

ABSTRAK

Manusia dan emosi adalah satu kesatuan yang tidak bisa terpisahkan. Umumnya raut wajah digunakan untuk mengenali emosi seseorang, namun dengan perkembangan teknologi saat ini, emosi seseorang dapat diketahui dengan mendengar intonasi suara yang diucapkan. Mayoritas penelitian dibidang pengenalan emosi pada suara, menggunakan dataset penelitian berbahasa asing yang mengakibatkan susahny proses divalidasi karena adanya perbedaan bahasa. Sehingga pada penelitian ini akan dilakukan pengenalan atau deteksi emosi pada suara menggunakan dataset berbahasa Indonesia.

Mel-Frequency Cepstral Coefficients (MFCC) merupakan metode ekstraksi ciri yang umum digunakan untuk pengenalan emosi pada suara karena memiliki prinsip kerja yang menyerupai telinga manusia. Prinsip kerja ini menjadikan MFCC mampu mengatasi sebagian besar penelitian yang berhubungan dengan pengenalan berdasarkan sinyal suara. Ada beberapa proses yang terjadi di dalam metode MFCC, yaitu *Pre-emphasize*, *Frame Blocking*, *Windowing*, *Fast Fourier Transform*, *Mel-scaled Filterbank*, dan *Discrete Cosine Transform*. *K-Nearest Neighbor* (KNN) merupakan algoritma *supervised machine learning* yang banyak digunakan untuk menyelesaikan permasalahan klasifikasi maupun deteksi karna kelebihanannya yang tangguh terhadap data latih yang *noise*. Data yang digunakan pada penelitian ini adalah data suara *podcast* yang diunduh dari beberapa *channel Youtube*.

Pengujian yang dilakukan pada penelitian ini yaitu *parameter tuning* dan pengujian kemampuan model dalam mendeteksi emosi terhadap data baru. Selain itu pengujian antara laki-laki dan wanita dilakukan secara terpisah karena laki-laki dan wanita memiliki karakter suara yang berbeda. Kombinasi antara metode ekstraksi ciri MFCC dan metode KNN mampu dalam mendeteksi emosi berdasarkan intonasi suara pada pria dengan akurasi sebesar 88,8% terhadap data *test* dan 80% terhadap data baru. MFCC dan KNN juga mampu mendeteksi emosi berdasarkan intonasi suara pada wanita dengan akurasi 92,5% terhadap data *test* namun terhadap data baru hanya mampu mencapai akurasi sebesar 50%. Model pada wanita mengalami kondisi *overfitting* yang disebabkan oleh data augmentasi *pitch shifting* yang tidak mampu menghasilkan sampel data yang cukup untuk secara akurat merepresentasikan semua kemungkinan nilai data *input*.

Kata kunci: *Mel-Frequency Cepstral Coefficient*, *K-Nearest Neighbor*, **Deteksi emosi, Pemrosesan Sinyal.**

ABSTRACT

Humans and emotions are inseparable. Generally, facial expressions are used to recognize a person's emotions, but with the development of technology today, a person's emotions can be known by hearing the intonation of the spoken voice. The majority of research in the field of voice emotion recognition uses foreign language research datasets, which makes the validation process difficult due to language differences. So that in this research, the recognition or detection of emotions in the voice will be carried out using Indonesian-language datasets.

Mel-Frequency Cepstral Coefficients (MFCC) is a commonly used feature extraction method for emotion recognition in sound because it has a working principle that resembles the human ear. This working principle makes MFCC able to overcome most of the research related to recognition based on sound signals. There are several processes that occur in the MFCC method, namely Pre-emphasize, Frame Blocking, Windowing, Fast Fourier Transform, Mel-scaled Filterbank, and Discrete Cosine Transform. K-Nearest Neighbor (KNN) is a supervised machine learning algorithm that is widely used to solve classification and detection problems due to its robustness against noise in data train. The data used in this research is podcast sound data downloaded from several Youtube channels.

The tests carried out in this study are tuning parameters and testing the ability of the model to detect emotions on new data. In addition, testing between men and women is done separately because men and women have different voice characters. The combination of MFCC feature extraction method and KNN method is able to detect emotions based on voice intonation in men with an accuracy of 88.8% on test data and 80% on new data. MFCC and KNN are also able to detect emotions based on voice intonation in women with 92.5% accuracy on test data but on new data can only achieve 50% accuracy. The model in women experienced an overfitting condition caused by pitch shifting augmentation data that was unable to generate enough data samples to accurately represent all possible input data values.

Keywords: Mel-Frequency Cepstral Coefficient, K-Nearest Neighbor, Emotion detection, Signal processing.