

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Indonesia merupakan negara dengan keberagaman suku dan budaya yang ada didalamnya. Keberagaman suku tentu tidak terlepas dari keberagaman bahasa yang dipakai suku tersebut dan kemudian menjadi bahasa suku atau sering disebut bahasa daerah. Bahasa daerah merupakan bagian penting bagi keberlangsungan suatu suku agar dapat terus menjaga eksistensinya, sehingga bahasa daerah menjadi warisan turun temurun dan wajib dipahami oleh setiap penerus generasi suku itu.

Pada zaman yang sudah serba modern ini, semakin mengubah cara hidup dan berkomunikasi seseorang. Bahasa daerah semakin ditinggalkan karena dianggap kuno dan tidak sesuai dengan pergaulan dalam generasi sekarang ini. Generasi sekarang yang diharapkan dapat melestarikan bahasa daerahnya masing-masing, justru memiliki bahasa tersendiri yang biasa disebut dengan bahasa gaul atau kekinian.

Banyak faktor yang membuat lemahnya kesadaran generasi sekarang ini untuk melestarikan bahasa daerahnya. Salah satunya adalah masih kurangnya sarana yang dapat digunakan untuk mengakses informasi dari bahasa daerah itu sendiri. Adapun akses yang dimaksud adalah agar seseorang dapat mengetahui dengan cepat apa terjemahan dari bahasa Indonesia yang ada menjadi bahasa daerah orang tersebut. Dengan adanya akses ini diharapkan mampu meningkatkan pengetahuan seseorang akan bahasa daerahnya masing-masing.

Dalam perkembangan ilmu pengetahuan saat ini, dapat membuat akses semua orang menjadi serba mudah. Salah satunya dalam masalah ini ilmu pengetahuan yang dapat dimanfaatkan adalah pengolahan citra digital. Pada penelitian ini, pengolahan citra digunakan untuk mengenali karakter dari sebuah gambar / citra teks bahasa Indonesia dan pada akhirnya teks ini dapat diterjemahkan menjadi bahasa daerah.

Metode untuk mengenali karakter pada citra teks biasanya disebut metode *Optical Character Recognition* (OCR). Proses yang dilakukan oleh metode ini umumnya yaitu dimulai dari praproses citra teks yang bertujuan untuk menghilangkan *noise*, biasanya citra ubah dalam bentuk *grayscale* atau biner. Kemudian dilakukan proses segmentasi untuk mengambil karakter dari teks, lalu mengekstrak fitur dari karakter tersebut dan barulah dilakukan pelatihan terhadap fitur ini sehingga dapat mengidentifikasi karakter tadi dan dikenali oleh sistem.

Penelitian terkait metode untuk identifikasi citra teks atau karakter banyak dilakukan menggunakan metode *Neural Network*. *Neural Network* memiliki cara kerja yang sama seperti sel otak manusia ketika memproses informasi, dimana *Neural Network* menyimpan dan menggunakan pengetahuan yang didapat dari pengalaman sehingga sangat cocok digunakan dalam masalah prediksi ataupun masalah klasifikasi yang sedang dibahas dalam penelitian ini. *Neural Network* memiliki beberapa metode yang banyak digunakan untuk klasifikasi karakter, yaitu *Multilayer Perceptrons* (MLP) dan *Convolutional Neural Network* (CNN). CNN merupakan bagian dari *Deep Learning* dan juga merupakan pengembangan dari MLP dimana CNN dapat memproses citra 2 dimensi sedangkan MLP hanya 1 dimensi.

Penelitian mengenai MLP dilakukan oleh Haryoko & Pramono (2016) untuk mengenali karakter plat kendaraan menggunakan metode *Canny* untuk deteksi tepian dan algoritma *Backpropagation* arsitektur MLP untuk mengenali karakter. Dari penelitian tersebut mampu mengenali karakter dengan tingkat keberhasilan 94,29%. Penelitian mengenai CNN dilakukan oleh Notonogoro, Jondri, & Arifianto (2018) untuk mengenali karakter pada plat kendaraan Indonesia dan mencapai akurasi hingga 95,45%. Penelitian lain dilakukan oleh Kim & Xie (2014) dengan melakukan pengenalan karakter pada tulisan tangan Korea dari 2 *database* berbeda dan mendapat nilai akurasi tertinggi yaitu 99,71%.

Dari kedua metode tersebut di atas, penulis menggunakan CNN untuk melakukan identifikasi atau klasifikasi citra teks untuk penerjemahan bahasa daerah. CNN memiliki tingkat akurasi yang lebih tinggi pada beberapa penelitian yang dilakukan untuk pengenalan karakter. CNN juga memiliki cara kerja yang hampir mirip dengan MLP dan CNN diyakini cenderung lebih efektif dan mudah

dilatih, karena pada CNN dilakukan konvolusi pada input untuk mereduksi ukuran citra (matriks). Selain itu pada penelitian ini, penulis juga menggunakan metode *Connected Component Analysis* (CCA) untuk proses segmentasi dan algoritma *Levenshtein Distance* untuk membantu proses penerjemahan. Data yang digunakan pada penelitian ini juga merupakan data primer yaitu data yang dibuat sendiri oleh penulis dan tim penelitian. Oleh karena itu penambahan metode lain dan penggunaan data primer menjadi pembeda antara penelitian ini dan penelitian mengenai OCR terdahulu.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya, maka dapat dirumuskan masalah sebagai berikut:

1. Apakah metode CNN dapat digunakan untuk mengenali teks pada suatu citra ?
2. Apakah hasil dari pengenalan teks pada citra dapat digunakan untuk proses penerjemahan teks bahasa Indonesia ke bahasa daerah ?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan metode CNN untuk pengenalan teks yang berbasis citra.
2. Mengimplementasikan proses penerjemahan teks bahasa Indonesia dari hasil pengenalan citra teks oleh CNN menjadi bahasa daerah.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk menumbuhkan kesadaran generasi sekarang ini akan bahasa daerahnya dengan membuat akses terjemahan melalui teks bahasa Indonesia berbasis citra / gambar.

1.5 Ruang Lingkup

Ruang lingkup dari penelitian ini adalah sebagai berikut :

1. Sistem ini dibuat hanya menerjemahkan tiga bahasa daerah yaitu bahasa Sunda (Bandung), bahasa Minang (Pariaman), bahasa Batak (Toba).
2. Sistem ini hanya mengenali teks cetak (*printed text*) dengan jenis *font* yaitu Times New Roman, Calibri, Arial, Bodoni, dan Helvetica.
3. Sistem hanya menerjemahkan 200 kata bahasa Indonesia yang dipilih dari daftar kata swadesh.

1.6 Luaran yang Diharapkan

Luaran yang diharapkan dari penelitian ini adalah terciptanya sistem yang dapat menerjemahkan suatu teks bahasa Indonesia pada citra menjadi teks dalam bahasa daerah.

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penyusunan proposal skripsi ini adalah sebagai berikut:

BAB 1 : PENDAHULUAN

Bab ini berisi tentang Latar Belakang, Rumusan Masalah, Tujuan Penelitian, Manfaat Penelitian, Ruang Lingkup, Luaran yang Diharapkan, dan Sistematika Penulisan.

BAB 2 : TINJAUAN PUSTAKA

Bab ini berisi dasar-dasar teori yang menjadi acuan dalam penyusunan laporan penelitian yang mendukung judul dari kegiatan yang penulis lakukan.

BAB 3 : METODOLOGI PENELITIAN

Bab ini berisi tentang tahapan penelitian, serta segala metode yang terdapat dalam penelitian ini.

BAB 4 : HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil perancangan sistem, implementasi algoritma, dan hasil pengujian model yang terdapat dalam penelitian ini.

BAB 5 : PENUTUP

Bab ini berisi tentang kesimpulan yang dapat diambil dari masalah-masalah yang telah diuraikan dan dibahas hasilnya, serta saran-saran yang dapat diberikan oleh penulis untuk penelitian berikutnya.

DAFTAR PUSTAKA

Berisikan sumber-sumber terkait penelitian ini.

RIWAYAT HIDUP

LAMPIRAN