ABSTRAK

IMPLEMENTASI MODEL *CONVOLUTIONAL RECURRENT NEURAL NETWORK* (CRNN) DENGAN *BIDIRECTIONAL LONG SHORT-TERM MEMORY* (BI-LSTM) UNTUK PENGENALAN EMOSI PADA SUARA

Oleh

M. RADITYA ADHIRAJASA

Pengenalan emosi pada suara merupakan tugas kompleks yang krusial untuk interaksi manusia-mesin. Penelitian ini mengimplementasikan dan mengevaluasi model *Convolutional Recurrent Neural Network* (CRNN) yang menggabungkan keunggulan *Convolutional Neural Network* (CNN) untuk ekstraksi fitur spasial dengan *Bidirectional Long Short-Term Memory* (BiLSTM) untuk analisis dependensi temporal. Model ini dilatih dan diuji menggunakan dataset CREMA-D, yang mencakup enam kelas emosi: marah, jijik, takut, bahagia, netral, dan sedih. Fitur diekstraksi dari sinyal audio menggunakan *log-Mel spectrogram* untuk merepresentasikan karakteristik suara yang relevan dengan persepsi manusia. Untuk meningkatkan variasi data dan generalisasi model, diterapkan teknik augmentasi data berupa *pitch shifting* dan *time stretching*. Hasil evaluasi menunjukkan bahwa model CRNN yang didukung dengan augmentasi data berhasil mencapai akurasi validasi sebesar 68.23%, sebuah peningkatan dari 66.08% yang dicapai oleh model tanpa augmentasi. Model ini menunjukkan performa yang menjanjikan, terutama dalam mengklasifikasikan emosi marah dan netral. Penelitian ini menyimpulkan bahwa arsitektur CRNN dengan BiLSTM efektif untuk tugas pengenalan emosi suara, dan augmentasi data berperan penting dalam meningkatkan kinerjanya.

Kata Kunci: Pengenalan Emosi Suara, CRNN, BiLSTM, *Log-Mel Spectrogram,* Augmentasi Data.

ABSTRACT

IMPLEMENTATION OF A CONVOLUTIONAL RECURRENT NEURAL NETWORK (CRNN) WITH BIDIRECTIONAL LONG SHORT-TERM MEMORY (BI-LSTM) FOR SPEECH EMOTION RECOGNITION

By

M. RADITYA ADHIRAJASA

Speech emotion recognition is a complex task crucial for human-machine interaction. This study implements and evaluates a Convolutional Recurrent Neural Network (CRNN) model that combines the advantages of a Convolutional Neural Network (CNN) for spatial feature extraction with Bidirectional Long Short-Term Memory (BiLSTM) for temporal dependency analysis. The model is trained and tested using the CREMA-D dataset, which includes six emotion classes: anger, disgust, fear, happiness, neutral, and sadness. Features are extracted from the audio signals using log-Mel spectrograms to represent sound characteristics relevant to human perception. To enhance data variation and model generalization, data augmentation techniques such as pitch shifting and time stretching were implemented. The evaluation results show that the CRNN model with data augmentation achieved a validation accuracy of 68.23%, an increase from the 66.08% achieved by the model without augmentation. The model shows promising performance, especially in classifying the anger and neutral emotions. This study concludes that the CRNN architecture with BiLSTM is effective for the task of speech emotion recognition, and that data augmentation plays a crucial role in enhancing its performance.

Keywords: *Speech Emotion Recognition, CRNN, BiLSTM, Log-Mel Spectrogram, Data Augmentation.*