

Perancangan model pengenalan emosi pada percakapan berbahasa indonesia dengan fitur mfcc, mel spectrogram dan chroma serta klasifikasi multi layer perceptron = Design of speech emotion recognition model for indonesian language with mfcc, mel spectrogram and chroma features and multi layer perceptron classifier

Jonathan, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20518997&lokasi=lokal>

Abstrak

Emosi atau perasaan manusia adalah salah satu faktor yang tidak dapat dikendalikan dalam aktivitas apapun. Tidak sedikit juga pekerjaan yang seringkali berkaitan dengan emosi manusia terutama di industri hiburan dan juga kesehatan. Oleh karena itu, 1 dekade kebelakang banyak riset yang dilakukan untuk mempelajari emosi manusia secara langsung maupun menggunakan teknologi. Pengembangan model speech emotion recognition berbahasa Indonesia masih sangat sedikit dan oleh karena itu dibutuhkan perbandingan secara spesifik pada penelitian ini diantara dua model classifier yaitu Convolutional Neural Network (CNN) dan juga Multilayer Perceptron (MLP) untuk menentukan model yang menghasilkan akurasi terbaik dalam memprediksi emosi dari suara manusia.

Dalam speech recognition secara umum, salah satu faktor penting dalam mendapatkan model dengan akurasi terbaik adalah metode ekstraksi fiturnya. Oleh karena itu, penelitian ini menggunakan 3 fitur untuk melakukan pelatihan terhadap model yaitu Mel-frequency Cepstral Coefficients (MFCC), Mel-Spectrogram dan chroma. Dari 3 fitur ini, divariasikan dan menghasilkan 7 metode ekstraksi yang berbeda untuk digunakan sebagai input pelatihan model.

Terakhir, untuk memastikan bahwa model sudah menggunakan parameter terbaik, dilakukan eksperimen dengan membandingkan model yang menggunakan batch size serta activation function yang berbeda. Ditemukan bahwa dengan menggunakan CNN dan fitur gabungan antara MFCC, mel-spectrogram dan juga chroma menghasilkan model dengan skor akurasi 50.6% sedangkan menggunakan MLP dengan fitur yang sama menghasilkan model dengan skor akurasi 58.47%.

Emotions or human feelings are one of the factors that cannot be controlled in any activity. There are also many jobs that are often related to human emotions, especially in the entertainment and health industries. The development of speech emotion recognition models in Indonesian is still very little and therefore a specific comparison is needed in this study between two classifier models, namely Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP) to determine the model that produces the best accuracy in predicting the emotion of the human voice.

In speech recognition in general, one of the important factors in acquiring a model with the best accuracy is the feature extraction method. Therefore, this study uses 3 features to train the model, namely Mel-frequency Cepstral Coefficients (MFCC), Mel-Spectrogram and chroma. From these 3 features, they were varied and resulted in 7 different extraction methods to be used as model training inputs.

Finally, to ensure that the model has used the best parameters, an experiment was conducted by comparing models using different batch sizes and activation functions. It was found that using CNN and the combined features of MFCC, mel-spectrogram and also chroma resulted in a model with an accuracy score of 50.6% while using MLP with the same features resulted in a model with an accuracy score of 58.47%.