

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Indonesia merupakan negara yang kaya akan budaya dan tradisi yang unik. Menurut Statistik Kebudayaan tahun 2023 yang dikeluarkan oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, jumlah warisan budaya tak benda di Indonesia adalah sebanyak 1.941[1]. Dari sekian banyak jumlah warisan budaya tak benda tersebut, aksara daerah merupakan salah satu budaya sastra yang beraneka ragam karena berasal dari berbagai suku yang ada di Indonesia. Aksara daerah merupakan sistem ortografi hasil masyarakat daerah yang meliputi aksara dan sistem pengaksaraan untuk menuliskan bahasa daerah[2]. Aksara Sunda termasuk ke dalam kategori Aksara dan Naskah Ka Ga Nga yang diakui sebagai warisan budaya tak benda pada tahun 2013[1].

Upaya pelestarian aksara daerah sudah marak dilakukan oleh berbagai lapisan masyarakat. Salah satu upaya melestarikan Aksara Sunda, Pemerintah daerah provinsi Jawa Barat (Jabar) mengeluarkan Peraturan Daerah (Perda) Provinsi Jabar nomor 5 tahun 2003, lalu dirubah oleh Perda Provinsi Jabar nomor 14 tahun 2014 tentang Pemeliharaan Bahasa, Sastra, dan Aksara Daerah[2]. Upaya lainnya adalah dari kalangan civitas akademik, khususnya bidang informasi dan teknologi, yang sudah banyak melakukan penelitian untuk melestarikan aksara daerah menggunakan media digital.

Banyak penelitian dengan fokus *Optical Character Recognition* (OCR) melalui metode *Convolutional Neural Network* (CNN) sudah dilakukan sebagai bentuk proses digitasi Aksara Sunda. Penelitian sebelumnya oleh Hayati K, et al[3] dilakukan eksplorasi metode CNN klasik untuk OCR Aksara Sunda. Fitur pada Aksara Sunda diekstraksi menggunakan *Histogram of Oriented Gradients* (HOG). Penelitian ini membandingkan metode *Support Vector*

Machine (SVM), *K-Nearest Neighbors* (*k*-NN), *Decision Tree*, *Naïve Bayes*, *Random Forest*, dan *Artificial Neural Network* (ANN) dalam proses OCR. Penelitian lainnya oleh Prameswari M, et al[4] dengan fokus *transfer learning* berbasis CNN membandingkan antara arsitektur ResNet-50, arsitektur VGG-19, dan arsitektur MobileNet. Penelitian ini dilakukan menggunakan 2.560 *training data* dan 640 *validation data*. Penelitian lainnya dilakukan oleh Akram R, et al[5] yang bertujuan untuk mencari *activation function* paling efektif dalam CNN untuk OCR Aksara Sunda. Penelitian ini membandingkan beberapa *activation function* dari *Rectified Linear Unit* (ReLU), yaitu *Leaky ReLU*, *Randomized Leaky ReLU* (RLReLU), dan *Optimized Leaky ReLU* (OLReLU). Penelitian ini menunjukkan bahwa OLReLU merupakan *activation function* terbaik dengan nilai akurasi 0.987.

Dilihat dari penelitian-penelitian diatas, penelitian yang sudah dilakukan masih berfokus pada klasifikasi karakter. Hal ini menunjukkan masih sedikit penelitian yang berfokus pada klasifikasi kata untuk OCR Aksara Sunda. Salah satu dari sedikitnya penelitian yang ada tentang OCR Aksara Sunda untuk klasifikasi kata dilakukan oleh Gerhana Yana, et al[6] yang menggunakan metode *Template Matching Algorithm* dan *Feature Extraction Algorithm*. Performa model pada penelitian ini untuk klasifikasi kata dari tulisan tangan secara spesifik mencapai akurasi dengan nilai 75% dari 30 sample kata. Oleh karena itu, diperlukan penelitian lebih lanjut pada pengklasifikasian kata sebagai langkah pengembangan berikutnya dalam OCR khususnya tulisan Aksara Sunda.

1.2. Rumusan Masalah

Permasalahan yang diangkat pada penelitian ini adalah masih sedikitnya penelitian tentang sistem *Optical Character Recognition* (OCR) pada tingkat kata, khususnya OCR Aksara Sunda. Kriteria yang menjadi fokus dalam permasalahan ini adalah menghasilkan model klasifikasi kata tulisan tangan Aksara Sunda tanpa menggunakan model segmentasi. Rumusan masalah diatas dapat dirinci menjadi sebagai berikut:

1. Bagaimana penerapan model berbasis EfficientNet untuk klasifikasi dalam *Optical Character Recognition* tingkat kata tanpa segmentasi pada tulisan tangan Aksara Sunda?
2. Bagaimana performa model yang dihasilkan dari penerapan model berbasis EfficientNet pada model klasifikasi tersebut?

1.3. Tujuan dan Manfaat

Tujuan dari penelitian ini yaitu sebagai berikut:

1. Mengimplementasikan EfficientNet sebagai model dasar pada *Optical Character Recognition* untuk mengenali pola kata tulisan tangan Aksara Sunda.
2. Menganalisis performa EfficientNet sebagai model dasar untuk *Optical Character Recognition* untuk mengenali pola kata Aksara Sunda.

1.4. Batasan Masalah

Batasan masalah pada penelitian Tugas Akhir ini yaitu:

1. Penelitian ini menggunakan dua jenis dataset.
 - a. Dataset karakter yang terdiri dari ngalagena dan swara. Terdiri dari 32 kelas.
 - b. Dataset kata yang terdiri dari gabungan antara ngalagena dan swara dengan rarangkén. Terdiri dari 220 kelas. Dataset kata yang dikumpulkan berasal dari kumpulan kalimat bahasa sunda secara acak. Dataset kata terdiri dari maksimal 3 karakter ngalagena atau swara dengan rarangkén
2. EfficientNet yang digunakan adalah EfficientNetB0.
3. Tidak ada model segmentasi yang digunakan dalam penelitian ini

1.5. Metode Penelitian

Metode dalam penelitian Tugas Akhir ini terdiri dari beberapa tahap yaitu sebagai berikut:

1. Kajian studi

Mencari materi yang bersangkutan, seperti Aksara Sunda dan tata cara penulisannya, *deep learning*, *Convolutional Neural Network (CNN)*, *Recurrent Neural Network (RNN)*, arsitektur EfficientNet dan arsitektur CNN ataupun RNN lainnya.

2. Pengumpulan data

Menggunakan data dengan sumber Direktori Aksara Sunda untuk Unicode, ditambah dengan data kata yang ditulis tangan dalam Aksara Sunda yang dikumpulkan.

3. Perancangan sistem

Sistem yang dirancang pada penelitian ini berupa sistem pendeteksian pola kata Aksara Sunda dengan model dasar berbasis EfficientNet.

4. Implementasi sistem

Implementasi rancangan sistem pada penelitian ini akan menggunakan bahasa pemrograman *python* dibantu dengan *library-library python* yang ada.

5. Evaluasi model

Evaluasi ini akan dilakukan dengan parameter acuan berupa akurasi, presisi, *recall* dan *F1-score*, disertai dengan evaluasi *Word Error Rate (WER)* dan *Character Error Rate (CER)*.

6. Pengambilan kesimpulan

Pengambilan kesimpulan penelitian ini dilakukan berdasarkan hasil dari pengujian sistem dan bertujuan untuk menjawab permasalahan yang diusulkan.