

Perancangan Pengenalan Emosi dari Percakapan Berbahasa Indonesia Menggunakan Model CNN-GRU dengan Augmentasi Data = Design of Speech Emotion Recognition for Indonesian Language Using CNN-GRU Model With Data Augmentation

Ali Alatas, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920564593&lokasi=lokal>

Abstrak

Dalam kehidupan sehari-hari, emosi memainkan peran penting dalam membentuk cara manusia berkomunikasi dan berinteraksi, baik dengan sesama maupun dengan teknologi. Dalam bidang Human-Computer Interaction (HCI), pengenalan emosi menjadi salah satu inovasi yang memungkinkan sistem komputer memahami perasaan manusia secara lebih mendalam. Penelitian ini bertujuan untuk mengembangkan sistem Speech Emotion Recognition (SER) berbasis suara percakapan berbahasa Indonesia menggunakan kombinasi model Convolutional Neural Network (CNN) dan Gated Recurrent Unit (GRU), dengan dukungan teknik augmentasi data untuk meningkatkan performa dan generalisasi model. Penelitian ini dilakukan melalui tiga tahap pengujian: pertama, eksperimen menggunakan benchmark dari TESS Dataset yang berbahasa Inggris; kedua, eksperimen terhadap jumlah augmentasi data untuk menentukan konfigurasi terbaik; dan ketiga, eksperimen membandingkan kinerja model CNN, GRU, dan CNN-GRU. CNN digunakan untuk mengekstraksi fitur suara utama, seperti MFCC, Chroma, Zero-Crossing Rate (ZCR), RMS, dan Spectral Contrast, sementara GRU menangkap pola temporal dalam data. Hasil penelitian menunjukkan bahwa kombinasi CNN-GRU dengan 6 jenis augmentasi data memberikan performa terbaik, dengan akurasi 94.49% dan loss 0.8136 pada dataset berbahasa Indonesia, serta akurasi 99.72% dan loss 0.1915 pada benchmark dari TESS Dataset yang berbahasa Inggris. Temuan ini menegaskan bahwa teknik augmentasi data efektif dalam meningkatkan stabilitas dan akurasi model, bahkan ketika dihadapkan pada variasi kualitas data. Penelitian ini berkontribusi pada pengembangan teknologi HCI yang lebih intuitif, dengan potensi penerapan dalam mendukung kesehatan mental, layanan berbasis suara, dan sistem pendidikan yang responsif terhadap emosi manusia.

.....In daily life, emotions play a crucial role in shaping how humans communicate and interact, both with each other and with technology. In the field of Human-Computer Interaction (HCI), emotion recognition has become one of the innovations that enable computer systems to deeply understand human feelings. This research aims to develop a Speech Emotion Recognition (SER) system based on Indonesian speech using a combination of Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) models, supported by data augmentation techniques to improve the performance and generalization of the model. The research was conducted through three stages of testing: first, an experiment using the TESS Dataset benchmark in English; second, an experiment on the number of data augmentations to determine the optimal configuration; and third, an experiment comparing the performance of CNN, GRU, and CNN-GRU models. CNN was utilized to extract key audio features, such as MFCC, Chroma, Zero-Crossing Rate, RMS, and Spectral Contrast, while GRU captured temporal patterns in the data. The results showed that the combination of CNN-GRU with 6 types of data augmentation provided the best performance, achieving an accuracy of 94.49% and a loss of 0.8136 on the Indonesian dataset, as well as an accuracy of 99.72% and a loss of 0.1915 on the benchmark TESS Dataset in English. These findings affirm that data augmentation

techniques are effective in improving model stability and accuracy, even when faced with variations in data quality. This research contributes to the development of more intuitive HCI technologies, with potential applications in supporting mental health, voice-based services, and educational systems that are responsive to human emotions.