

Abstract

Indonesian speech recognition system which are available uses Hidden Markov Model (HMM) as acoustic model. Previous research shows the latest breakthroughs against development of acoustic model by comparing the performances results between deep learning model and HMM model in noisy environments. This deep learning model does not require a dictionary of phonemes/ syllables and concepts of phonemes/syllables. Development of speech recognition systems in Indonesian using engine Automatic Speech Recognition (ASR) Mozilla DeepSpeech. The results obtained are based on testing several training and testing scenarios show that the 10-fold cross validation test produces an average word error rate (WER) of 1.78% lower than the 4-fold cross validation, with a WER average value of 7.32% and live streaming and inference testing is done with certain parameters resulting in good performance.

Keywords: deep learning, deep speech, ASR, WER, HMM, live streaming