BAB 1
Pendahuluan

1.1 Latar Belakang

Bahasa adalah alat komunikasi yang terorganisasi dalam bentuk satuan-satuan, seperti kata, kelompok kata, klausa, dan kalimat yang diungkapkan baik secara lisan maupun tulis[1].

Tantangan mendasar untuk penelitian saat ini tentang sains dan teknologi bicara adalah memahami dan memodelkan variasi individu dalam bahasa lisan. Individu memiliki gaya bicara mereka sendiri, tergantung pada banyak faktor, seperti dialek dan aksen mereka serta latar belakang sosial ekonomi mereka[2].

Di Indonesia terdapat banyak dialek dari berbagai daerah, dialek adalah varietas bicara dalam bahasa tertentu. Kamus Oxford Inggris (OED) menggambarkan dialek sebagai ”salah satu bentuk bawahan atau varietas bahasa yang timbul dari kekhasan lokal kosa kata, pengucapan dan idiom”. Variasi-variasi ini dapat ada di semua tingkat linguistik, yaitu kosakata, idiom, tata bahasa, dan pelafalan[3]. Rekognisi wicara dengan *multi-dialect* tetap merupakan masalah yang sulit, walaupun *dialect-specific acoustic model* bekerja baik secara general[4].

Pada penelitian terdahulu sistem yang dirancang hanya untuk satu dialek yang spesifik saja, oleh karena itu pada penelitian ini penulis akan mengimplementasikan sistem yang dirancang untuk merekognisi beberapa dialek yang ada di Indonesia.

1.2 Perumusan Masalah

Sistem rekognisi dialek yang berjumlah banyak memberikan akurasi rendah.

1.3 Tujuan

Tujuan pada penelitian ini adalah untuk merancang sistem rekognisi dialek di Indonesia.

1.4 Studi Terkait

1.4.1 *Uji Validasi Suara Berbasis Pengenalan Suara (Voice Recognition) Menggunakan EASY VR 3.0*

Pada penelitian[5] Terdapat dua metode untuk membangun sistem berbasis perintah suara (*Voice Command*) yaitu dengan menggunakan pengenalan wicara (*Speech Recognition*) dan pengenalan suara (*Voice Recognition*). *Speech Recognition* akan mengubah sinyal analog suara menjadi data digital yang akan dicocokkan dengan pola tertentu yang disimpan di dalam basis data. Sehingga hasil yang didapatkan berupa teks yang sesuai dengan pola ucapan yang diberikan. Sedangkan *Voice Recognition* akan mengenali suatu suara dengan membandingkan pola karakteristiknya dengan sinyal suara yang menjadi referensi atau acuan yang sudah disimpan sebelumnya. Jadi dengan kata lain *Speech Recognition* dapat mengerti kata apa yang dikatakan oleh seseorang dan *Voice Recognition* dapat mengidentifikasi seseorang melalui suaranya.

1.4.2 *Speech Recognition*

Pada penelitian[6][7] mengatakan bahwa *Speech Recognition* adalah sebuah proses untuk mengubah sinyal ucapan menjadi sebuah text dengan menggunakan algoritma yang diimplementasikan pada suatu program komputer. Teknologi ini memungkinkan suatu perangkat untuk mengenali dan memahami kata-kata yang di ucapkan dengan cara digitalisasi kata dan mencocokan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan dalam suatu perangkat. Kata-kata yang diucapkan diubah bentuknya menjadi sinyal digital dengan cara mengubah gelombang suara menjadi sekumpulan angka yang kemudian disesuaikan dengan kode-kode tertentu untuk mengidentifikasikan kata-kata tersebut. Hasil dari identifikasi kata yang diucapkan dapat ditampilkan dalam bentuk tulisan atau dapat dibaca oleh perangkat teknologi sebagai sebuah komando untuk melakukan suatu pekerjaan. Tujuan dari *Speech Recognition* area adalah membuat teknik dan sistem untuk menerima input berupa ucapan kepada sistem.Secara umum, model akustik khusus dialek bekerja dengan baik[4] . Namun, sulit untuk multi-dialek. Oleh karena itu, penelitian tentang pengenalan dialek menjadi penting. Sistem pengenalan dialek dapat dikembangkan menggunakan pembelajaran mesin konvensional tunggal: seperti mesin vektor dukungan (SVM) [8], [9], jaringan saraf tiruan (JST) [10], Gaussian Mixture Model (GMM) [11], ensemble learning [9], [12], [13], atau deep learning, seperti jaringan saraf

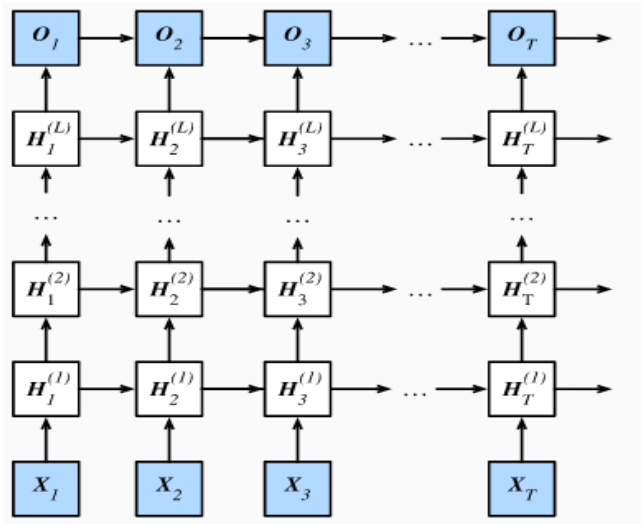
berulang (RNN) [14], [15] memori jangka pendek panjang (LSTM) [16], gabungan jaringan saraf konvolusional dan unit berulang dua arah gated (CNN-BiGRU) [17], LSTM dua arah (BLSTM) [18], kombinasi Hidden Markov Model dan LSTM (HMM-LSTM) [19].

1.4.3 *Mel Frequency Cepstrum Coefficients*

Pada penelitian[20][21] MFCC (*Mel Frequency Cepstrum Coefficients*) merupakan salah satu metode ekstraksi ciri untuk sinyal akustik terbaik. Analisis suara pada *mel-frequency* didasarkan pada persepsi pendengaran manusia, karena telinga manusia telah diamati dapat berfungsi sebagai filter pada frekuensi tertentu. MFCC merupakan satu metode yang banyak dipakai dalam bidang *speech recognition*. Metode ini digunakan untuk melakukan *feature extraction*, sebuah proses yang mengkonversikan sinyal suara menjadi beberapa parameter. Filter ini digunakan untuk menangkap karakteristik fonetis penting dari sebuah ucapan. MFCC digambarkan dalam skala *mel-frekuensi* yang merupakan frekuensi linier dibawah 1000Hz dan logaritmik di atas 1000Hz.

1.4.4 *Deep Recurrent Neural Network*

Pada penelitian[7][22] Pada umumnya, manusia tidak membuat keputusan secara tunggal setiap saat. Manusia akan selalu memperhitungkan masa lalu dalam membuat sebuah keputusan. Cara berpikir seperti ini adalah dasar dari pengembangan Recurrent Neural Network. Sama seperti analogi tersebut, RNN tidak membuang begitu saja informasi dari masa lalu dalam proses pembelajarannya. Hal inilah yang membedakan RNN dari Artificial Neural



Network biasa. Secara singkat, RNN adalah salah satu bagian dari keluarga Neural Network untuk memproses data yang bersambung (sequential data). Cara yang dilakukan RNN untuk dapat menyimpan informasi dari masa lalu adalah dengan melakukan looping di dalam arsitekturnya, yang secara otomatis membuat informasi dari masa lalu tetap tersimpan.Perbedaannya dengan Deep Recurrent Neural Network (DRNN) adalah jumlah hiddent layer pada DRNN lebih banyak. Pada penelitian ini, penguji menggunakan DRNN dengan single direction karena dirasa sudah cukup untuk mempelajari lima dialek di Indonesia.