

ABSTRAK

Emosi merupakan aspek penting dalam komunikasi verbal yang dapat memengaruhi makna dan respons suatu percakapan. Namun, pengenalan emosi dari suara masih menjadi tantangan, terutama ketika hanya mengandalkan karakteristik akustik. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem deteksi emosi berbasis teks hasil transkripsi suara dengan memanfaatkan metode *Speech-to-Text* dan *Natural Language Processing*. Sistem yang dikembangkan menggunakan model Whisper untuk mengonversi suara menjadi teks, kemudian menerapkan pendekatan *lexicon-based* dengan kamus mini berbobot untuk menentukan kategori emosi. Proses yang diterapkan meliputi *preprocessing audio*, konversi suara ke teks, normalisasi teks, dan analisis emosi. Sistem mampu menerima dua jenis input, yaitu teks manual dan suara real-time, sehingga memberikan fleksibilitas dalam penggunaannya. Evaluasi performa sistem dilakukan menggunakan *Word Error Rate* (WER) untuk mengukur kualitas transkripsi, serta *Precision*, *Recall*, dan *F1-score* untuk menilai akurasi deteksi emosi. Hasil pengujian menunjukkan bahwa sistem memiliki performa yang cukup baik dan stabil dengan rata-rata F1-score sebesar 86,8% dan nilai WER sebesar 4,7%, yang berada di atas standar umum kualitas transkripsi. Berdasarkan hasil tersebut, dapat disimpulkan bahwa pendekatan *lexicon-based* yang didukung oleh teknologi *Speech-to-Text Whisper* mampu menjadi solusi yang efektif, ringan, dan kompetitif dalam mendeteksi emosi berbasis teks pada sistem web interaktif.

Kata kunci: Deteksi emosi, *speech-to-text*, *natural language processing*, *lexicon-based*, *Whisper*

ABSTRACT

Emotion plays an important role in verbal communication, as it influences the meaning and interpretation of spoken interactions. However, recognizing emotion from speech remains a challenge, particularly when relying solely on acoustic features. This study aims to design and implement a text-based emotion detection system derived from speech transcription by utilizing Speech-to-Text and Natural Language Processing techniques. The proposed system employs the Whisper model to convert speech into text and applies a lexicon-based approach with a weighted mini-lexicon to determine emotion categories. The system workflow consists of audio preprocessing, speech-to-text conversion, text normalization, and emotion analysis. The system supports two types of input, namely manual text input and real-time speech input, providing flexibility for various usage scenarios. System performance is evaluated using Word Error Rate (WER) to measure transcription quality, and Precision, Recall, and F1-score to assess emotion classification accuracy. Experimental results indicate that the system achieves stable and satisfactory performance, with an average F1-score of 86.8% and a WER of 4.7%, which exceeds the commonly accepted transcription quality standard. These results demonstrate that a lexicon-based approach, supported by Whisper Speech-to-Text technology, can serve as an effective, lightweight, and competitive solution for text-based emotion detection in interactive web-based systems.

Keywords: *emotion detection, speech-to-text, natural language processing, lexicon-based, Whisper*