

MUHAMMAD SEPTIYANTO, 22.230.0013

**KLASIFIKASI EMOSI SUARA BAHASA INDONESIA MENGGUNAKAN
MFCC DAN BiLSTM DENGAN AUGMENTASI AUDIO, dibawah**

bimbingan Eko Budi Susanto, M.Kom. dan Devi Sugianti, M.Kom.

65 + xiii halaman / 24 gambar / 7 tabel / 20 lampiran / 31 pustaka (2019-2025)

ABSTRAK

Klasifikasi emosi suara menjadi teknologi penting dalam era kecerdasan buatan modern, namun penelitian untuk bahasa Indonesia masih terbatas. Penelitian ini menggunakan kombinasi ekstraksi fitur Mel-Frequency Cepstral Coefficients (MFCC) dan model Bidirectional Long Short-Term Memory (BiLSTM) dengan teknik augmentasi audio untuk klasifikasi emosi suara bahasa Indonesia. Dataset yang digunakan IndoWaveSentiment berisi 300 rekaman audio dari 10 responden dengan lima kelas emosi: netral, bahagia, terkejut, jijik, dan kecewa. Teknik augmentasi audio dengan rasio 2:1 menggunakan lima metode menghasilkan 900 sampel. Ekstraksi fitur MFCC menghasilkan 40 koefisien yang diproses menggunakan arsitektur BiLSTM dengan dua lapisan bidirectional (256 dan 128 unit). Model dilatih menggunakan Adam optimizer dengan early stopping. Hasil penelitian menunjukkan akurasi tertinggi 93.33% dengan precision 93.7%, recall 93.3%, dan F1-score 93.3%. Emosi "terkejut" mencapai performa sempurna (100%), sedangkan "bahagia" memiliki akurasi terendah (88.89%). Perbandingan dengan penelitian sebelumnya yang menggunakan Random Forest mencapai 90% dan Gradient Boosting 85%. Penelitian ini menunjukkan efektivitas kombinasi MFCC, BiLSTM, dan augmentasi audio dalam menangkap karakteristik emosi suara bahasa Indonesia untuk pengembangan sistem pengenalan emosi berbasis suara.

Kata kunci: *Augmentasi Audio, BiLSTM, Deep Learning, Klasifikasi Emosi, MFCC*

MUHAMMAD SEPTIYANTO, 22.230.0013

**INDONESIA SPEECH EMOTION CLASSIFICATION USING MFCC AND
BiLSTM WITH AUDIO AUGMENTATION**, dibawah bimbingan Eko Budi Susanto,
M.Kom. dan Devi Sugianti, M.Kom.

65 + xiii halaman / 24 gambar / 7 tabel / 20 lampiran / 31 daftar pustaka (2019-2025)

ABSTRACT

Emotion classification from speech has become an important technology in the modern artificial intelligence era, however research for the Indonesian language is still limited. This study uses a combination of Mel-Frequency Cepstral Coefficients (MFCC) feature extraction and Bidirectional Long Short-Term Memory (BiLSTM) model with audio augmentation techniques for Indonesian speech emotion classification. The IndoWaveSentiment dataset contains 300 audio recordings from 10 respondents with five emotion classes: neutral, happy, surprised, disgusted, and disappointed. Audio augmentation techniques with a 2:1 ratio using five methods generated 900 samples. MFCC feature extraction produced 40 coefficients that were processed using BiLSTM architecture with two bidirectional layers (256 and 128 units). The model was trained using Adam optimizer with early stopping. Research results show the highest accuracy of 93.33% with precision of 93.7%, recall of 93.3%, and F1-score of 93.3%. The "surprised" emotion achieved perfect performance (100%), while "happy" had the lowest accuracy (88.89%). Previous comparisons used Random Forest (90%) and Gradient Boosting (85%). This study demonstrates the effectiveness of combining MFCC, BiLSTM, and audio augmentation in capturing Indonesian speech emotion characteristics for the development of voice-based emotion recognition systems.

Keywords: *Augmentation Audio, BiLSTM, Deep Learning, Emotion
Classification, MFCC*