DEEP NEURAL NETWORK UNTUK PENGENALAN UCAPAN PADA BAHASA SUNDA DIALEK UTARA

DEEP NEURAL NETWORK FOR SPEECH RECOGNITION ON SUNDANESE LANGUAGE OF THE NORTHERN DIALECT

[1]Ghiffari Arwandani, [2]Andrew Briand Osmond, [3]Ratna Astuti Nugrahaeni

[123]Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom
[1]ghiffariarwandani@gmail.com, [2]abosmond@telkomuniversity.ac.id, [3]ratnaan@telkomuniversity.ac.id

Abstrak

Indonesia merupakan Negara dengan banyak ragam suku. Dari berbagai macam suku tadi, Indonesia mempunyai banyak Bahasa daerahnya masing-masing sebagai pembeda atau identitas dari daerah tersebut. Dalam hal ini pengenalan ucapan sangat penting untuk mempermudah pengenalan Bahasa yang digunakan. Pengenalan ucapan memiliki banyak metode sebagai pembelajaran, salah satunya menggunakan Deep Learning.

Deep learning sebuah model jaringan syaraf tiruan yang akhir-akhir ini mulai ramai dikembangkan. Pendekatan yang sering digunakan untuk mengimplementasikan Deep Learning adalah graphical methods atau Multilayer Representation, atau Multilayer Graphical model seperti Belief Network, Neural Network, Hidden Markov, dan lain-lain. Deep Learning telah menunjukkan hasil yang baik dalam meningkatkan akurasi pengenalan suara atau kasus-kasus lainnya yang serupa. Oleh karena itu pada penelitian ini penulis akan mencoba untuk mengimplementasikan Deep Neural Network pada Speech Recognition untuk mengklasifikasian Bahasa Sunda dialek Utara.

Dari hasil penelitian yang dilakukan, dari nilai parameter tertentu didapatkan akurasi sebesar 100%. Setelah mendapatkan parameter ideal dilakukan klasifikasi dengan rasio dari data latih : data data uji sebesar 50% : 50%, 60% : 40%, 70% : 30%, 80% : 20% dan 90 : 10%. Dari pengujian dengan rasio tesebut didapatkan kesimpulan bahwa, semakin banyak data latih semakin baik akurasi yang didapatkan.

Kata kunci : Deep learning, Speech Recognition, Deep Neural Network

Abstract

Indonesia is a country with many tribes. From various tribes earlier, Indonesia has many languages of their respective regions as a differentiator or identity of the region. In this case speech recognition is very important to facilitate the introduction of the language used. Speech recognition has many methods as learning, one of them using Deep Learning.

Deep learning of a model of artificial neural network which recently began to be developed. A common approach used to implement Deep Learning is graphical methods or Multilayer Representation, or Multilayer Graphical models such as Belief Network, Neural Network, Hidden Markov, and others. Deep Learning has shown good results in improving the accuracy of speech recognition or other similar cases. Therefore in this study the authors will try to implement Deep Neural Network on Speech Recognition to classify the Sundanese language of the Northern dialect.

From the results of research conducted, obtained accuracy by changing each parameter of 100%. After obtaining the ideal parameters are classified with the ratio of the training data: the test data data is 50%: 50%, 60%: 40%, 70%: 30%, 80%: 20% and 90: 10%. From the test with the ratio, it is concluded that, the more train data the better the accuracy obtained.

Keywords : Deep learning, Speech Recognition, Deep Neural Network

1. Pendahuluan

1.1 Latar Belakang

Indonesia merupakan neraga kepulauan yang menurut Kementerian Pertahanan RI menyebutkan jumlah Pulau yang dimiliki oleh NKRI tercatat 17.504 Pulau [1]. Dengan banyaknya pulau di Indonesia sudah pasti Indonesia juga memiliki banyak ragam suku. Salah satu yang identik dengan suku yaitu Bahasa.

Bahasa berkaitan erat dengan hubungan sosial kita dan merupakan media dimana kita berpartisipasi dalam berbagai aktivitas sosial [2]. Salah satu Bahasa yang di miliki Indonesia yaitu Bahasa Sunda. Hampir seluruh penduduk pulau Jawa khususnya Jawa Barat bisa/menguasai Bahasa Sunda. Bahasa Sunda juga mempunyai dialek yang beragam bergantung pada daerahnya. Itu juga yang menjadi pembeda atau menjadi ciri khas suatu daerah untuk pembeda Bahasa Sunda yang mereka gunakan.

Seiring berkembangnya teknologi, sekarang kita dapat menemukan aplikasi Speech Recognition (pengenalan suara). Speech Recognition adalah kemampuan program untuk mengidentifikasi kata dan frase dalam bahasa lisan dan mengkonversikannya ke format yang dapat dibaca oleh mesin [3]. Riset Speech Recognition untuk pengenalan ucapan dialek bahasa daerah masih terbilang sedikit bahkan hampir tidak ada.

Oleh karna itu di sini penulis ingin membuat suatu sistem Speech Recognition dengan menggunakan metode Deep Neural Network yang bertujuan untuk mempermudah pengenalan ucapan Bahasa Sunda dialek Utara dan diharapkan kedepannya dapat di kembangkan untuk penerjemah Bahasa antar daerah.

1.2 Tujuan

Adapun tujuan yang akan dicapai pada Tugas Akhir ini :
Membuat suatu sistem Speech Recognition untuk pengenalan Bahasa Sunda dialek Utara dengan menggunakan metode Deep Neural Network yang akan menghasilkan output berupa pengklasifikasian dari input yang nantinya akan menentukan apakah input memenuhi bobot standar untuk Bahasa Sunda dialek Bogor.

1.3 Identifikasi Masalah

Beberapa identifikasi masalah yang akan dibahas dalam Tugas Akhir ini adalah sebagai berikut :
Bagaimana membuat membuat sistem Speech Recognition untuk pengenalan Bahasa Sunda dialek Utara dengan menerapkan metode Deep Neural Network dan memproses suatu masukan berupa suara ke suatu sistem agar dimengerti oleh sistem tersebut dan dapat mengklasifikasikan masukan kedalam setiap class dengan dialek yang berbeda beda.

2. Dasar Teori

Bagian ini berisi tentang dasar teori yang digunakan untuk merancang speech recognition dengan mengimplementasikan metode deep neural network untuk klasifikasi Bahasa sunda dialek utara (bogor). Adapun teori-teori yang digunakan adalah sebagai berikut.

2.1 Speech Reognition

Speech Recognition atau yang kita kenal sebagai *Automatic Speech Recognition* (ASR) adalah kemampuan mesin atau program untuk mengidentifikasi kata dan frase dalam bahasa lisan dan mengkonversikannya ke format yang dapat dibaca oleh mesin [3]. Teknologi ini memungkinkan suatu perangkat dapat mengenali kata-kata dengan cara menganalisis spesifikasi kata yang disebutkan lalu men digitalisasi kata dan mencocokan sinyal digital tersebut dengan pola tertentu yang tersimpan untuk menyempurnakan pengenalan suara agar menghasilkan akurasi yang tinggi. Perancangan sistem ASR melalui dua fase yaitu fase pelatihan dan fase pengujian. Pada fase pelatihan, sistem akan menerima masukan berupa sample yang akan dijadikan sebagai sebagai data latih. Data latih akan disimpan didalam database dan akan dijadikan acuan dalam fase pengujian. Fase berikutnya adalah fase pengujian. Pada fase ini sistem akan di uji dengan cara memasukan sample dan dibandingkan dengan data latih yang ada, kemudian diputuskan keluaran berdasarkan kemiripan dari data latih. Adapun dua modul utama yang dibutuhkan dalam perancangan *speech recognition*, yaitu:

2.1.1 Ektrasi Ciri (feature extraction)

Ekstraksi ciri merupakan proses mengkonversi sinyal suara menjadi beberapa parameter, informasi yang didapatkan lebih rendah karena akan menghilangkan beberapa informasi yang kurang penting tanpa mengubah arti sesungguhnya.

2.1.2 Pencocokan Ciri (pattern matching)

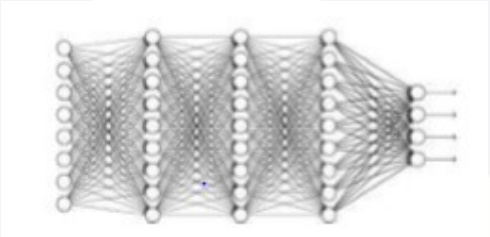
Dalam pencocokan ciri akan di lakukan pembandingan atau mencocokan data dari sinyal masukan dengan data latih yang telah ada di dalam database. Hasil dari pencocokan ciri ini akan menjadi keluaran sistem.

2.2 MFCC

MFCC merupakan salah satu metode ekstraksi ciri untuk sinyal akustik terbaik [4]. Analisis suara pada mel-frequency didasarkan pada persepsi pendengaran manusia, karena telinga manusia telah diamati dapat berfungsi sebagai filter pada frekuensi tertentu. *Mel Frequency Cepstrum Coefficients* (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.

2.3 Deep Neural Network

Algoritma DNN (*Deep Neural Networks*) adalah salah satu algoritma berbasis jaringan saraf yang dapat digunakan untuk pengambilan keputusan. *Deep Neural Network* memiliki tujuan meniru cara kerja otak manusia dengan metode *Multi Layer*. DNN ini terdiri dari beberapa Hidden Layer dengan koneksi antar Layer tetapi tidak ada koneksi antar units pada setiap layer-nya [6]. Pendekatan ini memungkinkan data yang komplek menjadi lebih mudah di modelkan [5].



Gambar 2. 1 DNN Layer

Metode ini memiliki arsitektur yang serupa dengan arsitektur pada Artificial Neural Networks (ANNs), dengan *Supervised Training*. Dengan mengidentifikasi masukan dan mencocokan nya dengan pola yang susah ada. Adapun kelebihan *Deep Lerning methods* untuk *Speech Recognition*, yaitu arsitektur jaringan lebih baik, Dapat mengoptimalkan segudang parameter, DNN sangat bagus untuk *Speech Recognition*, DNN lebih cepat dalam memahami banyak Bahasa/Dialek [7].

2.4 Auto Encoder

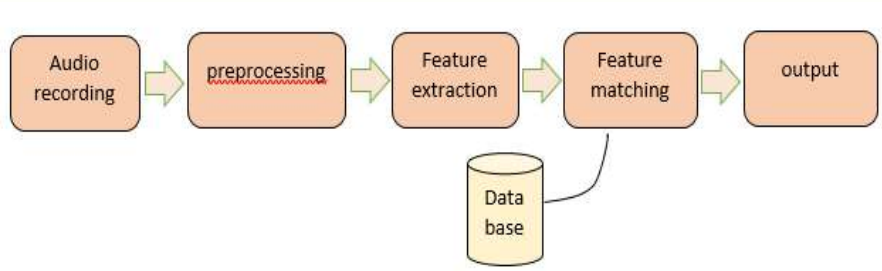
Autoencoder merupakan sebuah *neural network* yang digunakan untuk melatih data agar data yang dihasilkan pada *output* sama seperti data yang digunakan sebagai *input*nya. Tujuan dari *autoencoder* adalah untuk mempelajari representasi (*encoding*) untuk satu set data, biasanya untuk tujuan pengurangan dimensi. Ketika jumlah *neuron* di lapisan tersembunyi kurang dari ukuran input, *autoencoder* akan belajar representasi terkompresi dari input. *Autoencoder* dapat dibangun dengan membuat jaringan *feedforward*, dan kemudian memodifikasi beberapa pengaturan. Atur ukuran layer tersembunyi untuk *autoencoder*.

Baru-baru ini, konsep *autoencoder* telah menjadi lebih banyak digunakan untuk belajar model data generative [8].

3. Perancangan

3.1 Perancangan Sistem Secara Umum

Pada perancangan sistem ini menjelaskan secara umum mengenai tahapan sistem yang akan diteliti lebih lanjut. Berikut merupakan perancangan sistem secara umum yang ditujukan pada gambar 3.1. Sistem yang dibuat merupakan sistem yang dapat mengkonversikan masukan berupa sinyal suara menjadi teks Bahasa latin dari masukan tadi. Berikut adalah skema umum perancangan sistem speech recognition.



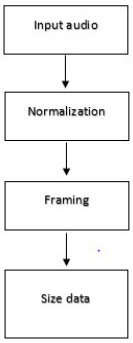
Gambar 3. 1 Perancangan Sistem Secara Umum

Pada gambar 3.1 menjelaskan bahwa secara umum, sistem *Speech Recognition* memproses sinyal suara yang masuk dan menyimpannya dalam bentuk digital. Hasil proses digitalisasi tersebut kemudian dikonversi dalam bentuk spektrum suara (cepstrum) yang akan dianalisis dengan cara membandingkannya dengan pola suara pada database sistem. Adapun tahap tahap ASR, yaitu:

1. Tahap Menerima Masukan: Pada tahap ini system mendapat masukan berupa gelombang suara. Suara berasal dari tangkapan mikrofon ataupun hasil rekaman.
2. Tahap Ekstrasi: Pada Tahap ini dilakukan penyimpanaan masukan yang berupa suara sekaligus pembuatan basis data sebagai pola. Data suara masukan diproses satu per satu berdasarkan urutanya.
3. Tahap Perbandingan/Pencocokan: Pada tahap ini sistem akan membandingkan atau mencocokan hasil ekstraksi dari masukan dengan data latih yang ada di dalam database yang tersedia.
4. Tahap Validasi: Pada tahap ini sistem akan mengambil keputusan terhadap input dan memasukan input kedalam class yang sesuai dengan bobot dari input.

3.2 Pre-Processing

Proses *pre-processing* ini merupakan tahap awal pemrosesan sinyal suara yang terdiri dari tahapan normalisasi, framing, serta terakhir proses menentukan besar size untuk prosses selanjutnya. Adapun diagram blok untuk proses *pre-processing* ditunjukkan pada gambar dibawah:



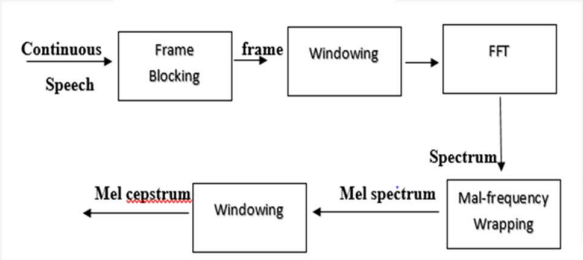
Gambar 3. 2 Block Diagram Pre-Processing

Berikut merupakan penjelasan alur *block diagram* dari prosses *pre-processing*:

1. Pada proses normalisasi akan :
   Prosses normalisasi bertujuan agar data tidak berpengaruh pada besar kecilnya amplitudo signal hasil perekaman, proses normalisasi juga tidak mengubah informasi yang terdapat pada signal. Proses normalisasi dilakukan dengan mencari nilai mutlak terendah/tertinggi dari signal dan digunakan untuk membagi signal aslinya.
2. Pada proses *framing* sistem akan :
   Sinyal masukan dipotong menjadi ukuran yang lebih kecil (dibuat menjadi frame-frame). Pemotongan sinyal (*framing*) dilakukan setiap 1 detik. Jadi hasil *framing* akan menghasilkan 16 frame untuk 1 narasumber.
3. Pada proses *size data* sistem akan :
   Dari hasil perekaman data, diketahui bahwa kalimat terpanjang menggunakan 4 kata (dapat dilihat di gambar table pengambilan data) dan kalimat yang lain hanya memiliki 3 kata,2 kata dan 1 kata. Untuk size data diambil dari kalimat yang mempunyai kata terbanyak dan untuk setiap kalimat yang memiliki kata kurang dari 4 akan disi dengan 0.

3.3 Feature Extraction

Ekstrasi Ciri atau *feature extraction* dilakukan pada dua proses, yaitu ektraksi ciri untuk pembuatan database sebagai template dan ektraksi ciri masukan data uji. Metode yang digunakan pada tahap ekstraksi ciri yaitu MFCC. Pada tahap ektrasi ciri ini MFCC akan memberikan 3 ciri untuk setiap frame, setiap ciri memiliki 12 nilai. Jadi jumlah untuk setiap *frame* selesai melewati ekstraksi ciri akan bernilai 3 x 12=36. Dari hasil *framing* diatas nilai untuk setiap *frame* yaitu 4 diambil dari jumlah kalimat yang memiliki jumlah kata terbanyak dan setelah di ektraksi ciri setiap *frame* memiliki 36 ciri. Jadi jumlah ciri untuk setiap *frame*nya yaitu 4 x 36=144.



Gambar 3. 3 Diagram Block MFCC

3.4 Feature Matching

Pada *feature matching* akan dilakukan 2 prosses yaitu fase pembelajaran dan fase pengujian. Setelah melakukan *feature extraction* data akan masuk kedalam fase pembelajaran. Pada fase pembelajaran ini, sistem akan melakukan pembelajaran agar mencapai akurasi tertinggi. Terdapat 3 parameter utama yang digunakan, yaitu bias, hiddensize dan epoch. Untuk nilai parameter dari *hiddensize* dan epoch akan diubah-ubah untuk mendapatkan akurasi tertinggi sedangkan untuk nilai parameter dari bias sudah tetap 1 (default). Selain dari parameter tersebut ada 2 parameter lagi dalam autoencoder untuk menentukan nilai error, yaitu *L2WeightRegulazitation* dan *sparsityRegulazitation*. Pada fase pengujian tidak terlalu berbeda dengan fase pembelajaran. Hanya saja pada fase pengujian input yang digunakan berbeda dengan data yang terdapat pada database. Untuk mendapatkan akurasi tinggi, hanya perlu merubah-ubah nilai dari parameter yang digunakan. Pada fase ini parameter yang digunakan sama dengan parameter pada saat fase pembelajaran.

4. Pengujian

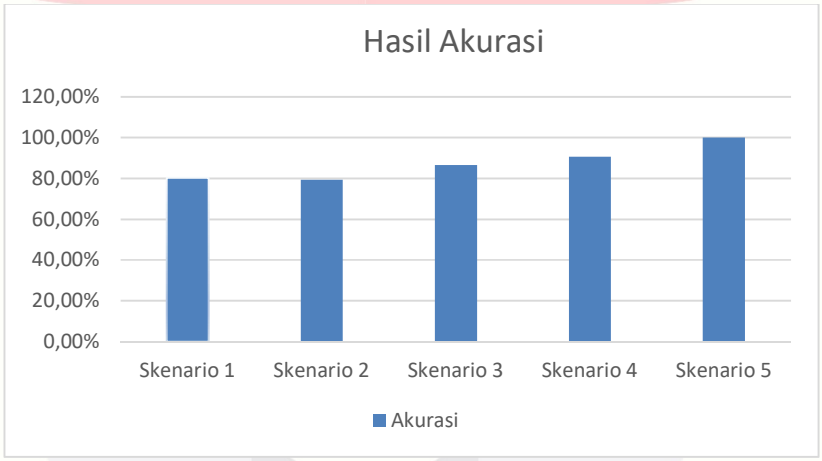
Terdapat beberapa skenario dari pengujian terhadap sistem yang dibuat, untuk mengetahui hal apa saja yang mempengaruhi tingkat akurasi sistem yang membuat sistem bekerja maksimal, adapun skenario pengujian sebagai berikut:

4.1 Pengujian Validasi Sistem

Validasi peforma sistem dengan mengubah setiap parameter agar mendapatkan akurasi terbaik yang nantinya parameter dengan akurasi terbaik akan digunakan untuk pengujian selanjutnya. Pada pengujian ini jumlah data uji sama dengan database. Pada table 4.1 dibawah ini merupakan skenarion pengujian dengan parameter uji.

Parameter	Nilai				
	Skenario 1	Skenario 2	Skenario 3	Skenario 4	Skenario 5
Hiddensize 1	50	100	100	144	200
Hiddensize 2	100	200	50	144	100
Epoch	100	100	100	100	400
L2WeightRegulazitation	0,02	0,0002	0,0002	0,0002	0,0002
sparsityRegulazitation	3	7	3	3	3

Table 4. 1 Pamater Validasi Sistem



Gambar 4. 1 Grafik Hasil Validasi Sistem

Dapat dilihat dari gambar 4.1 diatas, setiap parameter mempengaruhi tingkat akurasi. Dapat disimpulkan bahwa semakin kecil nilai (mendakati 0) untuk parameter *L2WeightRegulazitation* dan *sparsityRegulazitatio* akan meningkatkan akurasi, karena nilai dari kedua parameter tersebut menetukan nilai error. Sedangkan untuk nilai *hiddenzide* akan menghasilkan tingkat akurasi yang baik saat dibuat nilainya lebih kecil dari *inputsize*. Pada kasus ini *inputsize* yang diberikan adalah sebesar 144. Semakin banyak nilai dari parameter epoch akan semakin tinggi juga akurasi yang dihasilkan, karena epoch merupakan jumlah dari pembelajaran yang dilakukan. semakin banyak pembelajaran akan semakin pintar sistem dalam menentukan hasil klasifikasi.

4.2 Pengujian Peforma Sistem.

Pada pengujian selanjutnya akan dilakukan uji coba untuk peforma sistem dengan parameter diambil dari pengujian sebelumnya dengan tingkat akurasi tertinggi. Pada pengujian Validasi sistem diketahuin bahwa parameter terbaik yaitu

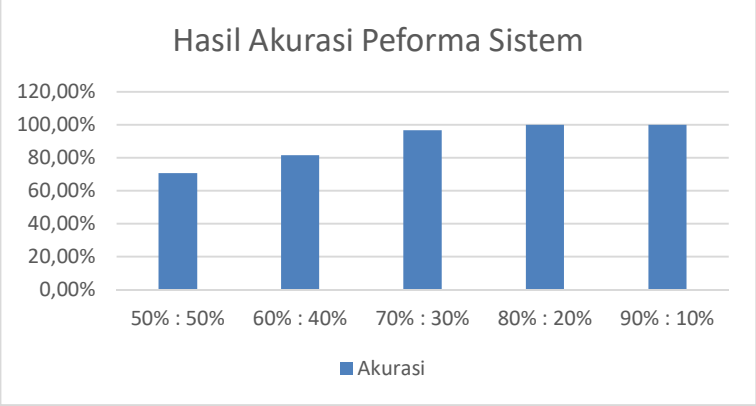
Parameter	Nilai
Hiddensize 1	100
Hiddensize 2	50
Epoch	400
L2WeightRegulazitation H	0,0002
sparsityRegulazitation H	3

Table 4. 2 Parameter Peforma Sistem

Dengan tingkat akurasi 100%. Pada pengujian ini data training dan data test akan bervariasi dengan rincian sebagai berikut:

Pengujian	Data *Training* (%)	Data *Testing* (%)
Skenario 1	50	50
Skenario 2	60	40
Skenario 3	70	30
Skenario 4	80	20
Skenario 5	90	10

Table 4. 3 Rasio Pengujian Peforma Sistem



Gambar 4. 2 Grafil Hasil Peforma Sistem

Dari gambar 4.2 diatas dapat dilihat hasil pengujian tingkat akurasinya semakin meningkat seiring meingkatnya jumlah data training. Pada pengujian 80% : 20% dan 90% : 10% mempunyai tingkat akurasi sempurna dengan 100% akurasi, itu disebabkan karena jumlah data training lebih banyak dibandingkan dengan data testing dan hasil dari klasifikasi data pada data testing ini memiliki lebih banyak kesamaan atau lebih dekat hasil klasifikasinya dengan data training.

5. Kesimpulan

a. Pada proses pengujian validasi sistem didapatkan bahwa tingkat akurasi tertinggi ditentukan dari nilai yang diberikan untuk setiap parameter.

b. Untuk mendapatkan akurasi yang tinggi nilai dari parameter *L2WeightRegulazitation* dan *sparsityRegulazitation* harus kecil (mendekati 0), untuk parameter dari *HiddenSize* lebih baik nilai yang diberikan lebih kecil dari *Input Size* dan nilai untuk epoch besar agar pembelajaran yang dilakukan semakin banyak.

c. Pada proses pengujian peforma sistem berdasarkan besar data didapatkan hasil akurasi yang maksimal yaitu 100% saat data latih 80% dan 90% lebih banyak dibandingkan dari data uji

Daftar Pustaka:

[1] Brigjen TNI Dody Usodo Hargo, S.IP,MM. ___. Jumlah Pulau DI Indonesia, [online] (https://dkn.go.id/ruang-opini/9/jumlah-pulau-di-indonesia.html diakses tanggal 3 Oktober 2017)

[2] Agha. "Language and Social Relations". Cambridge University Press. 2006.

[3] Margaret Rouse, "Speech Recognition" [online] (http://searchcrm.techtarget.com/definition/speech-recognition) diakses tanggal 3 Oktober 2017).

[4] R. S. Chavan dan G. S. Sable, "An Overview of Speech Recognition Using HMM," International Journal of Computer Science and Mobile Computing, vol. 2, no. 6, 2013.

[5] Chunyang Wu , Penny Karanasou , Mark J.F. Gales , Khe Chai Sim, "Stimulated Deep Neural Network for Speech Recognition", University of Cambridge, National University of Singapore, September, 2016.

[6] Geoffrey Hinton "Deep Neural Networks for Acoustic Modeling in Speech Recognition" 2012.

[7] Li Deng , Geoffrey Hinton, and Brian Kingsbury, "New Type of Deep Neural Network Learning for Speech Reocgnation and Related Applications: An Overview", IEE Internasional Conference, 2013"

[8] Diederik P Kingma, Welling, Max, "Auto-Encoding Variational Bayes", 2013

[9] Frank Seide, Gang Li, Xie Chen, and Dong Yu, "Feature Engineering in Context Dependent Deep Neural Networks for Conversational Speech Transcription", IEEE, 2011.

[10] Ronan Collobert, Jason Weston, "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning", published ICL, 2011.