

Perancangan Model Pengenalan Emosi pada Percakapan Berbahasa Indonesia dengan Metode Transfer Learning VGG-16 = Design of Speech Emotion Recognition Model for Indonesian Language with VGG-16 Transfer Learning Method

Arvalinno, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20518935&lokasi=lokal>

Abstrak

Kecerdasan buatan atau Artificial Intelligence (AI) banyak berkembang dalam sektor-sektor seperti: speech recognition, computer vision, Natural Language Processing, dll. Salah satu sektor penting yang banyak dikembangkan oleh peneliti adalah Speech Emotion Recognition atau pengenalan emosi berdasarkan suara manusia. Penelitian ini semakin berkembang karena timbul sebuah tantangan bagi manusia untuk memiliki interaksi mesin dan manusia yang lebih natural yaitu suatu mesin yang dapat merespon emosi manusia dengan memberikan balasan yang tepat juga. Perancangan Speech Emotion Recognition pada penelitian ini menggunakan dataset berupa fitur ekstraksi audio MFCC, Spectrogram, Mel Spectrogram, Chromagram, dan Tonnetz serta memanfaatkan metode Transfer Learning VGG-16 dalam pelatihan modelnya. Dataset yang digunakan diperoleh dari pemotongan audio dari beberapa film berbahasa Indonesia dan kemudian audio yang diperoleh diekstraksi fitur dalam kelima bentuk fitur yang disebut sebelumnya. Hasil akurasi model paling baik dalam penelitian ini adalah model transfer learning VGG-16 dengan dataset Mel Spectrogram yaitu dengan nilai akurasi 56.2%. Dalam pengujian model dalam pengenalan setiap emosi, f1-score terbaik diperoleh model transfer learning VGG-16 dengan dataset Mel Spectrogram dengan f1-score yaitu 55.5%. Skala mel yang diterapkan pada ekstraksi fitur mel spectrogram berpengaruh terhadap baiknya kemampuan model dalam mengenali emosi manusia.

Artificial Intelligence has been used in many sectors, such as speech recognition, computer vision, Natural Language Processing, etc. There was one more important sector that has been developed well by the scientists which are Speech Emotion Recognition. This research is developing because of the new challenge by human to have a better natural interaction between machines and humans where machines can respond to human's emotions and give proper feedback. In this research, to create the speech emotion recognition system, audio feature extraction such as MFCC, Spectrogram, Mel Spectrogram, Chromagram, and Tonnetz were used as input, and using VGG-16 Transfer Learning Method for the model training. The datasets were collected from the trimming of audio from several Indonesian movies, the trimmed audio will be extracted to the 5 features mentioned before. The best model accuracy is VGG-16 with Mel Spectrogram dataset which has reached 56.2% of accuracy. In terms of recognizing the emotion, the best f1-score is reached by the model VGG-16 with Mel Spectrogram dataset which has 55.5% of f1-score. Mel scale that is applied to the feature extraction of mel spectrogram affected the model's ability to recognize human emotion.