

ABSTRAK

Keandalan pengenalan suara pada *smart speaker* menjadi prasyarat penting bagi interaksi manusia–komputer yang natural, namun pada praktiknya kinerja *voice recognition* dapat menurun ketika perangkat digunakan pada variasi jarak dan kondisi konektivitas yang berbeda. Penelitian ini bertujuan mengevaluasi performa *voice recognition* secara komparatif pada Amazon Alexa dan *Interactive Speaker System* serta menguji keterkaitan variasi jarak pengguna terhadap keberhasilan pengenalan perintah suara. Pengujian dilakukan melalui eksperimen langsung dengan skenario pemberian perintah suara pada jarak 1 m, 2 m, dan 3 m. Data yang dikumpulkan berjumlah 1020 percobaan, dengan 715 percobaan dikategorikan gagal dan 305 percobaan dikategorikan berhasil. Untuk memperoleh representasi yang lebih informatif, data multimodal yang mencakup indikator audio, jaringan, lokasi, dan sinyal Wi-Fi diolah melalui ekstraksi fitur statistik, kemudian dimodelkan menggunakan Random Forest untuk klasifikasi kondisi berhasil/gagal. Hasil analisis menunjukkan adanya kecenderungan degradasi sinyal audio seiring meningkatnya jarak, yang berimplikasi pada penurunan probabilitas keberhasilan pengenalan suara. Model Random Forest memberikan performa yang sangat baik dengan nilai *accuracy*, *F1-macro*, dan *recall-macro* masing-masing 0,98. Temuan *feature importance* menegaskan dominasi fitur audio—terutama karakteristik distribusi dan variabilitas level suara—dibandingkan fitur jaringan, lokasi, maupun Wi-Fi. Secara keseluruhan, penelitian ini menegaskan bahwa kualitas dan kestabilan sinyal audio merupakan determinan utama performa *voice recognition* serta menunjukkan Random Forest efektif untuk evaluasi berbasis data multimodal pada skenario penggunaan nyata.

Kata Kunci: *Voice Recognition*, *Smart Speaker*, Random Forest, Evaluasi Performa, Data Multimodal

ABSTRACT

The reliability of voice recognition is a critical factor in enabling natural human–computer interaction through smart speakers; however, its performance in real-world usage is often affected by variations in user distance and connectivity conditions. This study aims to comparatively evaluate the performance of voice recognition on Amazon Alexa and an Interactive Speaker System, as well as to examine the influence of user distance on recognition success. Experimental testing was conducted by issuing voice commands at distances of 1 m, 2 m, and 3 m. A total of 1,020 trials were collected, consisting of 715 failed trials and 305 successful trials. Multimodal data comprising audio, network, location, and Wi-Fi indicators were processed through statistical feature extraction to obtain representative predictors, which were subsequently analyzed using a Random Forest algorithm to classify recognition outcomes. The results indicate a clear tendency of audio signal degradation as user distance increases, leading to a reduced probability of successful voice recognition. The Random Forest model demonstrated excellent classification performance, achieving accuracy, macro F1-score, and macro recall values of 0.98. Feature importance analysis reveals that audio-related features—particularly those describing the distribution and variability of sound levels—contribute most significantly to the model, exceeding the influence of network, location, and Wi-Fi features. Overall, the findings confirm that audio quality and signal stability are the primary determinants of voice recognition performance and demonstrate the effectiveness of Random Forest as an analytical approach for evaluating multimodal smart speaker performance under real-world conditions.

Keywords: Voice Recognition, Smart Speaker, Random Forest, Performance Evaluation, Multimodal Data