

Sistem Transliterasi Dokumen Aksara Jawa Menjadi Tulisan Romawi Menggunakan Metode Convolutional Neural Network

Theofilus Arifin





Pendahuluan

Bahasa dan Aksara Jawa

■ 3 Provinsi

Jawa Timur, Jawa Tengah, dan DIY

■ > 75 Juta

Orang merupakan penutur bahasa Jawa

■ < 1 %

Orang yang mampu membaca aksara Jawa



Rumusan Masalah



“Bagaimana cara membantu masyarakat untuk mengubah tulisan aksara Jawa pada suatu dokumen menjadi aksara Latin?”

Tujuan & Manfaat



Tujuan

Membuat sistem transliterasi tulisan aksara Jawa menjadi aksara Latin pada suatu dokumen



Manfaat

- Peneliti aksara Jawa
- Akademisi aksara Jawa
- Penerjemah aksara Jawa

Ruang Lingkup

- **Dataset** pengumpulan sendiri
- 20 carakan + 16 pasangan + 11 sandhangan
- **Aplikasi** berbasis website
- **Input** citra dokumen aksara jawa (3:4)
- **Output** berupa aksara Latin hasil transliterasi





Tinjauan Pustaka

Aksara Jawa

- Sistem penulisan masyarakat Jawa
- Tahun 1500 hingga sekarang
- Surat, naskah, dan piagam
- 20 Carakan
- 20 Pasangan
- 11 Sandhangan

Carakan Dan Pasangan				
Ha	Na	Ca	Ra	Ka
Da	Ta	Sa	Wa	La
Pa	Dha	Ja	Ya	Nya
Ma	Ga	Ba	Tha	Nga

Sandhangan		
Wulu	= i	
Suku	= u	
Taling	= e	
Taling Tarung	= o	
Pepet	= é	
Layar	= r (mati)	
Wignyan	= h (mati)	
Cecak	= ng (mati)	
Pangkun	= paten	
Pada Lingsa	= koma (.)	
Pada Lingsa	= titik	

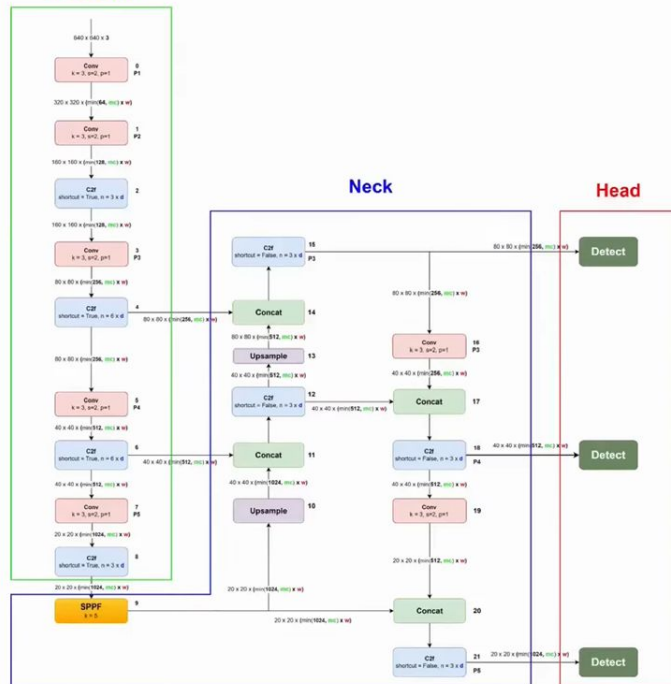
Image Preprocessing

1. Grayscaleing
2. Thresholding
3. Dilation
4. Contour Filtering
5. Projection Profile



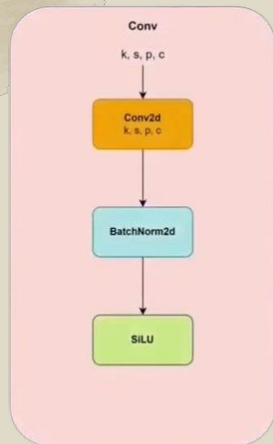
YOLOv8

Backbone

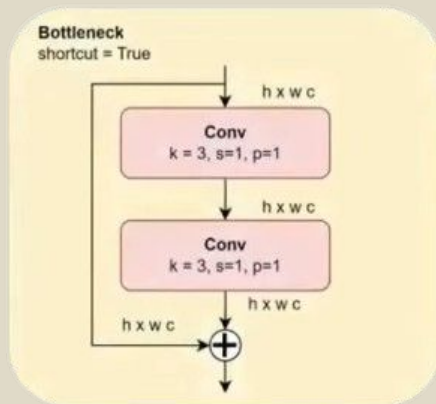


Model variant	d (depth_multiple)	w (width_multiple)	mc (max_channels)
n	0.33	0.25	1024
s	0.33	0.50	1024
m	0.67	0.75	768
l	1.00	1.00	512
xl	1.00	1.25	512

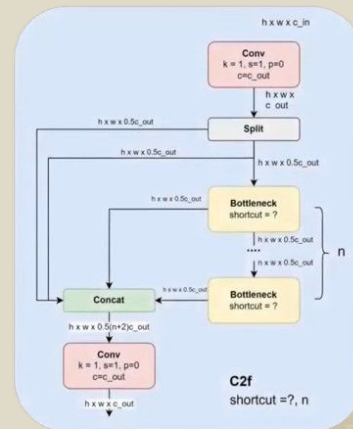
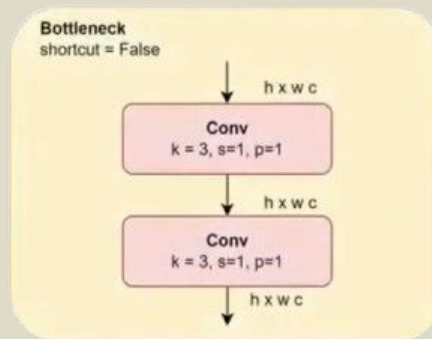
YOLOv8



Convolutional
Block

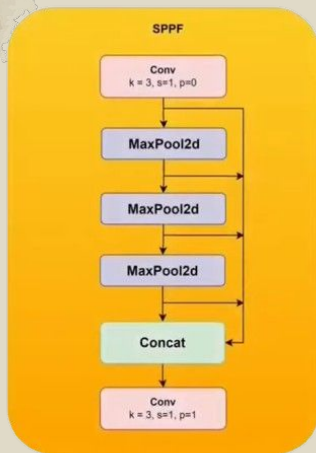


Bottleneck
Block

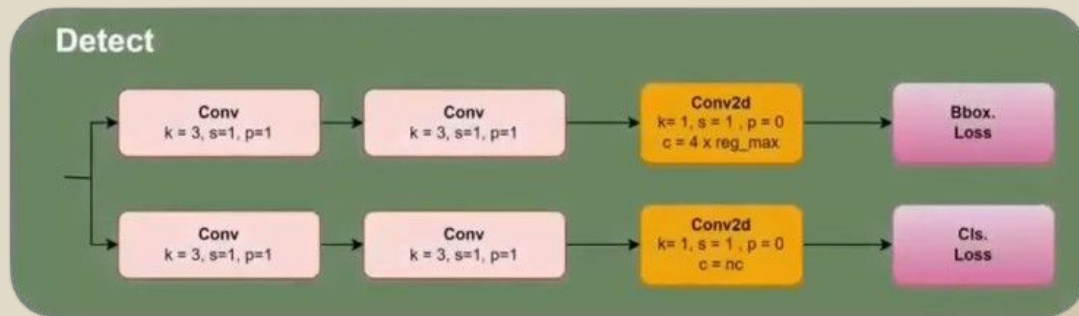


C2f
Block

YOLOv8



SPPF
Block



Detect
Block

Evaluasi Model

IoU

$$\text{IoU} = \frac{\text{Area (ground truth} \cap \text{predicted)}}{\text{Area (ground truth} \cup \text{predicted)}}$$

Precision

$$P = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

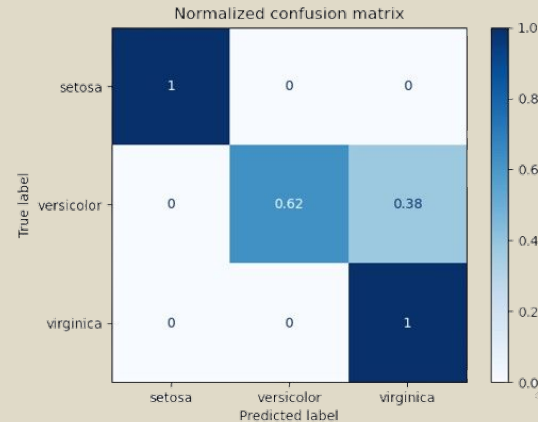
Recall

$$R = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

mAP

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$

Confusion Matrix



Library

Roboflow

Download Dataset

Ultralytics

Pembuatan model
YOLOv8

OpenCV

Pembacaan citra dan
preprocessing

Numpy

Array dan
operasi matematika

Streamlit

Python web framework

Pandas

Manipulasi data hasil
anotasi





Analysis

Kondisi Saat Ini

Kuesioner

40 Responden

Transliterasi Dokumen Aksara
Jawa ke Aksara Latin

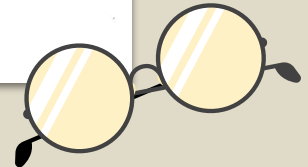
သၢၤတၢ်ဆၢၤတၢ်တၢ်တၢ်

တၢ်တၢ်တၢ်တၢ်တၢ်တၢ်တၢ်

သၢၤတၢ်ဆၢၤတၢ်တၢ်တၢ်

သၢၤတၢ်ဆၢၤတၢ်တၢ်တၢ်

သၢၤတၢ်ဆၢၤတၢ်တၢ်တၢ်



Kondisi Saat Ini

67.4%

Rata-rata akurasi responden untuk tiap baris

115 detik

Waktu untuk transliterasi tiap baris dokumen

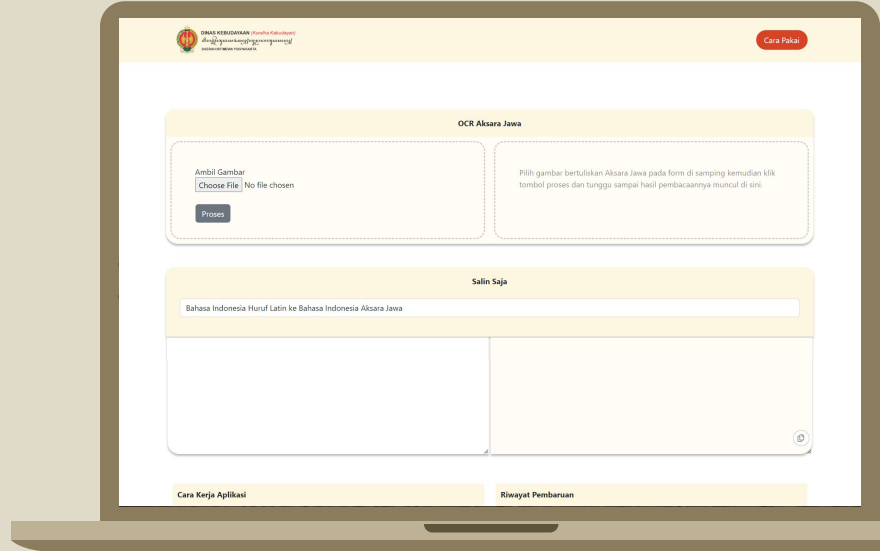
67.5%

mencari panduan melalui search engine
dan kemudian mencocokkan



Sistem Sejenis

Website **kongresaksarajawa.id** adalah sebuah website aplikasi transliterasi aksara Jawa yang dikembangkan oleh Kongres Aksara Jawa.



Penelitian Sejenis

1

Soebroto (2022)

- Akurasi Akhir 90.47%
- Menggunakan Arsitektur MobileNet
- Berfokus pada pembacaan kata saja



Ilham dan
Rochmawati (2018)

2

- Akurasi Akhir 94.23%
- Menggunakan Arsitektur menyerupai AlexNet
- Berfokus pada pembacaan kata saja

Masalah



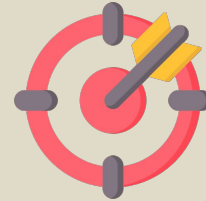
Proses

Dominan melakukan pencocokan dengan *search engine* secara berulang



Waktu

115 detik untuk transliterasi tiap baris pada dokumen



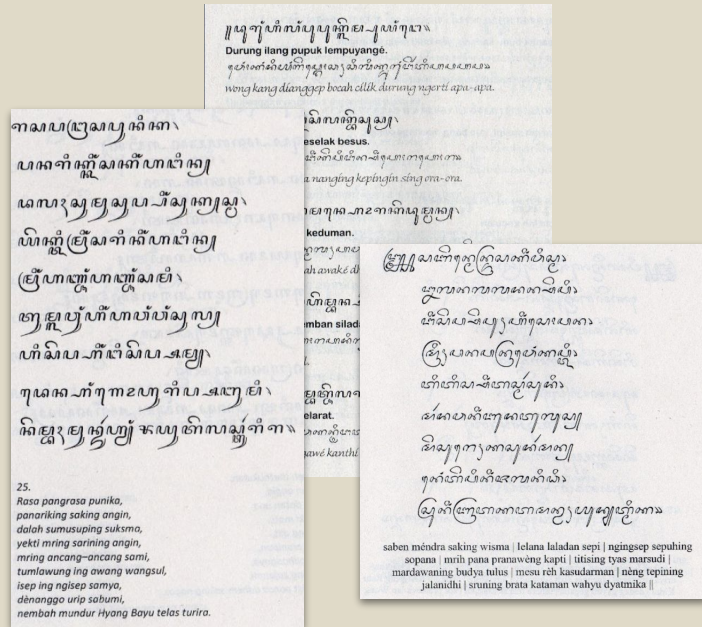
Akurasi

Rata-rata 67% untuk tiap baris pada dokumen



Desain dan Implementasi

Pengumpulan Dataset



164 Gambar (80, 10, 10)

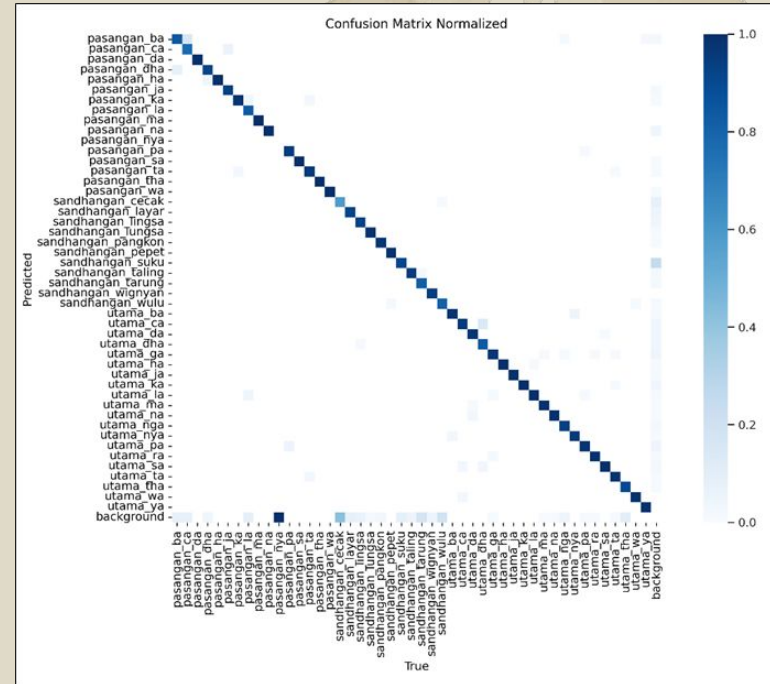


Pembuatan Model

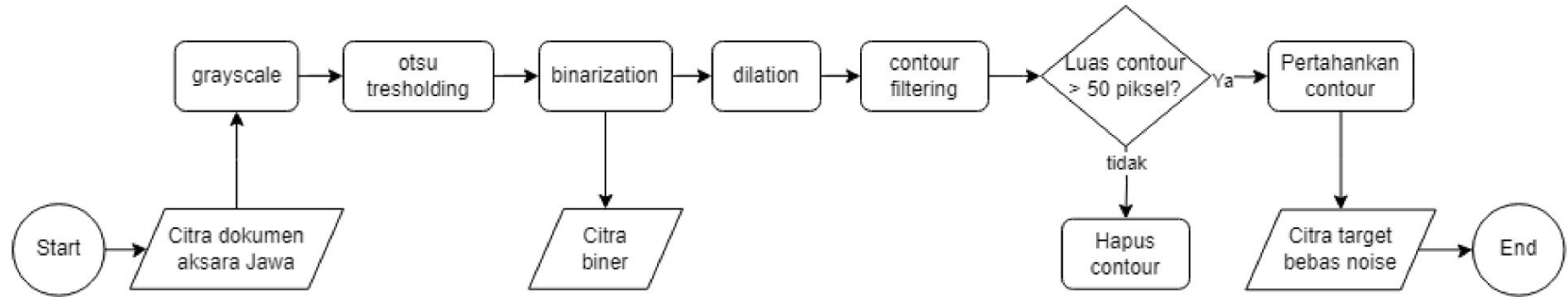
YOLOv8m

Ukuran Model 49.7 MB

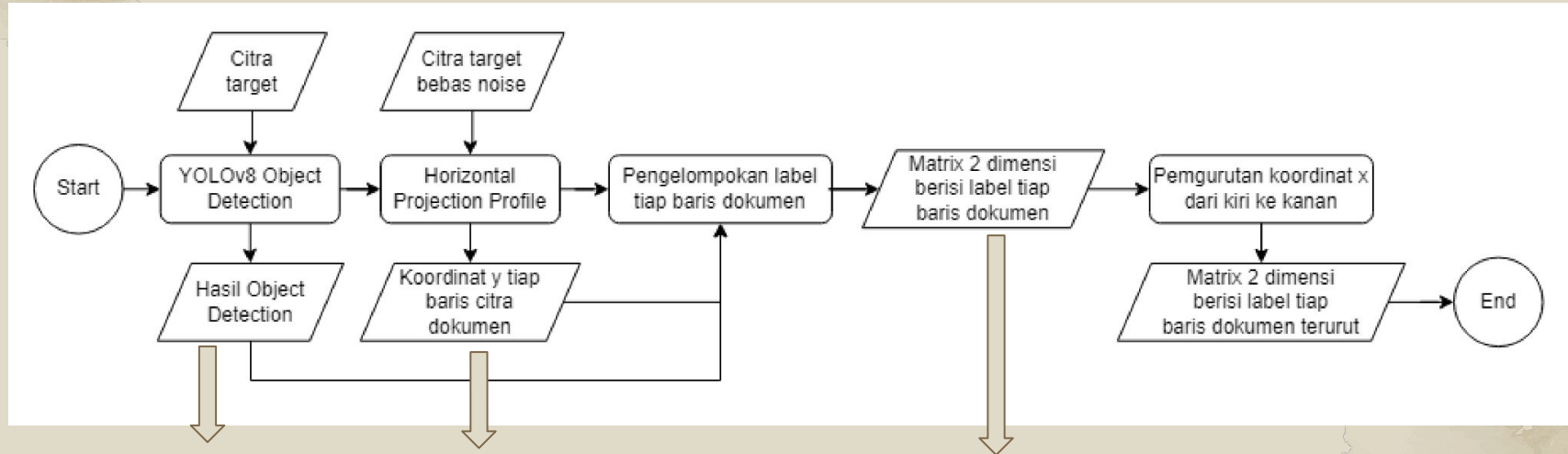
- Precision : 0.966
- Recall : 0.924
- mAP50 : 0.961
- mAP50-95 : 0.746



Proses Preprocessing



Proses Object Detection



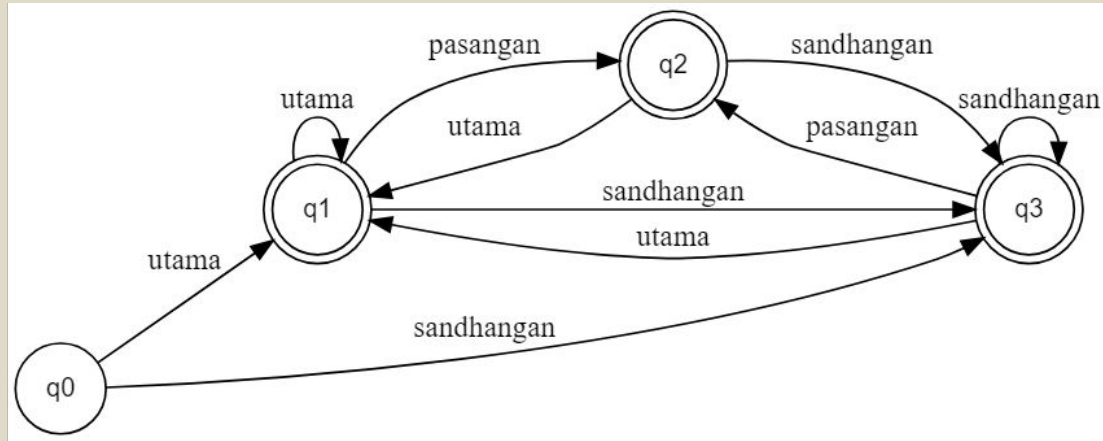
```
23 0.197869 0.473336 0.0310574 0.0325987
38 0.11397 0.322873 0.0518752 0.0174385
23 0.23787 0.0470776 0.0318226 0.0332257
```

Coordinates: (78, 169)

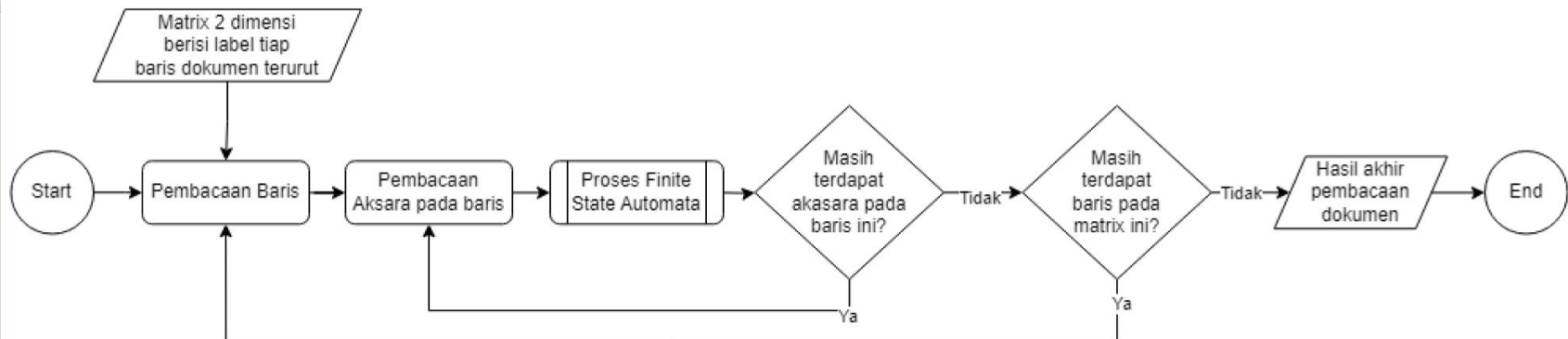
Coordinates: (221, 315)

```
['utama_ka', 'sandhangan_pepet', 'utama_la', 'utama_ca'],
['sandhangan_taling', 'utama_ha', 'sandhangan_tarung', 'utama_ra']]
```

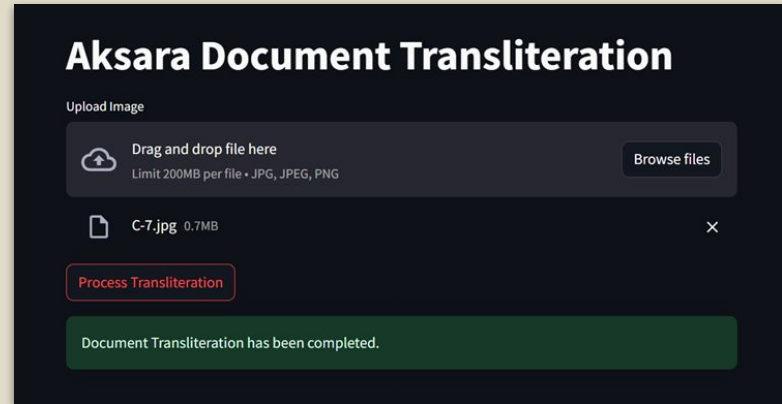
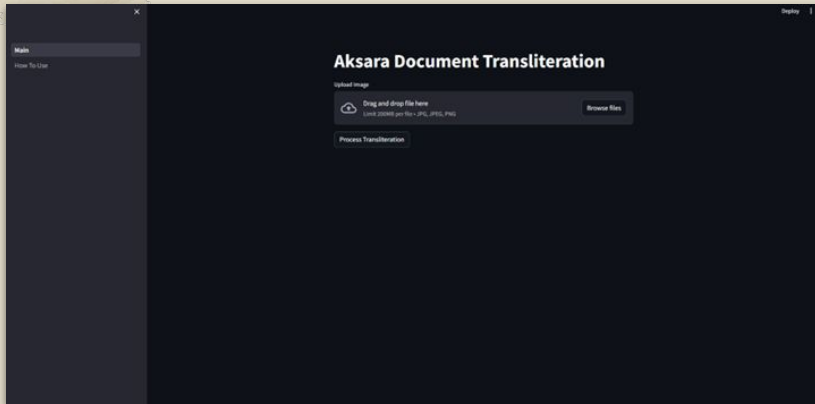
Finite State Automata



Proses Transliterasi



Website Aplikasi



[illegible]



Evaluasi

Validasi



Proses

Dominan melakukan pencocokan dengan *search engine* secara berulang



Transliterasi dokumen dalam satu aplikasi



Waktu

115 detik untuk tiap baris pada dokumen



Rata-rata 5 detik
20 kali lipat Lebih cepat



Akurasi

Rata-rata 67% untuk tiap baris pada dokumen



Rata-rata 75%
(mAP model YOLOv8)



Kesimpulan

Kesimpulan

Sistem transliterasi dibuat dengan pembacaan karakter menggunakan object detection melalui model YOLOv8 dan transliterasi menggunakan automata melalui finite state automata.

Sistem berhasil memenuhi tujuan utama sesuai dengan batasan yang telah ditentukan



Saran

- Memperluas dataset dengan menggunakan jenis aksara lain maupun jenis font aksara Jawa lain
- Mengembangkan metode untuk mengatasi *class imbalance*



Terima Kasih



CREDITS: This presentation template was
created by **Slidesgo**, including icons by
Flaticon, infographics & images by **Freepik**