ABSTRAK

Penelitian ini bertujuan untuk mengimplementasikan metode Long Short-Term Memory (LSTM) dalam menganalisis pola temporal pada data audio. Selain itu, ekstraksi fitur Log-Mel Spectrogram digunakan untuk merepresentasikan karakteristik spasial data audio guna meningkatkan performa model dalam klasifikasi suara derau lingkungan.

Penelitian dilakukan menggunakan dataset UrbanSound8k yang terdiri dari 8732 data audio yang dikelompokkan ke dalam 10 kelas suara lingkungan. Ekstraksi fitur dilakukan menggunakan Log-Mel Spectrogram, sementara model klasifikasi dibangun menggunakan metode LSTM dengan beberapa lapisan yang dioptimalkan. Pengujian hyperparameter dilakukan untuk menentukan kombinasi dropout rate dan learning rate terbaik yang mendukung kinerja model secara keseluruhan.

Hasil penelitian menunjukkan bahwa kombinasi dropout rate 0.20 dan learning rate 1e-3 menghasilkan performa terbaik. Akurasi pelatihan mencapai 97.60%, dengan categorical accuracy pengujian sebesar 88.44%, serta nilai evaluasi akurasi 88.49%, precision 89.14%, recall 88.75%, dan F1-score 88.87%. Pendekatan ini terbukti lebih efektif dalam mengenali pola temporal suara dibandingkan penelitian sebelumnya yang lebih banyak berfokus pada fitur spasial.

Kata Kunci : LSTM, Klasifikasi, Audio

ABSTRACT

This study aims to implement the Long Short-Term Memory (LSTM) method to analyze the temporal patterns in audio data. Additionally, Log-Mel Spectrogram feature extraction is utilized to represent the spatial characteristics of audio data, aiming to enhance the model's performance in classifying environmental noise.

The study was conducted using the UrbanSound8k dataset, which consists of 8,732 audio samples categorized into 10 environmental noise classes. Feature extraction was carried out using Log-Mel Spectrogram, while the classification model was built using LSTM with several optimized layers. Hyperparameter tuning was performed to determine the best combination of dropout rate and learning rate to improve the overall model performance.

The results showed that the combination of a 0.20 dropout rate and a learning rate of 1e-3 produced the best performance. Training accuracy reached 97.60%, with a categorical testing accuracy of 88.44%, and evaluation metrics yielded an accuracy of 88.49%, precision of 89.14%, recall of 88.75%, and F1-score of 88.87%. This approach proved to be more effective in recognizing temporal patterns in environmental noise compared to previous studies that primarily focused on spatial features.

Keywords : LSTM, Classification, Audio