

Deteksi emosi wicara bahasa Indonesia pada podcast dan talkshow dengan menggunakan metode recurrent neural network dan convolutional neural network = Emotion detection on Indonesian language in podcast and talkshow using recurrent neural network and convolutional neural network method

Yumna Pratista Tastaftian, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20516193&lokasi=lokal>

Abstrak

Speech Emotion Recognition adalah teknologi yang mampu bisa mendeteksi emosi lewat data suara yang diproses oleh sebuah mesin. Media yang sering digunakan untuk menjadi media interaksi antara 2 orang atau lebih yang saat ini sedang digunakan oleh banyak orang adalah Podcast, dan Talkshow. Seiring berkembangnya SER, penelitian terakhir menunjukkan bahwa penggunaan metode Deep Learning dapat memberikan hasil yang memuaskan terhadap sistem SER. Pada penelitian ini akan diimplementasikan model Deep Learning yaitu dengan Recurrent Neural Network (RNN) variasi Long Short Term Memory (LSTM) untuk mengenali 4 kelas emosi (marah, netral, sedih, senang). Penelitian ini menguji model yang digunakan untuk mengenali emosi dari fitur akustik pada data secara sekuensial. Skenario training dan testing dilakukan dengan metode one-against-all dan mendapatkan hasil (1) Dataset talkshow mengungguli dataset podcast untuk tipe 1 dan 2 dan untuk semua emosi yang dibandingkan; (2) Untuk dataset podcast pada emosi marah, senang, dan sedih didapatkan akurasi optimal pada dataset tipe 1 yaitu 67.67%, 71.43%, dan 68,29%, sedangkan untuk emosi netral didapatkan akurasi terbaik pada dataset tipe 2 dengan 77.91%; (3) Untuk dataset talkshow pada emosi marah, netral, dan sedih didapatkan akurasi terbaik pada dataset tipe 2 yaitu 78.13%, 92.0%, dan 100%. Dapat disimpulkan bahwa dataset talkshow secara garis besar memberikan hasil yang lebih optimal namun memiliki variasi data yang lebih sedikit dari dataset podcast. Dari sisi panjang data, pada penelitian ini didapatkan akurasi yang lebih optimum pada dataset dengan tipe 2.

.....Speech Emotion Recognition is a technology that is able to detect emotions through voice data that is processed by a machine. Media that is often used to be a medium of interaction between two or more people who are currently being used by many people are Podcasts, and Talkshows. As SER develops, recent research shows that the use of the Deep Learning method can provide satisfactory results on the SER system. In this study a Deep Learning model will be implemented, this study uses Long Short Term Memory (LSTM) as one of the variation of Recurrent Neural Network (RNN) to recognize 4 classes of emotions (angry, neutral, sad, happy). This study examines the model used to recognize emotions from acoustic features in sequential data. Training and testing scenarios are conducted using the one-against-all method and get results (1) The talkshow dataset outperforms the podcast dataset for types 1 and 2 and for all emotions compared; (2) For the podcast dataset on angry, happy, and sad emotions, the optimal accuracy in type 1 dataset is 67.67%, 71.43%, and 68.29%, while for neutral emotions the best accuracy is obtained in type 2 dataset with 77.91%; (3) For the talkshow dataset on angry, neutral, and sad emotions the best accuracy is obtained for type 2 datasets, namely 78.13%, 92.0%, and 100%. It can be concluded that the talkshow dataset in general gives more optimal results but has fewer data variations than the podcast dataset. In terms of data length, this study found more optimum accuracy in dataset with type 2.