

# **LAPORAN PROYEK DEEP LEARNING**

## **SIBI SIGN LANGUAGE**



**Oleh:**

- 1. Alfons Richardo / C14190069**
- 2. Dick Jovian / C14190076**
- 3. Kendrew Tandio / C14190070**
- 4. Nyoto Wijaya / C14190059**
- 5. Bryan Fernando Liauwan / C14190020**
- 6. Marcellino / C14190049**

**Universitas Kristen Petra**

## 1. Intro

Komunikasi manusia adalah bagian yang sangat penting dalam kehidupan kita, penting untuk berbagi pemikiran dan ide untuk kelangsungan hidup kita. Namun sayangnya bagi minoritas tuna rungu, sulit atau tidak mungkin untuk berkomunikasi melalui ucapan. Sebagian besar masyarakat awam tidak memiliki pengetahuan tentang bahasa isyarat dan interpretasinya. Hal ini adalah salah satu masalah utama yang dihadapi oleh orang-orang seperti ini selama berkomunikasi dengan minoritas tuna rungu apalagi mendapatkan juru bahasa tidak mudah setiap saat. Pengenalan bahasa isyarat adalah masalah yang telah dibahas dalam penelitian selama bertahun-tahun, model yang kami kembangkan menggunakan jaringan saraf untuk ejaan jari berdasarkan berbagai gerakan tangan.

Model berbasis RNN ini dilatih menggunakan dataset berupa video per abjad berdasarkan Sistem Bahasa Isyarat Indonesia (SIBI), disini kami tidak menggunakan gambar karena dibutuhkan gerakan tangan untuk beberapa abjad serta melalui video posisi dan struktur setiap jari dapat diidentifikasi lebih akurat daripada gambar. Dengan menggunakan model ini, seorang tuna rungu dapat dengan mudah berkomunikasi dengan orang yang tidak memiliki pengetahuan tentang itu. Masalah utama di sistem ini adalah masih banyak *false prediction* karena terdapat beberapa gerakan tangan yang mirip. Masalah ini dapat diselesaikan dengan menggunakan penghalusan video dari dataset dimana video dipotong ketika gerakan tangan tepat menunjukkan bahasa isyarat yang terbukti dapat meningkatkan akurasi model yang lebih tinggi sekitar 0.99.

## 2. Data Collecting and Preprocessing

Untuk mengembangkan proyek ini digunakan dataset berupa video dari 26 abjad dari Sistem Bahasa Isyarat Indonesia (SIBI) yang masing-masing abjadnya diberikan 21 video training. Video diambil dari youtube dan di potong-potong sesuai abjad yang berdurasi 1-3 detik. Pemotongan video ini dilakukan secara manual agar gerakan tangan yang tidak sesuai dengan abjad dapat dihapus dengan tepat. Contoh bagian yang dihapus adalah video saat gerakan tangan bergerak dari bawah ke atas. Dengan begitu, tiap video hanya menunjukkan gerakan tangan yang jelas dan sesuai dengan abjad yang ingin

dideteksi.

Untuk pengambilan data yang akan digunakan untuk pembuatan model, digunakan *library* mediapipe untuk mengambil *keypoints* tangan kanan. Per abjad video akan diambil 10 video kecil berdurasi 30 frame, dimana tiap video kecil tersebut akan dimajukan 6 frame. Apabila saat penambahan 6 frame tersebut sudah melebihi total frame video, maka akan diulang ke frame awal lagi. Tujuan dari ini adalah agar data yang digunakan untuk pembuatan model menjadi banyak dan bervariasi.

### 3. Deep Learning Model

Kita menggunakan *LSTM (Long Short Term Memory)* untuk pembuatan model. Layer yang kami gunakan untuk percobaan akhir adalah 1 layer *LSTM* dengan 64 nodes dan *activation function relu*. Kita juga menggunakan 2 *dense layer*, dimana layer ke 1 menggunakan 64 nodes dan *activation function relu*, dan layer ke 2 menggunakan 26 nodes (jumlah total abjad) dan *activation function softmax*.

Saat meng-*compile* model, kita menggunakan *adam optimizer* dan *categorical cross entropy* untuk *loss function* nya. Kita menggunakan *loss* tersebut karena model kita akan digunakan untuk mendeteksi data *categorical* (abjad A-Z).

```
model = Sequential()
model.add(LSTM(64, return_sequences=False, activation='relu', input_shape=(30,63)))
model.add(Dense(64, activation='relu'))
model.add(Dense(actions.shape[0], activation='softmax'))

model.compile(optimizer='Adam', loss='categorical_crossentropy', metrics=['categorical_accuracy'])
```

Model kami lakukan *fit* sebanyak 271 epoch dan menghasilkan akurasi *training* sebanyak 0.95. Untuk *testing*, diperoleh akurasi sebesar 0.9.

```
Epoch 266/2000
109/109 [=====] - 1s 6ms/step - loss: 0.2570 - categorical_accuracy: 0.9275
Epoch 267/2000
109/109 [=====] - 1s 6ms/step - loss: 0.2136 - categorical_accuracy: 0.9381
Epoch 268/2000
109/109 [=====] - 1s 6ms/step - loss: 0.2387 - categorical_accuracy: 0.9316
Epoch 269/2000
109/109 [=====] - 1s 7ms/step - loss: 0.2401 - categorical_accuracy: 0.9316
Epoch 270/2000
109/109 [=====] - 1s 7ms/step - loss: 0.2348 - categorical_accuracy: 0.9312
Epoch 271/2000
9/109 [=>.....] - ETA: 0s - loss: 0.1615 - categorical_accuracy: 0.9527
```

**Training**

```
] : accuracy_score(ytrue, yhat)  
]: 0.9008097165991903
```

---

### Testing

## 4. Progress Report

### Progress Report 0

Tanggal : 13 Juni 2022 (*before*)

1. **Total Model:** 10 model dengan beragam layer
2. **Hasil Accuracy:** accuracy di atas 0.9 secara menyeluruh.
3. **Kesimpulan:**
  - a. Gerakan terlalu sensitif sehingga gerakan tangan menghasilkan abjad yang lain.
  - b. Dataset perlu ditambah agar lebih banyak dan baik.
  - c. Dataset perlu di-*filter* ulang agar lebih baik dan cocok dengan abjad tertentu.
  - d. Perlu percobaan variasi layer-layer berikutnya.

### Progress Report 1

Tanggal : 14 Juni 2022

1. **Variabel awal:**
  - a. Total Video = 20
  - b. No\_sequences = 10
  - c. Sequence\_length = 30
2. **Waktu pengambilan keypoints:**
  - a. Belum mengambil tangan kiri
  - b. Min\_detection\_confidence = 0.4
  - c. Min\_tracking\_confidence = 0.4

d. Counter tiap video maju = 6 frames

### 3. Splitting data:

a. Train test split = 95% 5%

### 4. Model:

a. LSTM layer 1 = 64 nodes, Activation = relu

b. LSTM layer 2 = 128 nodes, Activation = relu

c. Dense layer 1 = 128 nodes, Activation = relu

d. Dense layer 2 = 64 nodes, Activation = relu

e. Dense layer 3: activation = softmax

f. Compile: optimizer = adam, loss = categorical\_crossentropy, metrics = categorical\_accuracy

g. Hasil training: epoch 401 = 0.9917

```
140/140 [=====] - 3s 18ms/step - loss: 0.0786 - categorical_accuracy: 0.9767
Epoch 397/2000
140/140 [=====] - 3s 18ms/step - loss: 0.1353 - categorical_accuracy: 0.9552
Epoch 398/2000
140/140 [=====] - 3s 18ms/step - loss: 0.1458 - categorical_accuracy: 0.9526
Epoch 399/2000
140/140 [=====] - 3s 18ms/step - loss: 0.1478 - categorical_accuracy: 0.9493
Epoch 400/2000
140/140 [=====] - 3s 18ms/step - loss: 0.0649 - categorical_accuracy: 0.9816
Epoch 401/2000
25/140 [====>.....] - ETA: 2s - loss: 0.0411 - categorical_accuracy: 0.9917
```

```
KeyboardInterrupt                                Traceback (most recent call last)
```

h. Hasil testing: 0.94

```
: accuracy_score(ytrue, yhat)
: 0.940677966101695
```

### 5. Percobaan RealTime:

a. Threshold = 0.9

b. Min\_detection\_confidence = 0.8

c. Min\_tracking\_confidence = 0.8

### 6. Kesimpulan:

a. Mediapipe masih kurang bisa deteksi apabila posisi tangan dibawah

- b. Mediapipe masih sering mendeteksi tangan walaupun tidak ada tangan di dalam frame
- c. Masih terjadi banyak false prediction, terutama waktu tangan sedang gerak tidak teratur (gabung dengan poin b)
- d. P dan Q masih kurang dapat dideteksi (lihat poin a)
- e. M dan N masih sering kebalik (posisi tangan sangat mirip)
- f. R, U, V masih sering kebalik (posisi tangan sangat mirip)
- g. T dan X masih sering kebalik (posisi tangan sangat mirip)

**7. Save Model:**

- a. Nama file data keypoints = MP\_Data21Vid2
- b. Nama file model = model21Vidlayer.h5

## **Progress Report 2**

**Tanggal : 15 Juni 2022**

**1. Preprocessing dataset manual:**

- a. Menghapus frame-frame video dataset yang tidak penting (gerakan tangan yang bukan abjad, frame tanpa tangan di layer, ada beberapa video abjad yang abjad setelahnya kelihatan di dalam frame (contoh: abjad A awalnya tapi frame belakang-belakang ada menunjukkan abjad B sepersekian detik))

**2. Variabel awal:**

- a. Total Video = 21
- b. No\_sequences = 10
- c. Sequence\_length = 30

**3. Waktu pengambilan keypoints:**

- a. Belum mengambil tangan kiri
- b. Min\_detection\_confidence = 0.4
- c. Min\_tracking\_confidence = 0.4
- d. Counter tiap video maju = 6 frames

**4. Splitting data:**

- a. Train test split = 95% 5%

**5. Model:**



- e. M dan N masih sering kebalik (posisi tangan sangat mirip)
- f. R dan U masih sering kebalik (posisi tangan sangat mirip)
- g. T dan X mulai jarang kebalik (posisi tangan sangat mirip)

#### 8. Save Model:

- a. Nama file data keypoints = MP\_Data21VidClean
- b. Nama file model = model21VidCleanlayer.h5

## Progress Report 3

#### 1. Splitting data:

- a. Train test split = 70% 30%

#### 2. Model:

- a. LSTM layer 1 = 64 nodes, Activation = relu
- b. Dense layer 2 = 64 nodes, Activation = relu
- c. Dense layer 3: activation = softmax
- d. Compile: optimizer = adam, loss = categorical\_crossentropy, metrics = categorical\_accuracy
- e. Hasil training: epoch 234 = 0.9926

```
Epoch 232/2000
109/109 [=====] - 1s 7ms/step - loss: 0.0390 - categorical_accuracy: 0.9872
Epoch 233/2000
109/109 [=====] - 1s 7ms/step - loss: 0.0539 - categorical_accuracy: 0.9863
Epoch 234/2000
48/109 [=====>.....] - ETA: 0s - loss: 0.0276 - categorical_accuracy: 0.9926
```

```
-----
KeyboardInterrupt                                Traceback (most recent call last)
<ipython-input-48-610fed43f9e2> in <module>
```

- f. Hasil testing: 0.95

```
accuracy_score(ytrue, yhat)
0.9568151147098516
```

#### 3. Save Model:

- a. Nama file data keypoints = MP\_Data21VidClean
- b. Nama file model = model21VidClean1layer.h5



#### **4. Kesimpulan:**

- a. Mediapipe masih kurang bisa deteksi apabila posisi tangan kebawah
- b. Mediapipe MULAI JARANG mendeteksi tangan walaupun tidak ada tangan di dalam frame
- c. Masih terjadi LEBIH SEDIKIT false prediction, terutama waktu tangan sedang gerak tidak teratur (gabung dengan poin b)
- d. P dan Q masih kurang akurat dideteksi namun dapat diberikan prediksi antara P dan Q (kacau) (lihat poin a)
- e. Angle untuk 3 poin di bawah harus agak cenderung ke bawah dan dekat kamera:
  - i. M dan N mulai bisa dideteksi
  - ii. V lebih sering keluar saat tangan sedang menunjukan U
  - iii. T dan X mulai bisa dideteksi

## **Progress Report 4**

#### **1. Variabel awal:**

- a. Total Video = 21
- b. No\_sequences = 10
- c. Sequence\_length = 30

#### **2. Waktu pengambilan keypoints:**

- a. Min\_detection\_confidence = 0.4
- b. Min\_tracking\_confidence = 0.4
- c. Counter tiap video maju = 6 frames

#### **3. Splitting data:**

- a. Train test split = 70% 30%

#### **4. Model:**

- a. LSTM layer 1 = 64 nodes, Activation = relu
- b. Dense layer 2 = 64 nodes, Activation = relu
- c. Dense layer 3: activation = softmax
- d. Compile: optimizer = adam, loss = categorical\_crossentropy, metrics = categorical\_accuracy
- e. Hasil training: epoch 127 = 0.99

```

216/216 [=====] - 2s /ms/step - loss: 0.1641 - categorical_accuracy: 0.9389
Epoch 121/2000
216/216 [=====] - 1s 7ms/step - loss: 0.1249 - categorical_accuracy: 0.9559
Epoch 122/2000
216/216 [=====] - 1s 7ms/step - loss: 0.0851 - categorical_accuracy: 0.9717
Epoch 123/2000
216/216 [=====] - 2s 7ms/step - loss: 0.0711 - categorical_accuracy: 0.9786
Epoch 124/2000
216/216 [=====] - 1s 7ms/step - loss: 0.3754 - categorical_accuracy: 0.8971
Epoch 125/2000
216/216 [=====] - 1s 7ms/step - loss: 0.0757 - categorical_accuracy: 0.9781
Epoch 126/2000
216/216 [=====] - 2s 7ms/step - loss: 0.0812 - categorical_accuracy: 0.9737
Epoch 127/2000
9/216 [>.....] - ETA: 1s - loss: 0.0558 - categorical_accuracy: 0.99

```

```

-----
KeyboardInterrupt                                Traceback (most recent call last)
<ipython-input-28-610fed43f9e2> in <module>

```

f. Hasil testing: 0.9249

```
accuracy_score(ytrue, yhat)
```

```
0.9249239093676023
```

## 5. Save Model:

- a. Nama file data keypoints = MP\_Data21VidCleanKananKiri
- b. Nama file model = model21VidCleanKananKirilayer.h5

## 6. Percobaan RealTime:

- a. Threshold = 0.9
- b. Min\_detection\_confidence = 0.8
- c. Min\_tracking\_confidence = 0.8

## 7. Kesimpulan:

- a. Mediapipe masih kurang bisa deteksi apabila posisi tangan kebawah
- b. Mediapipe SERING mendeteksi tangan walaupun tidak ada tangan di dalam frame
- c. Masih terjadi BANYAK false prediction, terutama waktu tangan sedang gerak tidak teratur (gabung dengan poin b)
- d. V tidak dapat dideteksi (selalu keluar U) di tangan kanan dan kiri
- e. Akurasi tangan kanan jadi berkurang daripada model-model sebelumnya
- f. T tidak bisa dideteksi di tangan kanan

# Progress Report 5

### 1. Variabel awal:

- a. Total Video = 21
- b. No\_sequences = 10
- c. Sequence\_length = 30

### 2. Waktu pengambilan keypoints:

- a. Mengambil tangan kanan kiri
- b. Min\_detection\_confidence = 0.8
- c. Min\_tracking\_confidence = 0.8
- d. Counter tiap video maju = 10 frames

### 3. Splitting data:

- b. Train test split = 70% 30%

### 4. Model:

- a. LSTM layer 1 = 64 nodes, Activation = relu
- b. Dense layer 2 = 64 nodes, Activation = relu
- c. Dense layer 3: activation = softmax
- d. Compile: optimizer = adam, loss = categorical\_crossentropy, metrics = categorical\_accuracy
- e. Hasil training: epoch 127 = 0.99

```
216/216 [=====] - 2s /ms/step - loss: 0.1641 - categorical_accuracy: 0.9389
Epoch 121/2000
216/216 [=====] - 1s 7ms/step - loss: 0.1249 - categorical_accuracy: 0.9559
Epoch 122/2000
216/216 [=====] - 1s 7ms/step - loss: 0.0851 - categorical_accuracy: 0.9717
Epoch 123/2000
216/216 [=====] - 2s 7ms/step - loss: 0.0711 - categorical_accuracy: 0.9786
Epoch 124/2000
216/216 [=====] - 1s 7ms/step - loss: 0.3754 - categorical_accuracy: 0.8971
Epoch 125/2000
216/216 [=====] - 1s 7ms/step - loss: 0.0757 - categorical_accuracy: 0.9781
Epoch 126/2000
216/216 [=====] - 2s 7ms/step - loss: 0.0812 - categorical_accuracy: 0.9737
Epoch 127/2000
9/216 [>.....] - ETA: 1s - loss: 0.0558 - categorical_accuracy: 0.99
```

```
KeyboardInterrupt                                Traceback (most recent call last)
<ipython-input-28-610fed43f9e2> in <module>
```

- f. Hasil testing: 0.9249

```
accuracy_score(ytrue, yhat)
```

```
0.9249239093676023
```

## **5. Save Model:**

- a. Nama file data keypoints = MP\_DataPR5
- b. Nama file model = model21VidCleanKananKirilayer.h5

## **6. Percobaan RealTime:**

- a. Threshold = 0.9
- b. Min\_detection\_confidence = 0.8
- c. Min\_tracking\_confidence = 0.8

## **7. Kesimpulan:**

- a. Mediapipe masih kurang bisa deteksi apabila posisi tangan kebawah
- b. Mediapipe SERING mendeteksi tangan walaupun tidak ada tangan di dalam frame
- c. Masih terjadi BANYAK false prediction, terutama waktu tangan sedang gerak tidak teratur (gabung dengan poin b)
- d. V tidak dapat dideteksi (selalu keluar U) di tangan kanan dan kiri
- e. Akurasi tangan kanan jadi berkurang daripada model-model sebelumnya
- f. T tidak bisa dideteksi di tangan kanan

# **Progress Report 6**

**Tanggal : 19 Juni 2022**

## **1. Preprocessing dataset manual:**

- a. Menghapus frame-frame video dataset yang tidak penting (gerakan tangan yang bukan abjad, frame tanpa tangan di layer, ada beberapa video abjad yang abjad setelahnya kelihatan didalam frame (contoh: abjad A awalnya tapi frame belakang-belakang ada menunjukan abjad B sepersekian detik))

## **2. Variabel awal:**

- a. Total Video = 21
- b. No\_sequences = 10
- c. Sequence\_length = 30

## **3. Waktu pengambilan keypoints:**

- a. Belum mengambil tangan kiri

- b. Min\_detection\_confidence = 0.8
- c. Min\_tracking\_confidence = 0.8
- d. Counter tiap video maju = 6 frames

### 1. Splitting data:

- a. Train test split = 70% 30%

### 2. Model:

- a. LSTM layer 1 = 64 nodes, Activation = relu
- b. LSTM layer 2 = 128 nodes, Activation = relu
- c. Dense layer 1 = 128 nodes, Activation = relu
- d. Dense layer 2 = 64 nodes, Activation = relu
- e. Dense layer 3: activation = softmax
- f. Compile: optimizer = adam, loss = categorical\_crossentropy, metrics = categorical\_accuracy
- g. Hasil training: epoch 507 = 0.93

```
Epoch 500/2000
79/79 [=====] - 1s 18ms/step - loss: 0.2110 - categorical_accuracy: 0.9361
Epoch 501/2000
79/79 [=====] - 1s 18ms/step - loss: 0.2035 - categorical_accuracy: 0.9362
Epoch 502/2000
79/79 [=====] - 1s 19ms/step - loss: 0.2003 - categorical_accuracy: 0.9384
Epoch 503/2000
79/79 [=====] - 1s 18ms/step - loss: 0.2307 - categorical_accuracy: 0.9270
Epoch 504/2000
79/79 [=====] - 1s 18ms/step - loss: 0.1965 - categorical_accuracy: 0.9387
Epoch 505/2000
79/79 [=====] - 1s 18ms/step - loss: 0.1873 - categorical_accuracy: 0.9432
Epoch 506/2000
79/79 [=====] - 1s 18ms/step - loss: 0.1891 - categorical_accuracy: 0.9422
Epoch 507/2000
55/79 [=====>.....] - ETA: 0s - loss: 0.1952 - categorical_accuracy: 0.9382
```

```
-----
KeyboardInterrupt                                Traceback (most recent call last)
<ipython-input-22-610fed43f9e2> in <module>
      1 model.fit(X_train, y_train, epochs=2000, callbacks=[th_callback])
```

- h. Hasil testing: 0.9

```
1]: accuracy_score(ytrue, yhat)
1]: 0.9021237303785781
```

### 3. Save Model:

- a. Nama file data keypoints = MP\_PR6
- b. Nama file model = modelPR6.h5

**4. Percobaan RealTime:**

- a. Threshold = 0.9
- b. Min\_detection\_confidence = 0.8
- c. Min\_tracking\_confidence = 0.8

## **Progress Report 7**

**Tanggal : 19 Juni 2022**

**4. Preprocessing dataset manual:**

- a. Menghapus frame-frame video dataset yang tidak penting (gerakan tangan yang bukan abjad, frame tanpa tangan di layer, ada beberapa video abjad yang abjad setelahnya kelihatan didalam frame (contoh: abjad A awalnya tapi frame belakang-belakang ada menunjukan abjad B sepersekian detik))

**5. Variabel awal:**

- a. Total Video = 21
- b. No\_sequences = 10
- c. Sequence\_length = 30

**6. Waktu pengambilan keypoints:**

- a. Belum mengambil tangan kiri
- b. Min\_detection\_confidence = 0.8
- c. Min\_tracking\_confidence = 0.8
- d. Counter tiap video maju = 6 frames

**5. Splitting data:**

- b. Train test split = 70% 30%

**6. Model:**

- a. LSTM layer 1 = 64 nodes, Activation = relu
- b. Dense layer 2 = 64 nodes, Activation = relu
- c. Dense layer 3: activation = softmax
- d. Compile: optimizer = adam, loss = categorical\_crossentropy, metrics = categorical\_accuracy

- e. Hasil training: epoch 507 = 0.92 (digambar hanya 13 epoch karena tidak sengaja ke fit lagi.)

```

109/109 [=====] - 1s /ms/step - loss: 0.2507 - categorical_accuracy: 0.9237
Epoch 8/2000
109/109 [=====] - 1s 6ms/step - loss: 0.2894 - categorical_accuracy: 0.9101
Epoch 9/2000
109/109 [=====] - 1s 6ms/step - loss: 0.3228 - categorical_accuracy: 0.9034
Epoch 10/2000
109/109 [=====] - 1s 6ms/step - loss: 0.2902 - categorical_accuracy: 0.9115
Epoch 11/2000
109/109 [=====] - 1s 6ms/step - loss: 0.2924 - categorical_accuracy: 0.9135
Epoch 12/2000
109/109 [=====] - 1s 6ms/step - loss: 0.2338 - categorical_accuracy: 0.9297
Epoch 13/2000
49/109 [=====>.....] - ETA: 0s - loss: 0.2350 - categorical_accuracy: 0.9298

-----
KeyboardInterrupt                                Traceback (most recent call last)
<ipython-input-51-610fed43f9e2> in <module>

```

- f. Hasil testing: 0.89

```

: accuracy_score(ytrue, yhat)
: 0.8906882591093117

```

## 7. Save Model:

- Nama file data keypoints = MP\_PR6
- Nama file model = modelPR6\_2.h5

## 8. Percobaan RealTime:

- Threshold = 0.95
- Min\_detection\_confidence = 0.8
- Min\_tracking\_confidence = 0.8