

ABSTRAK

Emosi berperan penting dalam kehidupan manusia. Speech Emotion Recognition (SER) adalah topik penelitian yang berkembang pesat dan diterapkan di berbagai bidang, termasuk telekomunikasi untuk meningkatkan layanan customer service. Kepuasan pelanggan sangat penting bagi kesuksesan perusahaan, sehingga penting adanya sistem yang mendeteksi emosi secara otomatis melalui suara. Klasifikasi merupakan solusi praktis untuk permasalahan tersebut. Sebelumnya sudah ada banyak penelitian mengenai deteksi emosi akan tetapi prosesnya hanyalah ekstraksi fitur dan klasifikasi saja. Sehingga peneliti berinisiatif untuk menambahkan reduksi fitur menggunakan *Principal Component Analysis* (PCA) untuk mendapatkan fitur yang lebih baik. Ekstraksi fitur yang digunakan adalah *Mel Frequency Cepstral Coefficient* (MFCC), *Zero Crossing Rate* (ZCR), *Root Mean Square Energy* (RMSE), *Spectral Centroid*, dan *Spectral Rolloff*. *Random Forest* dipilih sebagai metode klasifikasi yang akan digunakan dalam penelitian ini dikarenakan *Random Forest* dapat beradaptasi dengan fitur yang kompleks. Penelitian ini menggunakan dataset dari *The Ryerson Audio-Visual Dataset of Emotional Speech and Song* (RAVDESS), berisi 24 aktor (12 laki-laki dan 12 perempuan) dari Amerika Serikat berdurasi 3-4 detik per data, Dimana tiap aktor menyuarakan 60 data sehingga didapat total 1440 data dengan 8 kelas emosi yaitu netral, tenang, senang, sedih, marah, takut, jijik, dan kaget. Akurasi tertinggi tanpa dilakukan reduksi dimensi diperoleh sebesar 66.67%. Kemudian akurasi tertinggi dengan reduksi dimensi diperoleh sebesar 70.83% dimana $n_estimator$ 150 dengan max_depth 100 dan memiliki 54 fitur yang direduksi menjadi 24 fitur. Sehingga adanya reduksi dimensi PCA membuat peningkatan akurasi sebesar 4.16%.

Kata Kunci : *Speech Emotion Recognition*, Klasifikasi, *Principal Component Analysis*, *Random Forest*

ABSTRACT

Emotions play an important role in human life. Speech Emotion Recognition (SER) is a rapidly growing research topic and is applied in various fields, including telecommunications to improve customer service. Customer satisfaction is very important for the success of a company, so it is important to have a system that automatically detects emotions through voice. Classification is a practical solution to this problem. Previously, there have been many studies on emotion detection, but the process was only feature extraction and classification. So the researcher took the initiative to add feature reduction using Principal Component Analysis (PCA) to get better features. The feature extraction used is Mel Frequency Cepstral Coefficient (MFCC), Zero Crossing Rate (ZCR), Root Mean Square Energy (RMSE), Spectral Centroid, and Spectral Rolloff. Random Forest was chosen as the classification method to be used in this study because Random Forest can adapt to complex features. This study uses a dataset from The Ryerson Audio-Visual Dataset of Emotional Speech and Song (RAVDESS), containing 24 actors (12 males and 12 females) from the United States with a duration of 3-4 seconds per data, where each actor voices 60 data so that a total of 1440 data are obtained with 8 classes of emotions, namely neutral, calm, happy, sad, angry, afraid, disgusted, and surprised. The highest accuracy without dimensionality reduction is obtained at 66.67%. Then the highest accuracy with dimensionality reduction is obtained at 70.83% where $n_estimator$ 150 with max_depth 100 and has 54 features that are reduced to 24 features. So that the PCA dimensionality reduction increases accuracy by 4.16%.

Keywords : *Speech Emotion Recognition, Classification, Principal Component Analysis, Random Forest*