

1. Pendahuluan [10 pts/Bold]

Latar Belakang

Penelitian sistem pengenalan ucapan dalam lima tahun terakhir menunjukkan bahwa pembangunan sistem pengenalan ucapan yang terdiri dari beberapa algoritma dan tahapan pemrosesan tersebut dapat tergantikan dengan model Deep Learning yang dapat disebut dengan "Deep Speech". Deep Speech merupakan sistem end-to-end yang dapat dilatih pada tingkat karakter. Penggunaan Recurrent Neural Network (RNN) yang terdapat pada deep speech merupakan metode yang dianggap sebagai pengembangan terhadap acoustic model yang terdapat pada sistem pengenalan ucapan menggunakan metode Hidden Markov Model (HMM) yang dianggap model tradisional sehingga performansi yang didapat melampaui performansi penggunaan model [tradisional\[14\]](#).

Sistem pengenalan ucapan yang menggunakan model tradisional memiliki faktor yang mempengaruhi performansi, antara lain: (i) tingkat kebisingan dan (ii) keterkaitan konsep kamus fonem yang menyebabkan kata yang [terbatas\[14\]](#). Sebaliknya sistem pengenalan ucapan dengan model deep learning mampu menghilangkan konsep kamus yang membatasi suatu kata dan mampu menangani pengenalan suara pada lingkungan bising tersebut. Penelitian tentang algoritma deep learning semakin berkembang pesat terutama untuk masalah pengenalan suara ini. Pencapaian tersebut memungkinkan untuk meningkatkan performansi pengenalan suara pada lingkungan bising dan menghilangkan konsep kamus fonem/silabel. Mozilla DeepSpeech merupakan salah satu proyek deep speech yang bersifat open source berdasarkan penelitian milik "Baidu's Deep Speech Research".

Dalam penelitian ini proyek deep speech dari Mozilla akan digunakan sebagai engine Automatic Speech Recognition (ASR) dalam sistem pengenalan ucapan dalam bahasa Indonesia..

Topik dan Batasannya

Pada umumnya pengenalan teks kalimat bahasa Indonesia menggunakan acoustic model dengan metode Hidden Markov Model (HMM), sistem end-to-end dengan metode deep speech merupakan pengembangan acoustic model berbasis metode HMM yang mengatasi permasalahan yang dimiliki oleh metode HMM sehingga tingkat error yang dihasilkan lebih kecil.

Pembangunan sistem pengenalan ucapan dalam bahasa Indonesia menggunakan engine ASR Mozilla, data berupa speech corpus beserta teks corpus dalam bahasa Indonesia, perangkat central processing unit (CPU) untuk tahap testing dan perangkat graphic processing unit (GPU) dengan VRAM 8Gb untuk tahap training.

Tujuan

Mengembangkan sistem pengenalan dalam bahasa Indonesia dengan menggunakan Mozilla DeepSpeech dengan model deep learning yang memiliki tingkat performansi lebih tinggi dibandingkan dengan model tradisional maka dilakukan pembuktian dengan beberapa skenario pengujian yang telah ditentukan, mengetahui hasil prediksi pada lingkungan bising, mengetahui beberapa nilai parameter yang mempengaruhi nilai performansi dan mengetahui pengaruh kosakata ejaan lama terhadap model yang dilatih.

Organisasi Tulisan

Penulisan Tugas Akhir dibagi menjadi lima bagian. Bagian 1 menjelaskan permasalahan, topik dan batasan dan solusi yang ditawarkan. Bagian 2 menjelaskan referensi atau studi terkait yang digunakan dalam menulis Tugas Akhir. Bagian 3 berisi tentang rincian model yang dibangun untuk menyelesaikan permasalahan dan data yang digunakan dalam eksperimen. Bagian 4 menyajikan hasil eksperimen pengujian dan analisis pengujian terhadap model yang telah dibangun. Bagian 5 merupakan kesimpulan dan saran yang diambil dari eksperimen yang telah dilakukan dan saran untuk peneliti selanjutnya.