

## ABSTRAK

Suara tidak dapat langsung diolah oleh mesin tanpa adanya proses ekstraksi ciri yang dilakukan terlebih dahulu. Saat ini terdapat begitu banyak pilihan metode ekstraksi ciri yang dapat digunakan, sehingga menentukan metode ekstraksi ciri yang tepat merupakan hal yang tidak mudah. Salah satu metode ekstraksi ciri pada sinyal suara yang kerap kali digunakan yaitu *Mel-Frequency Cepstral Coefficient* (MFCC). MFCC memiliki prinsip kerja menyerupai sistem pendengaran manusia yang menyebabkan banyak digunakan pada berbagai tugas yang berhubungan dengan pengenalan berdasarkan sinyal suara. Penelitian ini akan menggunakan metode MFCC untuk mengekstrak ciri pada sinyal suara dan *Support Vector Machine* sebagai metode klasifikasi emosi pada dataset RAVDESS.

MFCC terdiri dari beberapa tahap yaitu *Pre-emphasize*, *Frame Blocking*, *Windowing*, *Fast Fourier Transform*, *Mel-Scaled Filterbank*, *Discrete Cosine Transform*, dan *Cepstral Liftering*. Jenis rancangan pengujian yang akan dilakukan pada penelitian ini adalah *parameter tuning*. *Parameter tuning* dilakukan dengan tujuan untuk memperoleh parameter-parameter yang menghasilkan akurasi terbaik pada model *machine learning*. Parameter-parameter yang akan di-*tuning* antara lain, nilai  $\alpha$  pada proses *Pre-Emphasis*, panjang *frame* beserta panjang *overlap* pada proses *Frame Blocking*, jumlah filter mel pada proses *Mel-Scaled Filterbank*, jumlah *cepstral coefficient* pada proses *Discrete Cosine Transform* dan nilai C pada SVM.

Akurasi terbaik pada laki-laki sebesar 85,71% diperoleh dengan kombinasi parameter filter *pre-emphasize* sebesar 0.95, panjang *frame* sebesar 0.023 ms, *overlap frame* yang bersebelahan sebesar 40%, jumlah filter *mel* pada proses *mel-scaled filterbank* sebesar 24 mel, jumlah *cepstral coefficient* sebesar 24 *coefficient* dan nilai 'C' pada SVM sebesar 0,01. Akurasi terbaik pada wanita sebesar 92,21% diperoleh dengan kombinasi parameter filter *pre-emphasize* sebesar 0.95, panjang *frame* sebesar 0.023 ms, *overlap frame* yang bersebelahan sebesar 40%, jumlah filter *mel* pada proses *mel-scaled filterbank* sebesar 24 mel, dan jumlah *cepstral coefficient* sebesar 13 *coefficient* dan nilai 'C' pada SVM sebesar 0,01. Dari kedua hasil pengujian parameter *tuning* antara laki-laki dan wanita terdapat kesamaan nilai parameter pada semua parameter pengujian, kecuali pada jumlah *cepstral coefficient*. Jumlah *cepstral coefficient* pada laki-laki sebanyak 24 *coefficient* sedangkan jumlah *cepstral coefficient* pada wanita sebanyak 13 *coefficient*.

Berdasarkan penelitian yang dilakukan, terdapat kesimpulan sebagai berikut, kombinasi metode MFCC dengan SVM mampu digunakan untuk klasifikasi emosi berdasarkan data masukan berupa intonasi suara dengan akurasi sebesar 85,71% pada laki-laki dan 92,21% pada wanita. Perbedaan akurasi yang didapat antara model laki-laki dan wanita disebabkan karena data yang digunakan berbeda. Model laki-laki dilatih dengan data suara laki-laki dan model wanita dilatih dengan data suara wanita, hal ini dilakukan karena laki-laki dan wanita memiliki rentang frekuensi suara yang berbeda.

**Kata kunci:** *Mel-Frequency Cepstral Coefficient*, *Support Vector Machine*, **Klasifikasi emosi, Pemrosesan sinyal**

## ABSTRACT

*Sounds cannot be directly processed by machines without a feature extraction process being carried out first. Currently, there are so many choices of feature extraction methods that can be used, so determining the right feature extraction method is not easy. One method of feature extraction on sound signals that is often used is Mel-Frequency Cepstral Coefficient (MFCC). MFCC has a working principle that resembles the human hearing system, which causes it to be widely used in various tasks related to recognition based on sound signals. This research will use the MFCC method to extract characteristics on voice signals and Support Vector Machine as a method of emotion classification on the RAVDESS dataset.*

*MFCC consists of several stages, namely Pre-emphasize, Frame Blocking, Windowing, Fast Fourier Transform, Mel-Scaled Filterbank, Discrete Cosine Transform, and Cepstral Liftering. The type of test design that will be carried out in this research is parameter tuning. Parameter tuning is carried out with the aim of obtaining parameters that produce the best accuracy in the machine learning model. The parameters that will be tuned include the  $\alpha$  value in the Pre-Emphasis process, frame length and overlap length in the Frame Blocking process, the number of mel filters in the Mel-Scaled Filterbank process, the number of cepstral coefficients in the Discrete Cosine Transform process and the C value in SVM.*

*The best accuracy in males of 85.71% was obtained with a combination of filter parameter pre-emphasize of 0.95, frame length of 0.023 ms, overlap of adjacent frames of 40%, number of mel filters in the mel-scaled filterbank process of 24 mel, number of cepstral coefficient of 24 coefficient and the value of 'C' in SVM of 0.01. The best accuracy in women of 92.21% was obtained with a combination of filter parameters pre-emphasize of 0.95, frame length of 0.023 ms, overlap of adjacent frames of 40%, the number of mel filters in the mel-scaled filterbank process of 24 mel, and the number of cepstral coefficient of 13 coefficient and 'C' value in SVM of 0.01. From the two test results of tuning parameters between men and women, there are similar parameter values in all test parameters, except for the number of cepstral coefficients. The number of cepstral coefficient in men is 24 coefficient while the number of cepstral coefficient in women is 13 coefficient.*

*Based on the research conducted, there are the following conclusions, the combination of MFCC and SVM methods can be used for emotion classification based on input data in the form of voice intonation with an accuracy of 85.71% in men and 92.21% in women. The difference in accuracy obtained between male and female models is due to the different data used. Male models are trained with male voice data and female models are trained with female voice data, this is done because men and women have different voice frequency ranges.*

***Keywords: Mel-Frequency Cepstral Coefficient, Support Vector Machine, Classification Emotion, Signal Processing***