

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan oleh penulis terkait pengenalan emosi berdasarkan suara manusia menggunakan metode *Convolutional Neural Network* (CNN) dengan menggunakan data citra representasi *MFCC Spectrogram*, kesimpulan yang akan penulis berikan sebagai berikut:

1. Penerapan fitur ekstraksi *Mel-Frequency Cepstral Coefficients* (MFCC) yang direpresentasikan kedalam bentuk citra *spectrogram* dan metode *Convolutional Neural Network* (CNN) dalam pengenalan emosi berdasarkan suara manusia dapat mengenali emosi *angry* (marah), *disgust* (jijik), *fear* (takut), *happy* (senang), *sad* (sedih), *neutral* (netral), dan *surprise* (terkejut) dengan baik.
2. Ketika menentukan kombinasi dari *hyperparameter* arsitektur CNN, maka perlu melakukan *trial* dan *error* dengan berbagai uji coba kombinasi *parameter*, karena banyak sekali kombinasi yang dapat dilakukan, dalam penelitian ini kombinasi yang dilakukan yaitu *convolutional layer*, *batch size*, *hidden layer*, *learning rate*, *epoch*, dan pembagian rasio data latih, validasi dan uji. Dari hasil menentukan kombinasi *hyperparameter*, didapatkan model terbaik yaitu ketika menggunakan layer konvolusi sebanyak 4 yaitu (8, 16, 32, 128) serta menggunakan *Max Pooling* dengan ukuran matriks 2 x 2 sebagai *pooling layer*, lalu menggunakan fungsi aktivasi untuk *input layer* yaitu ReLU, fungsi aktivasi pada *output layer* yaitu Softmax dengan jumlah 7 *neuron*, lalu model di *compile* dengan *optimizer* Adam, menggunakan nilai *learning rate* default dari keras yaitu 0.001,

menggunakan *epoch* sebanyak 50 kali untuk pelatihan, kemudian penggunaan ukuran *batch size* yang digunakan adalah 32 *batch* dan pembagian rasio pada data latih, validasi, dan uji sebesar 80%:10%:10%, sebagai konfigurasi parameter terbaik yang digunakan untuk penelitian ini.

3. Kemampuan model yang dihasilkan pada penelitian ini memiliki performa yang optimal dalam pengenalan emosi. Model ini dapat dengan baik membedakan kelas emosi dengan akurasi kelas *Angry* (Marah) yaitu 96,80%, kelas *Disgust* (Jijik) yaitu 97,86%, kelas *Fear* (Takut) yaitu 96.16%, kelas *Happy* (Senang) yaitu 95,52%, kelas *Neutral* (Netral) yaitu 96,37%, kelas *Sad* (Sedih) yaitu 96,80%, dan kelas *Surprise* (Terkejut) yaitu 95,52%. Sehingga rata-rata dari hasil akurasi model sebesar 96.42%.

5.2 Saran

Berdasarkan hasil penelitian ini, berikut ini saran yang penulis berikan untuk pengembangan penelitian kedepannya:

1. Dataset yang digunakan dalam penelitian ini masih menggunakan bahasa Inggris, sehingga perlu dilakukan penelitian selanjutnya terkait penggunaan bahasa Indonesia, sebagai dataset untuk pembangunan model pengenalan emosi berdasarkan suara manusia.
2. Menggunakan teknik pengolahan data yang lebih baik, dengan menerapkan normalisasi atau *augmentasi* pada data, untuk dapat meningkatkan performa model pengenalan emosi berdasarkan suara manusia.
3. Menyamaratakan durasi waktu setiap data suara, untuk mendapatkan kualitas *dataset* yang lebih baik, agar model yang dibangun lebih optimal.

4. Memperhatikan penggunaan kombinasi *hyperparameter* yang akan digunakan dalam membangun atau melatih model, agar model yang sudah dibangun tidak *overfitting*.
5. Model dapat dikembangkan menjadi aplikasi berbasis *android* atau *web*, untuk pengenalan emosi berdasarkan suara manusia secara *realtime*.