1. Pendahuluan

Speech Emotion Recognition (SER) merupakan suatu cara untuk mempelajari formasi dan perubahan keaadan emosi seseorang dari sinyal ucapan perspektif sehingga komputer bisa memiliki kecerdasan dalam melakukan interaksi dengan manusia [1]. *SER* juga salah satu teknologi yang cukup menantang sebagai media untuk penelitian karena pada *SER* ini terkadang memiliki keterbatasan dalam *training data* sehingga mengakibatkan akurasi prediksi yang rendah [1].

Permasalahan yang diangkat pada penelitian ini adalah kekurangan data. Permasalahan tersebut merupakan permasalahan yang sering dialami pada suatu penelitian yang mengakibatkan akurasi dari suatu percobaan tidak optimal. Untuk penelitian ini, penulis menggunakan 4 kategori emosi yang terdapat pada dataset Berlin Emo-DB [2]. 4 kategori emosi tersebut memiliki total 339 data. Dari jumlah data tersebut dirasa masih memiliki jumlah yang sangat minim untuk dijadikan suatu objek penelitian. Untuk menyelesaikan masalah tersebut, beberapa metode augmentasi data akan digunakan. Beberapa metode juga digunakan pada beberapa penelitian untuk menyelesaikan masalah kekurangan data. Etienne [3] menggunakan metode augmentasi data pada penelitiannya yang menggunakan dataset IEMOCAP dengan melakukan *rescaling* terhadap spectrogram. Niu [1] juga menggunakan augmentasi data yang mirip dengan Etienne [3].

Pada suatu penelitian juga melibatkan metode yang akan digunakan pada prroses klasifikasi. Berdasarkan metode klasifikasi untuk SER dapat dibedakan menjadi 2 kategori yaitu *traditional machine learning* dan *deep learning* [1].

Beberapa metode *traditional machine learning* yang sering digunakan pada beberapa penelitian adalah *Hidden Markov Model* [4], *Gaussian Mixture Model* [5], *Artificial Neural Network* [6] dan *Support Vector Machine* [7] [8] [6]. Di sisi lain, *deep learning* adalah metode *machine learning* yang dirancang untuk tugas tertentu [9]. Dalam SER, beberapa peneliti telah menerapkan metode *deep learning*. Sebagai contoh, Vladimir [10] dalam penelitiannya menerapkan *Recurrent Neural Network (RNN)* pada SER dengan mendapatkan nilai akurasi sebesar 70%. Sementara itu, Niu [1] pada penelitiannya menggunakan Deep-CNN yang menghasilkan nilai akurasi sebesar 99%, dimana pada penelitian tersebut menggunakan augmentasi data dengan cara *rescaling* pada spectrogram. Metode augmentasi yang digunakan telah menambah jumlah data dari 7.527 data menjadi 310.067 data dan mampu meningkatkan nilai akurasi secara signifikan dari 41,54% menjadi 99%