

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model *voice emotion recognition* berbasis *Convolutional Neural Network* (CNN) guna mendukung interaksi *customer service* dalam mengenali emosi pelanggan melalui suara. Penelitian menggunakan dataset *IndoWaveSentiment* yang terdiri dari 300 rekaman audio berbahasa Indonesia dan telah dilabeli ke dalam lima kategori emosi: marah, senang, netral, sedih, dan terkejut. Metodologi yang digunakan meliputi pra-pemrosesan data (konversi mono, normalisasi, *trimming*, *padding*, dan augmentasi), ekstraksi fitur menggunakan *Mel-Frequency Cepstral Coefficients* (MFCC), serta pembagian data menjadi data pelatihan dan data pengujian. Model CNN dirancang untuk mengekstraksi pola spasial dari fitur audio dan mengklasifikasikan emosi secara otomatis. Hasil evaluasi menunjukkan bahwa model mencapai akurasi 84% pada data uji, dengan nilai *precision*, *recall*, dan *f1-score* rata-rata sebesar 84%. Kinerja terbaik diperoleh pada kategori netral dengan *recall* 95% dan *f1-score* 88%, sedangkan pada kategori marah dan senang, *recall* masing-masing sebesar 76% dan 77%. Hasil ini membuktikan bahwa model CNN efektif dalam mendeteksi emosi utama pada layanan customer service Berbahasa Indonesia.

Kata kunci: *audio, convolutional neural network, customer service, voice emotion recognition*

ABSTRACT

This study aims to develop and evaluate a voice emotion recognition model based on Convolutional Neural Network (CNN) to support customer service interactions in recognizing customer emotions through speech. The research uses the IndoWaveSentiment dataset, which consists of 300 Bahasa audio recordings labeled with five emotion categories: angry, happy, neutral, sad, and surprised. The methodology includes data pre-processing (mono conversion, normalization, trimming, padding, and augmentation), feature extraction using Mel-Frequency Cepstral Coefficients (MFCC), and splitting data into training and testing sets. The CNN model is designed to extract spatial patterns from audio features and classify emotions automatically. Model performance is evaluated using accuracy, confusion matrix, precision, recall, and f1-score. The results show that the model achieves an accuracy of 84% on the test set, with average precision, recall, and f1-score of 84%. The best performance is observed in the neutral category, with a recall of 95% and f1-score of 88%, while the angry and happy categories have lower recall values of 76% and 77%, respectively. These findings demonstrate that the CNN-based model is effective in detecting primary emotions from speech in customer service contexts.

Kata Kunci: *audio, convolutional neural network, customer service, voice emotion recognition*



KARAWANG