

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bahasa merupakan sebuah sistem komunikasi yang digunakan dalam berinteraksi untuk menyampaikan informasi. Salah satu bentuk komunikasi dapat melalui tulisan atau aksara. Aksara menjadi salah media atau alat komunikasi yang masih digunakan hingga saat ini. Berdasarkan data Ethnologue tahun 2023, Indonesia menempati peringkat kedua negara dengan jumlah bahasa terbanyak yang digunakan dengan tujuh ratus dua puluh (720) bahasa. Diantara banyaknya bahasa yang ada di Indonesia terdapat beberapa bahasa dengan aksara yang cukup populer seperti Jawa, Bali, Sunda dan Lontara. Keberagaman bahasa yang ada di Indonesia, membuat setiap bahasa memiliki aksara dan karakteristik yang unik [1] [2]. Hal ini menjadi tantangan tersendiri dalam pengenalan karakter aksara bagi masyarakat Indonesia kedepannya.

Berdasarkan data Badan Pusat Statistik tahun 2023, penggunaan bahasa daerah mengalami penurunan terutama pada generasi Post-gen Z, dari data tersebut hanya 61,7% generasi Post-gen Z yang memiliki kemampuan berbahasa daerah di lingkungan tetangga atau kerabat. Sebanyak 38,3% generasi Post-gen Z tidak memiliki kemampuan dalam berbahasa daerah dan mengenali aksara daerahnya masing-masing. Menurunnya kemampuan generasi Post-gen-Z dalam mengenali aksara daerah dapat menyebabkan hilangnya identitas lokal dan keberagaman bahasa yang di Indonesia kedepannya. Menurunnya kemampuan berbahasa daerah pada generasi Post-gen Z disebabkan karena ketidakpahaman bentuk aksara daerah yang salah satunya dapat di edukasi melalui teknologi informasi. Teknologi informasi tersebut dapat digunakan dengan melakukan klasifikasi aksara. Dengan adanya klasifikasi aksra daerah diharapkan dapat meningkatkan pengetahuan bagi generasi Post-gen Z dalam mengenali akasara daerah yang ada di Indonesia. Hal ini bertujuan agar aksara daerah di Indonesia tidak dilupakan terutama pada generasi kedepannya. Terdapat salah satu bidang di teknologi informasi yaitu menggunakan

metode *deep learning* yang menjadi salah satu metode terbaik dalam klasifikasi jenis aksara daerah yang ada di Indonesia.

Deep Learning merupakan cabang ilmu pengembangan dari *machine learning* yang menggunakan jaringan saraf tiruan mendalam untuk mengenali pola dari data yang kompleks [3]. *Deep learning* memiliki karakteristik menyerupai jaringan saraf pada manusia yang terdiri dari banyak lapisan neuron (*hidden layer*) untuk melakukan klasifikasi perintah input sehingga dapat menghasilkan output yang diinginkan [4]. Pemanfaatan *deep learning* mencakup berbagai pekerjaan dalam kehidupan seperti melakukan prediksi, klasifikasi, pengenalan pola dan objek, diagnosa dan lain sebagainya [5]. Beberapa metode *deep learning* yang dimanfaatkan dengan sangat baik dalam melakukan klasifikasi citra adalah *Convolutional Neural Network* (CNN), *Multi Layer perceptron* (MLP), dan *Deep Belief Network* (DBN).

Convolutional Neural Network (CNN) merupakan sebuah algoritma jaringan saraf tiruan yang dapat mengenali informasi prediktif dari suatu objek seperti, teks, gambar, suara dan video [6]. CNN memiliki kemampuan yang sangat baik dalam melakukan klasifikasi terutama dalam klasifikasi citra digital. Hal ini dikarenakan CNN bekerja berdasarkan sistem pengenalan gambar pada korteks visual manusia [5] [7]. Adapun algoritma *Multi Layer Perceptron* (MLP) merupakan jaringan saraf tiruan yang paling sederhana, namun cukup umum digunakan. Mirip seperti CNN, MLP terdiri dari beberapa lapisan neuron dan saling terhubung secara keseluruhan. MLP dapat digunakan dalam berbagai tugas, seperti klasifikasi, regresi serta *clustering* [7] [8]. Selain dua algoritma tersebut terdapat algoritma *Deep Belief Network* (DBN) yang dapat digunakan dalam klasifikasi citra. DBN merupakan jaringan saraf tiruan yang terdiri dari beberapa lapisan unit tersembunyi. DBN dapat digunakan dalam melakukan berbagai tugas, seperti klasifikasi, regresi, clustering dan feature learning [10].

Terdapat beberapa keunggulan dari ketiga algoritma di atas untuk melakukan klasifikasi aksara. Pada algoritma CNN memiliki keunggulan model sangat mudah dilatih untuk memproses data yang berbentuk gambar atau citra serta dapat dilatih dengan dataset yang relatif kecil. Algoritma MLP memiliki keunggulan kemudahan

dalam melatih model dan dapat digunakan dalam berbagai tugas [8]. Sedangkan pada algoritma DBN memiliki keunggulan dapat mempelajari fitur kompleks dari data dan bisa digunakan dalam mempelajari pembelajaran tanpa pengawasan [11] [12].

Dibalik kelebihan beberapa algoritma tersebut, terdapat kekurangan pada algoritma tersebut dalam melakukan klasifikasi aksara. Pertama, pada algoritma MLP memiliki kekurangan sangat mudah mengalami *overfitting* terutama pada dataset yang relatif kecil serta cukup sulit untuk diinterpretasikan [13]. Kemudian pada algoritma DBN memiliki kekurangan pada proses pelatihan model yang membutuhkan waktu yang lama untuk melatih, terutama pada dataset gambar yang berdimensi besar [12]. Selain itu DBN dapat mengalami *overfitting* pada dataset yang berukuran kecil dan DBN cukup sulit untuk diinterpretasikan.

Kekurangan dari dua algoritma di atas dapat ditangani oleh CNN karena kelebihan dalam memproses data gambar. Namun, CNN memiliki sedikit kelemahan ketika melakukan pelatihan model sehingga terjadi kondisi *overfitting* [14]. *Overfitting* merupakan kondisi dimana model memberikan prediksi yang akurat pada data pelatihan namun tidak untuk data baru. Kekurangan pada CNN dapat diatasi dengan cara optimasi *hyperparameter*. Optimasi *hyperparameter* memungkinkan optimasi pada model CNN sehingga mendapatkan hasil yang lebih baik dan mencegah *overfitting*.

Berdasarkan uraian di atas, penelitian ini akan melakukan optimasi *hyperparameter* pada CNN untuk mendapatkan hasil yang akurat dalam melakukan klasifikasi aksara. Dalam penelitian ini menggunakan tiga *hyperparameter*, yaitu *learning rate* untuk meningkatkan akurasi dari model, *filter kernel* untuk meningkatkan efisiensi model selama pelatihan dan *dropout* guna menghindari model dari kemungkinan *overfitting*. Penelitian ini juga akan memaparkan mengenai hasil dari analisis loss dan akurasi antara model yang menggunakan optimasi *hyperparameter* dan tanpa optimasi *hyperparameter*

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan di atas, dapat diketahui permasalahan dalam penelitian ini sebagai berikut :

Berdasarkan data BPS Sebanyak 38,3% generasi Post-gen Z mengalami penurunan kemampuan dalam berbahasa daerah dan mengenali aksara daerahnya masing-masing. Pada bidang teknologi informasi terdapat algoritma CNN yang digunakan untuk klasifikasi aksara. Namun, algoritma CNN memiliki kelemahan dalam melakukan pelatihan model sehingga sering terjebak dalam kondisi *overfitting* yang merupakan kondisi dimana model memberikan prediksi yang akurat pada data pelatihan namun tidak untuk data baru.

1.3 Pertanyaan Penelitian

Berdasarkan rumusan masalah di atas, maka pertanyaan penelitian dalam melakukan penelitian ini yaitu:

1. Bagaimana implementasi optimasi *hyperparameter* pada CNN dalam klasifikasi aksara ?
2. Bagaimana pengaruh hyperparameter pada tingkat keberhasilan dalam klasifikasi aksara ?
3. Berapa perbandingan hasil akurasi model CNN menggunakan hyperparameter dan tanpa menggunakan hyperparameter pada klasifikasi aksara ?

1.4 Batasan Masalah

Berdasarkan rumusan masalah dan tujuan penelitian, maka untuk mewujudkan penelitian yang sesuai dengan masalah yang ada diperoleh batasan-batasan masalah penelitian sebagai berikut :

1. Dataset yang digunakan didapatkan dengan cara *crawling* melalui internet.
2. Dataset yang digunakan terbatas pada penggunaan citra aksara Jawa, Bali, Sunda, dan Lontara. Dataset yang dikumpulkan berjumlah 400 dataset training dan 400 dataset pengujian.

1.5 Tujuan Penelitian

Berdasarkan rumusan masalah, dapat dijabarkan tujuan penelitian sebagai berikut:

1. Mengimplementasikan *hyperparameter* pada CNN dalam mengklasifikasikan aksara agar mendapatkan akurasi yang tinggi dan mencegah terjadinya *overfitting*.
2. Membandingkan hasil pelatihan model CNN sebelum dilakukan optimasi *hyperparameter* dan setelah dilakukan optimasi *hyperparameter* dalam klasifikasi aksara.

1.6 Manfaat Penelitian

Berdasarkan rumusan masalah, batasan masalah dan tujuan penelitian yang sudah diuraikan diatas, maka dapat diketahui manfaat dari penelitian ini adalah:

1. Dapat menambah pengetahuan mengenai arsitektur *convolutional neural network* dan optimasi *hyperparameter*.
2. Dapat menambah pemahaman tentang klasifikasi aksara.
3. Dapat digunakan sebagai acuan mengenai optimasi *hyperparameter* pada CNN.