

PEMBANGUNAN MODEL PENGENALAN SUARA MENGGUNAKAN METODE *DEEP LEARNING*

ABSTRAK

Seiring perkembangan teknologi pengenalan suara seperti pada *Google voice*, *Siri* dan *Cortana*, penggunaan suara kini dapat dianggap sebagai suatu kebutuhan bagi kalangan pengguna teknologi. Masalah utama pengenalan suara adalah suara mempunyai tingkat variabilitas sangat tinggi, yaitu suara dan objek yang sama dapat menghasilkan data berbeda, terlebih lagi untuk suara dan objek yang berbeda. Permasalahan tersebut menjadi tantangan dalam hal pengenalan suara, sehingga pada penelitian ini melakukan pengenalan suara dengan membangun model pengenalan suaranya. Eksperimen dilakukan mulai dari perekaman suara, praproses audio, konversi, pelabelan, praproses citra dan pengenalan dengan mengimplementasikan metode *Convolutional Neural Network* (CNN) untuk ekstraksi ciri data suara dan pembentukan *dataset* (data latih dan uji) serta proses pelatihan dan proses pengujian menggunakan empat skenario. Skenario pertama adalah memisahkan data berdasarkan masing-masing responden dengan tujuan untuk mengetahui model pengenalan suara terbaik yang dihitung melalui jumlah pengenalan suara yang diucapkan dan teknik pengucapannya. Skenario proses ini jumlah *epoch* yang digunakan berkelipatan 200 sebanyak 5 kali yaitu 200, 400, 600, 800, 1000. Sedangkan *batch*-nya setiap *epoch* adalah sesuai *data training*. Skenario kedua adalah menerapkan peningkatan nilai *epoch* dari skenario pertama. Peningkatan nilai *epoch* tersebut menjadi 2000, 4000, 6000, 8000 dan 10000. Skenario ketiga adalah menggunakan *cross validation k-fold* (10) untuk membangun model pengenalan yang dilakukan tanpa membedakan data berdasarkan responden dan teknik pengucapannya. Langkah yang dilakukan dengan membagi seluruh data menjadi sepuluh bagian (*fold*) untuk dilakukan sepuluh kali pengujian, masing-masing bagian mencakup semua kelas data. Selanjutnya data di *training* 1000 *epoch*. Sedangkan *batch* untuk setiap *epoch* adalah sesuai data *training*. Skenario keempat adalah menerapkan peningkatan nilai *epoch* dari skenario ketiga. Peningkatan nilai *epoch* tersebut menjadi 10000 *epoch*. Berdasarkan penelitian yang telah dilakukan tersebut, hasil rata-rata akurasi pengenalan adalah 27,32% pada skenario 1, 40,59% pada skenario 2, 16,66% pada skenario 3, dan 43,70% pada skenario 4.

Kata Kunci: Pengenalan suara, Citra spektrogram, *Deep Learning*, *Convolutional Neural Network*.

DEVELOPMENT OF VOICE RECOGNITION MODEL USING DEEP LEARNING METHOD

ABSTRACT

Along with the development of voice recognition technology as in Google voice, Siri and Cortana, voice usage can now be regarded as a necessity for the users of technology. The main problem of voice recognition is that the sound has a very high level of variability, that is, the sound and the same object can produce different data, especially for different sounds and objects. The problem becomes a challenge in terms of speech recognition, so in this study to do voice recognition by building the voice recognition model. Experiments were conducted from sound recording, audio preview, conversion, labeling, image preview and recognition by implementation the Convolutional Neural Network (CNN) method for extracting voice data features and dataset formation (training and test data) as well as training and testing process using four scenarios. The first scenario is to separate each respondent's data in order to know the best voice recognition model calculated by the number of spoken speech recognition and the pronunciation technique. The scenario of this process is the number of epoch used 200 multiplication of 5 times that is 200, 400, 600, 800, 1000. While the batch of each epoch is appropriate training data. The second scenario is to apply the epoch value increase from the first scenario. The increase of the epoch value to 2000, 4000, 6000, 8000 and 10000. The third scenario is to use cross-validation k-fold (10) to construct an introduction model undertaken without discriminating the data based on the respondent and the pronunciation technique. The steps are done by dividing all data into ten (fold) sections for ten tests, each covering all the data classes. Furthermore the data in training 1000 epoch. While the batch for each epoch is appropriate training data. The fourth scenario is to apply an increase in the epoch value of the third scenario. The increase of epoch value becomes 10000 epoch. Based on the research that has been done, the average result of recognition accuracy is 27.32% in scenario 1, 40,59% in scenario 2, 16,66% in scenario 3 and 43,70% in scenario 4.

Keyword: Voice Recognition, Spectrogram Image, Deep Learning, Convolutional Neural Network.