

**Pengembangan Sistem Pengenalan dan Penerjemahan Aksara
Jawa Menggunakan CNN dan Transformer**

Riset Informatika (D081)



Di Susun Oleh :

Muhamad Fauzi 21081010233

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
SURABAYA
2024**

BAB I

PENDAHULUAN

1.1 Latar Belakang

Aksara Jawa, yang juga dikenal sebagai Hanacaraka, merupakan salah satu warisan budaya Indonesia yang kaya dan memiliki nilai historis yang tinggi. Namun, seiring perkembangan zaman, penggunaan aksara Jawa semakin berkurang dan cenderung tergantikan oleh aksara Latin. Dalam upaya melestarikan dan mempermudah pembelajaran aksara Jawa, diperlukan suatu sistem yang mampu mengenali dan menerjemahkan aksara Jawa secara otomatis.

Teknologi pengenalan karakter telah berkembang pesat dengan adanya algoritma berbasis pembelajaran mendalam (deep learning) seperti Convolutional Neural Network (CNN) dan Transformer. CNN terkenal dalam pengolahan data berbentuk gambar, sedangkan Transformer unggul dalam pemrosesan urutan data, termasuk teks. Kombinasi keduanya memberikan peluang untuk mengembangkan sistem yang efektif dalam mengenali aksara Jawa dan menerjemahkannya ke dalam teks Latin atau bahasa lainnya.

Penelitian ini bertujuan untuk mengembangkan sistem pengenalan dan penerjemahan aksara Jawa menggunakan metode CNN dan Transformer. Sistem ini diharapkan dapat membantu melestarikan aksara Jawa sekaligus memberikan kontribusi dalam bidang pengolahan bahasa alami dan pengenalan pola.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang diajukan dalam penelitian ini adalah:

1. Bagaimana merancang dan mengimplementasikan sistem pengenalan aksara Jawa menggunakan CNN?
2. Bagaimana mengintegrasikan model Transformer untuk menerjemahkan hasil pengenalan aksara Jawa ke dalam aksara Latin?
3. Sejauh mana sistem yang dikembangkan dapat mencapai tingkat akurasi yang tinggi dalam pengenalan dan penerjemahan aksara Jawa?

1.3 Batasan Masalah

Agar penelitian ini terfokus, batasan masalah yang diterapkan adalah sebagai berikut:

1. Dataset yang digunakan hanya mencakup gambar aksara Jawa (Hanacaraka) dalam format standar dan kombinasi aksara Jawa dan sandhangan
2. Model CNN digunakan untuk pengenalan aksara pada level karakter.
3. Model Transformer digunakan untuk penerjemahan teks aksara Jawa ke aksara Latin.
4. Penelitian ini tidak mencakup aksara lain selain aksara Jawa.

1.4 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah :

1. Mengembangkan sistem pengenalan aksara Jawa menggunakan metode CNN.
2. Mengembangkan modul penerjemahan aksara Jawa ke aksara Latin menggunakan Transformer.
3. Mengevaluasi kinerja sistem berdasarkan akurasi pengenalan dan penerjemahan.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

- Penelitian ini dapat menjadi referensi bagi pengembangan teknologi pengenalan karakter berbasis deep learning, khususnya dalam pengolahan aksara tradisional
- Membantu melestarikan aksara Jawa dengan menyediakan alat digital yang dapat mengenali dan menerjemahkan aksara tersebut.
- Mempermudah proses pembelajaran aksara Jawa bagi masyarakat umum.
- Mendorong inovasi teknologi dalam pengolahan bahasa dan pelestarian budaya.

BAB II

TINJAUAN PUSTAKA

2.1 Pengenalan Aksara Jawa

Aksara Jawa adalah salah satu aksara tradisional di Indonesia yang memiliki bentuk unik dan kompleks. Pengenalan aksara Jawa secara otomatis membutuhkan pendekatan yang dapat menangani variasi bentuk aksara, termasuk aksara dasar, pasangan, dan sandhangan. Beberapa penelitian sebelumnya menggunakan metode klasifikasi berbasis fitur untuk mengenali aksara Jawa, namun pendekatan ini memiliki keterbatasan pada data yang bervariasi.

2.2 Convolutional Neural Network (CNN)

CNN adalah salah satu algoritma deep learning yang dirancang untuk memproses data berbentuk grid, seperti gambar. CNN bekerja dengan mengekstraksi fitur dari gambar melalui lapisan konvolusi dan pooling. CNN telah terbukti efektif untuk tugas-tugas pengenalan citra, termasuk pengenalan karakter tulisan tangan. Penelitian seperti [Nama Peneliti, Tahun] menunjukkan bahwa CNN mampu mencapai akurasi tinggi dalam klasifikasi aksara Jawa.

2.3 Transformer

Transformer adalah arsitektur model yang banyak digunakan dalam pemrosesan bahasa alami (NLP). Model ini diperkenalkan oleh Vaswani et al. (2017) dan dikenal karena mekanisme attention-nya yang memungkinkan model memahami hubungan antar elemen dalam urutan data. Transformer juga telah diadaptasi untuk tugas transliterasi dan penerjemahan teks, menunjukkan hasil yang unggul dibandingkan pendekatan sebelumnya.

2.4 Penelitian Terkait

Beberapa penelitian juga telah dilakukan mengenai pengenalan aksara Jawa

1. Penelitian oleh Rahmawati et al. (2023) menggunakan metode **Convolutional Neural Network (CNN)** untuk mengklasifikasikan aksara Jawa secara otomatis. Penelitian ini mencapai akurasi sebesar **97,42%**. Hasilnya adalah sebuah aplikasi berbasis web yang mampu memprediksi aksara Jawa dengan benar
2. Penelitian oleh Jonathan dan Wasito (2023) menggunakan algoritma Convolutional Neural Network (CNN) untuk mengenali aksara Jawa tulisan tangan digital. Dataset yang digunakan terdiri dari 30 kelas karakter aksara Jawa dengan total 1.050 gambar. Penelitian ini mencapai akurasi sebesar 84% dalam pengenalan gambar aksara Jawa tulisan tangan digital. Selain itu, penelitian ini juga memanfaatkan teknologi Computer Vision untuk melakukan klasifikasi gambar dengan input berupa webcam, meskipun hasil pengenalan aksara tulisan tangan murni masih kurang optimal.

BAB III

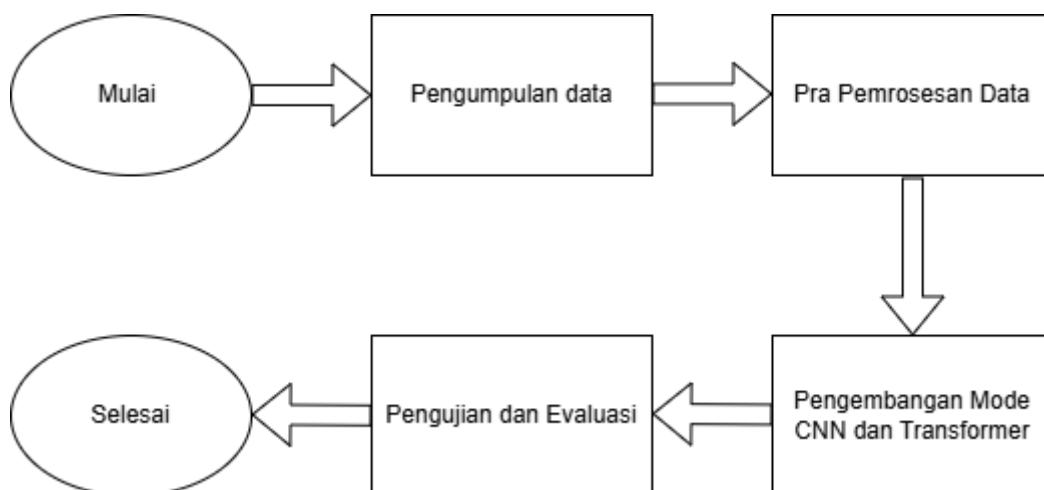
METODOLOGI PENELITIAN

3.1 Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimental untuk mengembangkan dan mengevaluasi sistem pengenalan dan penerjemahan aksara Jawa. Model CNN digunakan untuk pengenalan pola dalam gambar aksara Jawa, sedangkan Transformer digunakan untuk menerjemahkan hasil pengenalan ke aksara Latin. Tahapan penelitian meliputi pengumpulan data, preprocessing, implementasi model, dan evaluasi kinerja sistem.

3.2 Tahapan Penelitian

Untuk melakukan sebuah penelitian, diperlukan adanya tahapan-tahapan yang tersusun dengan baik dan sistematis agar pelaksanaan penelitian tepat mencapai tujuan yang diharapkan. Tahapan penelitian tersebut dapat dilihat digambar berikut :



Gambar 1. Tahapan Penelitian

Penjelasan tahapan penelitian pada gambar diatas adalah sebagai berikut :

- **Pengumpulan Dataset:** Mengumpulkan dataset gambar aksara Jawa yang didapat dari kaggle, yaitu Aksara Jawa/Hanacaraka yang mencakup aksara jawa dasar serta dataset yang dibuat sendiri terdiri dari kombinasi aksara jawa dan sandhangan dalam berbagai variasi gaya dan kondisi.
- **Preprocessing Data:** Melakukan resize dan normalisasi ukuran gambar agar seragam,
- **Pengembangan Model:**
 - Melatih CNN untuk pengenalan karakter aksara Jawa.
 - Menggunakan Transformer untuk penerjemahan hasil pengenalan ke aksara Latin.
- **Pengujian dan Evaluasi:**
 - Mengevaluasi model berdasarkan metrik kinerja seperti akurasi, bleu score dan CER atau WER

3.3 Arsitektur Sistem

1. **CNN untuk Pengenalan Aksara Jawa:**
 - CNN digunakan untuk mengenali karakter aksara Jawa dari gambar.
 - Input: Gambar aksara Jawa.
 - Output: Teks aksara Jawa digital.
2. **Transformer untuk Penerjemahan:**
 - Transformer digunakan untuk menerjemahkan teks aksara Jawa ke teks Latin.
 - Input: Teks aksara Jawa digital.
 - Output: Teks Latin hasil terjemahan.

3.4 Evaluasi Sistem

Model nantinya akan di evaluasi dengan metrik-metrik pengujian sebagai berikut :

1. **Akurasi CNN:**

Mengukur keberhasilan model CNN dalam mengenali karakter aksara Jawa dengan membandingkan hasil prediksi dengan label sebenarnya.

2. **BLEU Score untuk Transformer:**

Mengukur kualitas terjemahan teks Latin dibandingkan dengan referensi terjemahan.

3. **Character Error Rate (CER) dan Word Error Rate (WER):**

Mengukur tingkat kesalahan karakter dan kata dalam hasil terjemahan teks Latin.

Hasil evaluasi digunakan untuk menilai efektivitas sistem yang dikembangkan dan memberikan rekomendasi untuk perbaikan di masa depan.