan untuk mengidentifikasi seluruh kebutuhan sistem, meliputi kebutuhan perangkat keras (*hardware*), perangkat lunak (*software*), serta metode yang akan diterapkan dalam pengembangan sistem pengenalan gerakan bahasa isyarat.

2. Pengambilan *Dataset*

   Dataset yang digunakan diperoleh dari platform Kaggle, yaitu Word-Level BISINDO (WL-BISINDO) yang memuat data video gerakan bahasa isyarat Indonesia. Meskipun *dataset* awal mencakup 32 kelas *gesture*, pada penelitian ini dilakukan proses seleksi menjadi 12 kelas yang lebih representatif guna meningkatkan fokus pembelajaran model dan memperoleh hasil klasifikasi yang lebih optimal.

3. Ekstraksi dan Pra-pemrosesan Data

   Data video diproses menggunakan pustaka MediaPipe untuk mengekstraksi koordinat titik landmark tangan pada setiap frame. Hasil ekstraksi selanjutnya dinormalisasi dan dikonversi menjadi data sekuensial dengan panjang maksimum tertentu (MAX_LEN), sehingga dapat digunakan sebagai masukan bagi model Bi-LSTM dalam proses klasifikasi.

4. Pelatihan dan Evaluasi Model Bi-LSTM

   Model Bi-LSTM dilatih menggunakan data sekuensial hasil pra-pemrosesan untuk mempelajari pola gerakan bahasa isyarat secara temporal. Proses pelatihan dilaksanakan dengan memanfaatkan platform Google Colab. Selanjutnya, model dievaluasi menggunakan beberapa metrik pengukuran, yaitu *accuracy*, *loss*, dan *confusion matrix*, guna mengetahui performa model dalam melakukan klasifikasi *gesture* secara menyeluruh.

5. Pengembangan dan Integrasi Sistem Web

   Tahap ini mencakup perancangan dan pembangunan aplikasi berbasis web sebagai antarmuka (*interface*) sistem. Model Bi-LSTM yang telah melalui proses pelatihan kemudian diintegrasikan ke dalam aplikasi, sehingga sistem mampu menerima masukan berupa video dari kamera secara langsung dan menghasilkan prediksi gerakan isyarat secara *real-time*.

6. Pengujian Sistem

   Pengujian sistem dilaksanakan untuk memverifikasi bahwa aplikasi mampu beroperasi dengan baik dalam mengenali gerakan bahasa isyarat secara *real-time*. Pengujian dilakukan dengan memberikan masukan berupa gerakan isyarat melalui kamera perangkat, kemudian hasil prediksi ditampilkan pada antarmuka sistem sebagai keluaran akhir.

7. Kesimpulan

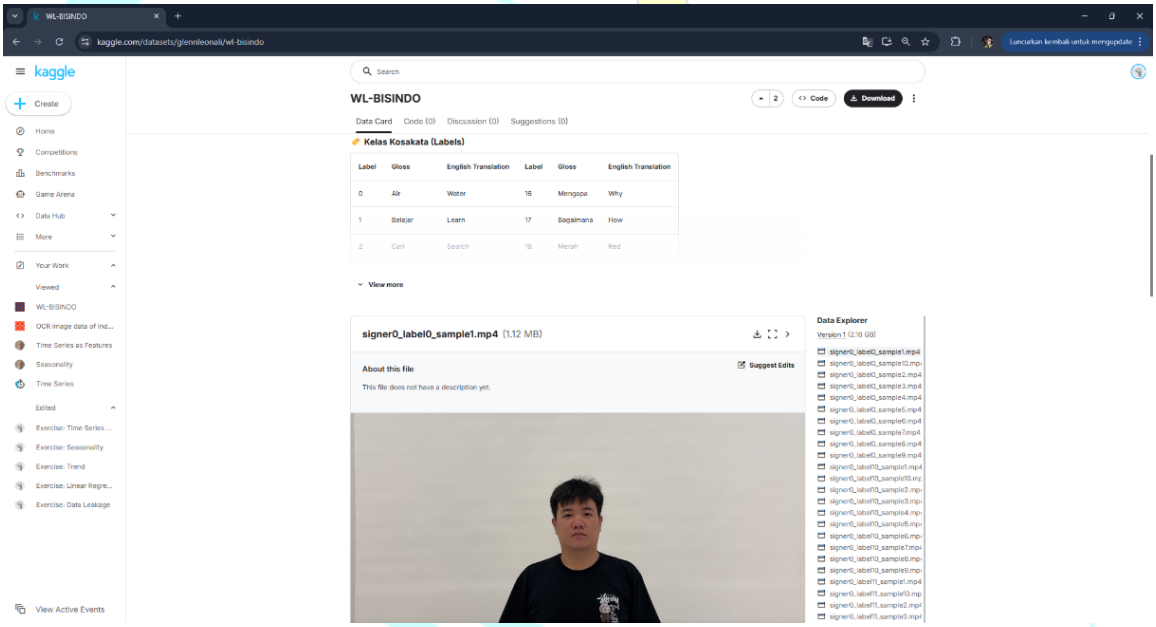
   Tahap akhir penelitian adalah menarik kesimpulan dari pengujian dan evaluasi yang ekstensif. Kesimpulan ini mencakup tujuan penelitian dan memberikan evaluasi lengkap terhadap sistem pengenalan isyarat bahasa isyarat.

3.4 Dataset dan Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, yaitu *dataset* Word-Level BISINDO (WL-BISINDO) yang dapat diakses melalui tautan berikut: https://www.kaggle.com/datasets/glennleonali/wl-bisindo. WL-BISINDO merupakan *dataset* video Bahasa Isyarat Indonesia (BISINDO) yang berfokus pada varian regional Banten, dengan ukuran sekitar 2,16 GB dan terdiri dari 1.600 video RGB yang mencakup 32 kelas gerakan isyarat terisolasi. Setiap

kelas direpresentasikan oleh sejumlah video yang diperagakan oleh beberapa penandatangan (*signer*) dengan variasi gerakan yang beragam.

Dengan mempertimbangkan jumlah data pada setiap kelas yang relatif terbatas, yaitu sekitar 50 video per kelas, serta tingginya kompleksitas klasifikasi pada keseluruhan *dataset*, penelitian ini melakukan proses seleksi dengan mereduksi jumlah kelas menjadi 12 kelas *gesture* yang lebih representatif. Pemilihan kelas dilakukan berdasarkan ketersediaan data yang memadai serta kejelasan perbedaan pola gerakan antar kelas, sehingga proses pelatihan model dapat berlangsung secara lebih optimal dan terarah.



Gambar 3.2 Struktur Dataset Word-Level BISINDO

Sumber: https://www.kaggle.com/datasets/glennleonali/wl-bisindo

Model *Bidirectional Long Short-Term Memory* (Bi-LSTM) dilatih dan diuji menggunakan dataset yang diberikan. Gambar 3.2 menunjukkan bahwa dataset tersebut diorganisir dengan baik berdasarkan kelas gestur dalam folder, sehingga memudahkan pemuatan dan pemrosesan data. Data video kemudian diproses melalui tahap ekstraksi fitur menggunakan pustaka MediaPipe untuk memperoleh koordinat titik *landmark* tangan pada setiap *frame*. Hasil ekstraksi tersebut selanjutnya dinormalisasi dan dikonversi menjadi data sekuensial dengan panjang maksimum tertentu (*MAX_LEN*), yang digunakan sebagai masukan bagi model Bi-LSTM dalam menjalankan proses klasifikasi gerakan bahasa isyarat.

3.5 Pra-Pemrosesan Data

Pra-pemrosesan data dilakukan untuk mempersiapkan *dataset* video gerakan bahasa isyarat sebelum digunakan pada tahap pelatihan model *Bidirectional Long Short-Term Memory* (Bi-LSTM). Tahap ini mengubah data video mentah menjadi representasi numerik terstruktur untuk *machine learning*.

Pada penelitian ini, pra-pemrosesan data dilaksanakan menggunakan pustaka MediaPipe *Hands* untuk mengekstraksi fitur gerakan tangan dari setiap video dalam *dataset*. Pendekatan ini dipilih karena gerakan bahasa isyarat didominasi oleh pergerakan tangan, sehingga penggunaan titik *landmark* tangan dinilai representatif sekaligus efisien dalam proses pengolahan data. Adapun tahapan pra-pemrosesan data adalah sebagai berikut:

1. Pembacaan Video

   Setiap berkas video dalam *dataset* dibaca menggunakan pustaka OpenCV untuk memperoleh *frame* secara berurutan. Seluruh *frame* digunakan sebagai representasi gerakan tanpa dilakukan proses *frame sampling*, sehingga informasi temporal tetap terjaga secara utuh.
2. Deteksi *Landmark* Tangan

   Setiap *frame* diproses menggunakan MediaPipe *Hands* untuk mendeteksi posisi tangan secara otomatis. Sistem dikonfigurasi untuk mendeteksi maksimal dua tangan pada setiap *frame*, sehingga gerakan tangan kanan dan kiri dapat ditangkap secara bersamaan.
3. Ekstraksi *Landmark*

   Berdasarkan hasil deteksi, sistem mengekstraksi 21 titik *landmark* untuk setiap tangan, di mana setiap titik direpresentasikan oleh tiga nilai koordinat (x, y, z). Dengan demikian, satu tangan menghasilkan 63 nilai (21×3), sehingga total fitur yang digunakan adalah 126 nilai untuk kedua tangan. Apabila hanya satu tangan yang terdeteksi, sistem akan mengisi nilai nol (*zero padding*) pada posisi tangan yang tidak terdeteksi guna menjaga konsistensi dimensi data.

4. Normalisasi Data

   Koordinat *landmark* yang diperoleh selanjutnya dinormalisasi dengan menjadikan titik pergelangan tangan (*wrist*) sebagai titik acuan. Proses ini bertujuan untuk mereduksi pengaruh posisi absolut tangan terhadap hasil klasifikasi sekaligus meningkatkan kestabilan model dalam mengenali pola gerakan.

5. Pembentukan Data Sekuensial

   Data *landmark* dari setiap *frame* disusun menjadi data sekuensial berdasarkan urutan waktu (*temporal order*). Setiap video direpresentasikan sebagai satu *sequence* yang terdiri dari kumpulan vektor *landmark* secara berurutan. Guna menjaga keseragaman panjang data antarvideo, dilakukan proses penyesuaian panjang *sequence* menggunakan metode *padding* hingga mencapai panjang maksimum yang telah ditentukan, yaitu sebanyak 100 *frame*.

6. Pembagian *Dataset*

   Dataset yang telah diproses sebelumnya dibagi 80:20 antara data pelatihan (*training set*) dan data pengujian (*test set*). Pembagian ini dilakukan untuk menguji kemampuan generalisasi model terhadap data non-pelatihan.

7. Augmentasi Data

   Augmentasi data dilakukan pada rangkaian data untuk meningkatkan varians data dan mengurangi potensi *overfitting* pada model. Teknik augmentasi yang diterapkan meliputi pembalikan horizontal (*horizontal flipping*) dan penambahan *Gaussian noise*. Proses augmentasi ini bertujuan untuk meningkatkan kemampuan generalisasi model terhadap variasi gerakan yang beragam. Augmentasi hanya diterapkan pada data pelatihan (*training set*) setelah proses pembagian dataset guna menghindari terjadinya *data leakage* antara data pelatihan dan data pengujian.

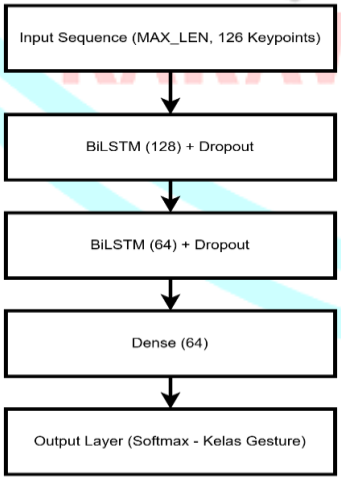
8. Penyimpanan Data

   Setiap *sequence* hasil pemrosesan disimpan dalam format berkas numerik (*.npy*). Setiap berkas merepresentasikan satu sampel data yang memuat rangkaian koordinat *landmark* tangan secara sekuensial, yang selanjutnya digunakan sebagai masukan bagi model Bi-LSTM pada tahap pelatihan.

3.6 Arsitektur Model Bi-LSTM

Arsitektur model *Bidirectional Long Short-Term Memory* (Bi-LSTM) digunakan untuk mempelajari pola gerakan bahasa isyarat berdasarkan data sekuensial hasil proses pra-pemrosesan. Model ini dirancang untuk menangkap hubungan temporal antar *frame* sehingga mampu merepresentasikan perubahan gerakan secara berurutan dan terstruktur.

Data masukan berupa rangkaian *sequence* dengan panjang maksimum (*MAX_LEN*) sebesar 100 *frame*, di mana setiap *frame* direpresentasikan dalam bentuk vektor *keypoints* sebanyak 126 fitur hasil ekstraksi *landmark* menggunakan pustaka MediaPipe. Model Bi-LSTM dipilih karena kemampuannya memproses data sekuensial dari dua arah secara bersamaan, yaitu *forward* (maju) dan *backward* (mundur), sehingga model dapat memanfaatkan konteks informasi secara lebih lengkap dari keseluruhan rangkaian *frame* dan meningkatkan akurasi dalam mengenali pola gerakan isyarat. Struktur arsitektur model yang digunakan ditunjukkan pada Gambar 3.3, yang disusun secara berurutan dari lapisan masukan hingga lapisan keluaran sebagai berikut:



Gambar 3.3 Arsitektur Model Bidirectional Long Short-Term Memory

1. *Input Layer*

   Lapisan masukan menerima data dalam bentuk *sequence* dengan ukuran (100, 126), yang merepresentasikan rangkaian *keypoints* dari setiap *frame* hasil ekstraksi menggunakan pustaka MediaPipe.

2. *Bidirectional LSTM Layer (128 units) + Dropout (0,5)*

   Lapisan ini berfungsi untuk menangkap pola temporal awal dari data sekuensial dengan memproses data dari dua arah secara bersamaan. Lapisan *dropout* dengan nilai 0,5 diterapkan setelahnya untuk mereduksi risiko *overfitting* dengan menonaktifkan sebagian neuron secara acak selama proses pelatihan.

3. *Bidirectional LSTM Layer (64 units) + Dropout (0,5)*

   Lapisan ini digunakan untuk memperdalam pembelajaran terhadap pola sekuensial yang lebih kompleks berdasarkan representasi fitur dari lapisan sebelumnya. Lapisan *dropout* kembali diterapkan dengan nilai yang sama untuk meningkatkan kemampuan generalisasi model terhadap data baru.

4. *Dense Layer (64, ReLU)*

   Lapisan *dense* berfungsi untuk mengolah representasi fitur yang dihasilkan oleh lapisan Bi-LSTM menjadi bentuk yang lebih terstruktur. Fungsi aktivasi ReLU (*Rectified Linear Unit*) memperkenalkan non-linearitas untuk membantu model mempelajari korelasi fitur yang lebih kompleks.

5. *Output Layer (Softmax)*

   Fungsi aktivasi *softmax* pada lapisan keluaran memberikan distribusi probabilitas atas 12 kelas gerakan bahasa isyarat untuk memprediksinya. Prediksi model diakhiri dengan kelas yang memiliki kemungkinan tertinggi.

Dengan arsitektur tersebut, model Bi-LSTM mampu mempelajari pola gerakan bahasa isyarat secara efektif melalui pemodelan data sekuensial yang mempertimbangkan hubungan temporal antar *frame*. Model yang telah melalui proses pelatihan selanjutnya diintegrasikan sebagai komponen inti sistem pengenalan gerakan bahasa isyarat berbasis web guna menghasilkan prediksi secara *real-time*.

3.7 Perancangan Sistem

Perancangan sistem merupakan tahapan yang dilaksanakan untuk merancang struktur dan alur kerja sistem pengenalan gerakan bahasa isyarat berbasis *computer vision* yang diintegrasikan ke dalam aplikasi berbasis web. Perancangan ini bertujuan untuk memastikan bahwa seluruh proses, mulai dari penerimaan masukan video hingga menghasilkan prediksi gerakan isyarat, dapat berjalan secara terstruktur, sistematis, dan terintegrasi dengan baik.



Gambar 3.4 Perancangan Antarmuka Sistem Berbasis Web

Berdasarkan Gambar 3.4, sistem terdiri dari beberapa komponen utama yang dapat diuraikan sebagai berikut:

1. Halaman Utama Sistem

   Halaman utama (*home page*) menampilkan judul sistem beserta deskripsi singkat mengenai fungsi aplikasi sebagai sistem pengenalan gerakan bahasa isyarat berbasis *computer vision*. Halaman ini berfungsi sebagai titik masuk utama bagi pengguna sebelum menjalankan proses pengenalan.
2. Area Hasil Pengenalan

   Bagian ini digunakan untuk menampilkan hasil prediksi gerakan isyarat yang dilakukan oleh pengguna. Keluaran yang dihasilkan berupa teks label kelas hasil klasifikasi *gesture* dari model Bi-LSTM, sehingga pengguna

dapat mengetahui gerakan isyarat yang berhasil dikenali oleh sistem secara langsung.

3. Masukan Kamera (Visual/Gesture Input)

   Sistem menerima masukan berupa gerakan isyarat yang ditangkap secara langsung melalui kamera perangkat pengguna. Komponen ini berfungsi sebagai sumber data utama yang akan diproses lebih lanjut oleh sistem dalam tahap ekstraksi fitur dan klasifikasi.

4. Tombol Aksi (Mulai Pengenalan)

   Tombol ini digunakan untuk memulai proses pengenalan gerakan isyarat secara manual oleh pengguna. Setelah tombol diaktifkan, sistem akan menjalankan proses ekstraksi titik *landmark* menggunakan pustaka MediaPipe, dilanjutkan dengan proses klasifikasi menggunakan model Bi-LSTM untuk menghasilkan prediksi gerakan isyarat.

Secara umum, alur kerja sistem diawali dari pengguna mengakses aplikasi berbasis web, kemudian memberikan masukan berupa gerakan isyarat melalui kamera perangkat. Selanjutnya, sistem menjalankan proses ekstraksi fitur titik *landmark* menggunakan pustaka MediaPipe, yang kemudian dilanjutkan dengan proses klasifikasi menggunakan model Bi-LSTM. Hasil prediksi yang diperoleh selanjutnya ditampilkan pada area hasil pengenalan sebagai keluaran akhir sistem.

3.8 Skenario Pengujian Sistem

Pengujian sistem dilaksanakan untuk mengevaluasi kinerja model *Bidirectional Long Short-Term Memory* (Bi-LSTM) dalam mengenali gerakan bahasa isyarat berbasis *computer vision*, baik pada tahap pengujian model maupun implementasi sistem berbasis web. Pengujian menggunakan *dataset* BISINDO dari platform Kaggle yang telah melalui tahap pra-pemrosesan dan dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20. Sebagian data pelatihan selanjutnya digunakan sebagai data validasi selama proses pelatihan melalui mekanisme validation split.

1. Pengujian Model

   Pengujian model bertujuan untuk mengevaluasi performa algoritma Bi-LSTM dalam mengklasifikasikan pola gerakan bahasa isyarat berdasarkan

data sekuensial hasil ekstraksi titik *landmark*. Model dilatih menggunakan data *training*, divalidasi menggunakan data *validation*, kemudian diuji menggunakan data *testing* yang sepenuhnya terpisah dari proses pelatihan. Evaluasi dilakukan menggunakan tiga metrik, yaitu *accuracy* untuk mengukur ketepatan prediksi, *loss* untuk mengukur tingkat kesalahan model, serta *confusion matrix* untuk menganalisis distribusi prediksi pada setiap kelas *gesture* dan mengidentifikasi kelas yang mengalami kesalahan prediksi (*misclassification*).

2. Pengujian Sistem Aplikasi Berbasis Web

   Pengujian aplikasi berbasis web dilakukan untuk memverifikasi bahwa sistem dapat beroperasi dengan baik dalam kondisi penggunaan nyata. Pengguna melakukan gerakan bahasa isyarat melalui kamera perangkat, kemudian sistem memproses masukan video secara *real-time* melalui tahapan ekstraksi titik *landmark* menggunakan MediaPipe, pembentukan data sekuensial, klasifikasi menggunakan model Bi-LSTM, dan penampilan hasil prediksi pada antarmuka (*interface*) sistem. Aspek yang dievaluasi meliputi kemampuan pengenalan secara *real-time*, kesesuaian hasil prediksi, serta stabilitas dan keandalan sistem secara keseluruhan.

Dengan kedua skenario pengujian tersebut, diharapkan sistem yang dikembangkan tidak hanya unggul secara teoretis, tetapi juga dapat diimplementasikan secara optimal dalam kondisi penggunaan nyata.