

Abstrak

Sistem pengenalan ucapan dalam bahasa Indonesia yang tersedia menggunakan Hidden Markov Model (HMM) sebagai acoustic model. Penelitian sebelumnya menunjukkan adanya terobosan terbaru terhadap perkembangan acoustic model yaitu membandingkan hasil performansi model deep learning dan model HMM pada lingkungan yang bising. Model deep learning ini tidak memerlukan adanya kamus fonem/silabel dan konsep fonem/silabel. Pembangunan sistem pengenalan ucapan dalam bahasa Indonesia menggunakan engine Automatic Speech Recognition (ASR) Mozilla DeepSpeech. Hasil yang didapatkan berdasarkan pengujian beberapa skenario training dan testing menunjukan bahwa pengujian 10-fold cross validation menghasilkan nilai rata-rata word error rate (WER) 1.78% lebih rendah dibandingkan 4-fold cross validation, dengan hasil nilai rata-rata WER 7.32% serta pengujian inference test dan live streaming dilakukan dengan parameter tertentu menghasilkan performansi yang baik.

Kata kunci : deep learning, deep speech, ASR, WER, HMM, live streaming