

Perancangan Model Pengenalan Emosi pada Percakapan Berbahasa Indonesia dengan Fitur MFCC dan Metode Machine Learning = Design of Speech Emotion Recognition Model for Indonesian Language with MFCC Features and Machine Learning Methods

Philipus Kristian Renaldy, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20518944&lokasi=lokal>

Abstrak

Emosi merupakan hal penting yang dimiliki oleh manusia. Banyak riset yang sudah dilakukan untuk menganalisis emosi seseorang secara langsung maupun tidak langsung. Salah satu topik dari machine learning yang berkembang adalah sistem yang mampu mempelajari isi suara manusia untuk menentukan emosi seseorang yang dinamakan speech emotion recognition. Banyak riset yang sudah dilakukan masih menggunakan dataset berbahasa Inggris, untuk itu diperlukan penelitian speech emotion recognition dengan menggunakan dataset berbahasa Indonesia. Pada penelitian ini dilakukan analisa speech emotion recognition menggunakan 4 model berbeda yaitu Convolutional Neural Network (CNN), Support Vector Machines (SVM), K-Nearest Neighbor (KNN), dan Logistic Regression (LR). Penelitian ini dilakukan dengan menggunakan hasil ekstraksi dari Mel-frequency Cepstral Coefficient (MFCC) yang dimasukkan ke dalam bentuk matriks 2D sebagai input menuju model percobaan. Dataset yang digunakan merupakan cuplikan dialog berbahasa Indonesia dengan karakteristik emosi tertentu yang sudah dikelompokkan terlebih dahulu. Dari percobaan yang telah dilakukan, didapatkan hasil bahwa model SVM memiliki tingkat rata-rata akurasi tertinggi jika dibandingkan dengan model lainnya, yaitu sebesar 59%. Sedangkan untuk model LR, KNN, dan CNN didapatkan tingkat akurasi rata-rata secara berurutan sebesar 54,5%; 53,5%; dan 47,7%.

Emotions are important things in human life. A lot of research had been done to analyze persons' emotions directly or indirectly. One of the topics of machine learning that is developing is a system that could understand the content of the human voice to determine a person's emotions called speech emotion recognition. Much of the research that had been done still uses English datasets. Therefore, speech emotion recognition research using Indonesian language datasets is needed. In this study, Speech Emotion Recognition analysis was performed using 4 different models, such as Convolutional Neural Network (CNN), Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and Logistic Regression (LR). This study was conducted using the extraction outputs from the Mel-frequency Cepstral Coefficient (MFCC) which was converted into a 2D matrix. The output would be used as an input to the model. The dataset used was a snippet of Indonesian dialogue with several emotional characteristics that had been grouped. Based on this study, the results showed that the SVM model had the highest average level of accuracy around 59%. Meanwhile, for the LR, KNN, and CNN models, the average accuracy rate were 54.5%; 53.5%; and 47.7%.