ABSTRAK

DETEKSI SUARA KECERDASAN BUATAN BERBAHASA INDONESIA DENGAN *CROSS-LINGUAL SPEECH REPRESENTATION* WAV2VEC2 DAN *CONVOLUTIONAL NEURAL NETWORK*

Oleh:

Ahmad Juaeni Yunus

1227050011

Perkembangan teknologi *voice cloning* dan *deepfake* audio berbasis kecerdasan buatan telah menimbulkan ancaman nyata terhadap keamanan digital di Indonesia, di mana kerugian akibat penipuan berbasis suara sintetis telah mencapai Rp7,8 triliun sejak akhir 2024, sementara penelitian deteksi suara AI masih sangat terbatas untuk Bahasa Indonesia sebagai *low-resource language*. Penelitian ini bertujuan mengembangkan dan mengevaluasi model deteksi suara AI berbahasa Indonesia (Indo-AD) yang mampu membedakan suara asli manusia dari suara sintetis secara otomatis. Metodologi yang digunakan mengikuti kerangka CRISP-DM, dengan dataset berupa rekaman suara asli dan suara buatan dari platform ElevenLabs, FishAudio, dan MinimaxAudio, diproses menggunakan model *Self-Supervised Learning* berbasis XLSR-Wav2Vec2 sebagai ekstraktor representasi akustik yang dikombinasikan dengan *Convolutional Neural Network* (CNN) sebagai *classifier* dua kelas, dan dievaluasi dengan skema *Leave-One-Speaker-Out* (LOSO) untuk menguji kemampuan generalisasi. Hasil eksperimen menunjukkan bahwa model Indo-AD mencapai akurasi dan F1-score sebesar 90,5%, membuktikan bahwa pendekatan *Self-Supervised Learning* efektif mengekstraksi pola akustik kompleks dari sinyal audio mentah, dengan potensi implementasi sebagai sistem keamanan digital berbasis audio untuk mendeteksi *deepfake* suara berbahasa Indonesia.

Kata kunci: *Deepfake audio*, *Self-Supervised Learning*, XLSR-Wav2Vec2, CNN, deteksi suara AI, Bahasa Indonesia.

ABSTRACT

ARTIFICIAL INTELLIGENCE VOICE DETECTION IN INDONESIAN USING CROSS-LINGUAL SPEECH REPRESENTATION WAV2VEC2 AND CONVOLUTIONAL NEURAL NETWORK

By:

Ahmad Juaeni Yunus

1227050011

Advances in AI-based voice cloning and deepfake audio technology have posed a real threat to digital security in Indonesia, where losses from synthetic voice-based fraud have reached Rp7.8 trillion since the end of 2024, while research on AI voice detection remains very limited for Indonesian as a low-resource language. This study aims to develop and evaluate an Indonesian-language AI voice detection model (Indo-AD) capable of automatically distinguishing authentic human voices from synthetic ones. The methodology follows the CRISP-DM framework, utilizing a dataset comprising authentic and synthetic voice recordings from the ElevenLabs, FishAudio, and MinimaxAudio platforms, processed using a Self -Supervised Learning model based on XLSR-Wav2Vec2 as an acoustic representation extractor, combined with a Convolutional Neural Network (CNN) as a binary classifier, and evaluated using the Leave-One-Speaker-Out (LOSO) scheme to test generalization capabilities. Experimental results show that the Indo-AD model achieves an accuracy and F1-score of 90.5%, proving that the Self-Supervised Learning approach is effective at extracting complex acoustic patterns from raw audio signals, with the potential for implementation as an audio-based digital security system to detect Indonesian-language voice deepfakes.

Keywords*: Audio deepfake, Self-Supervised Learning, XLSR-Wav2Vec2, CNN, AI voice detection, Indonesian language.*