Penentuan Dialek Jawa Menggunakan Metode Deep Neural Network

Determination of the Javanese Dialect Using the Deep Neural Network Method

[1]Mustofa Restu Adi, [2]Andrew Briand Osmond, [3]Anggunmeka Luhur Prasasti

[123]Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom
[1]restumustofa0603@gmail.com, [2]abosmond@telkomuniversity.ac.id, [3]anggunmeka@gmail.com

Abstrak

Indonesia merupakan negara dengan banyak ragam suku bahasa dan budaya. Dari berbagai macam suku tersebut Indonesia mempunyai banyak bahasa daerahnya masing-masing sebagai ciri khas dan pembeda bahasa dari masing-masing daerah. Dalam hal ini untuk mempermudah tiap orang memahami intisari dari masing-masing bahasa dari berbagai macam suku dan daerah di Indonesia maka pengenalan ucapan sangatlah penting.Pengenalan ucapan memiliki banyak metode sebagai pembelajaran, salah satunya menggunakan *Deep Learning*.

***Deep learning* merupakan sebuah model jaringan syaraf tiruan yang akhir-akhir ini mulai ramai dikembangkan. *Deep Learning* telah menunjukkan hasil yang baik dalam meningkatkan akurasi pengenalan suara atau kasus-kasus lainnya yang serupa. *Deep Learning* sendiri mempunyai berbagai macam pendekatan, tetapi pada penelitian ini penulis hanya menigmplementasikan salah satu pendekatan dari *Deep Learning* yaitu *Deep Neural Network* pada *Speech Recognition*.**

Algoritma DNN (*Deep Neural Networks*) adalah salah satu algoritma berbasis jaringan saraf yang dapat digunakan untuk pengambilan keputusan. Aplikasi yang dibuat ini hampir sama seperti aplikasi *google voice* dan *translate* pada umumnya, namun dalam fitur yang dibuat dalam *google voice* dan *translate* pengenalan ucapannyahanya bisa dalam bahasa umum tiap-tiap negara tidak bisa dalam bahasa daerah masing-masing tiap suku. Maka, dalam aplikasi yang akan dibuat nanti aplikasi bisa menentukan pengenalan bahasa tiap suku dan daerah tapi dalam aplikasi nanti hanya fokus dalam pengerjaan dialek berbahasa Jawa. Dalam aplikasi ini *input* yang berupa suara bahasa berdialek Jawaakan diolah menjadi *output* berupa teks berdialek Jawa. Tidak hanya mengenali satu dialek Jawa saja akan tetapi juga dapat menentukan dialek Jawa dari daerah mana yang di hasilkan.

Kata Kunci :*Deep Learning, Speech Recognition,Deep Neural Network,* dialek

Abstract

Indonesia is a country with many ethnic and cultural ethnicities. Of the various ethnic groups, Indonesia has many regional languages as distinctive and distinctive languages of each region. In this case, to make it easier for everyone to understand the essence of each language from various kinds of tribes and regions in Indonesia, speech recognition is very important. Speech recognition has many methods as learning, one of which uses Deep Learning.

Deep learning is a model of artificial neural network that has recently been developed. Deep Learning has shown good results in improving speech recognition accuracy or other similar cases. Deep Learning itself has a variety of approaches, but in this study the authors only implement one of the approaches of Deep Learning, namely the Deep Neural Network in Speech Recognition.

The DNN (Deep Neural Networks) algorithm is one of the neural network-based algorithms that can be used for decision making. This application is almost the same as the Google Voice and Translate application in general, but in the features made in Google Voice and the speech recognition translation it can only be in the general language, each country cannot be in the local language of each tribe. So, in the application that will be made later the application can determine the recognition of the language of each tribe and region but in the application will only focus on the work of the Javanese language dialect. In this application the input in the form of the language of the Javanese dialect will be processed into an output in the form of Javaanese dialect text. Not only recognize one Javaanese dialect but also can determine the Javaanese dialect from which area is produced.

Keywords : Deep Learning, Speech Recognition,Deep Neural Network, dialect.

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 bangsa.Banyaknya pulau tersebut juga terdapat berbagai macam-macam suku, bahasa sehari-hari, adat isiadat serta ciri khas yang berbeda dari tiap suku dan derah. Tetapi salah satu identik yang membedakan antara tiap daerah adalah Bahasa atau dialek yang berbeda-beda.

Jumlah bahasa daerah yang dimiliki Indonesia sangatlah banyak. Menurut Lembaga Ilmu Pengetahuan Indonesia (LIPI) terdapat lebih dari 700 bahasa daerah akan tetapi baru 617 bahasa yang teridentifikasi dan 139 bahasa terancam punah [2]. Salah satu bahasa yang terancam punah dan menjadi ciri khas bangsa adalah Bahasa Jawa, karena bahasa Jawa sendiri merupakan rata-rata bahasa yang dimiliki oleh provinsi Jawa Timur dan Jawa Tengah. Bahasa ini terancam punah karena era globalisasi jika diperkotaan masyarakat lebih sering menggunakan logat bahasa resmi yakni Bahasa Indonesia untuk berkomunikasi dengan orang lain daripada menggunakan logat bahasa daerahnya sendiri. Mungkin saja, bahasa daerah lebih sukar untuk dipahami bagi orang awam untuk berkomunikasi dengan orang asli dari daerah Jawa Timur dan Jawa Tengah.

Bahasa Jawa juga mempunyai bermacam-macam dialek antar daerah yang berbeda-beda serta pemahaman dan artinya yang berbeda. Banyak orang awam yang masih mengalami kesulitan dalam memahami atau mengenali bahasa antar suku atau daerah yang ada khususnya dareah Jawa Timur dan Jawa Tengah. Sehingga masih sering terjadi kesalahpahaman dalam memahami maksud dan tujuan arti dari bahasa daerah tersebut.

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 karena itu pada penelitian ini penulis ingin membuat sebuah aplikasi atau sebuah sistem yang akan menentukan bahasa atau dialek Jawa sebagai masukan dan keluaran yang berupa teksdan dialek dari bahasa Jawa yang bermacam-macam logat tiap daerahnya.

Dengan menggunakan sebuah sistem *Speech Recognition* dan menggunakan metode *Deep Neural Network*yang bertujuan untuk memudahkan seseorang yang tidak bisa berbahasa Jawa atau bisa berbahasa Jawa tetapi kurang mengerti makna bahasa dari tiap daerah tersebut. Sehingga ketika saat berkomunikasi bisa tau apa yang dimaksud dari logat dialek Jawa tersebut serta tahu dialek Jawa daerah mana yang sedang diucapkan dan diharapkan kedepannya dapat dikembangkan untuk penerjemah Bahasa serta tahu dan dapat menentukan bahasa dari seluruh suku bangsa yang ada di Indonesia.

1.2 Tujuan

Tujuan dari pembuatan penelitian tugas akhir ini adalah membuat suatu system *Speech Recognition* untuk pengenalan Bahasa Jawa dengan menggunakan metode *Deep Neural Network* untuk menampilkan keakuratan masukan dan keluaran bahasa serta pengimplementasian metode *Deep Neural*

Network untuk menentukan keluaran bahasa Jawa mana yang diucapkan narasumber seperti yang sudah ditetapkan dalam sistem.

1.3 Identifikasi Masalah

Identifikasi masalah dalam tugas akhir ini yaitu bagaimana membuat system *Speech Recognition* dengan logat dialek Jawa yang berbeda beda dengan menerapkan metode *Deep Neural Network* untuk memproses suatu masukan berupa suara ke suatu sistem agar dimengerti oleh sistem tersebut dan cara menampilkan hasil keluaran dalam bentuk teks angka persentase serta menentukan dialek dari Jawa mana yang sedang diucapkan.

2. Dasar Teori

2.1 Speech Recognition

Speech Recognition atau yang kita kenal sebagai *Automatic Speech Recognition* (ASR)adalahkemampuan mesin atau program untuk mengidentifikasi kata dan frase dalam bahasa lisan dan mengkonversikannya ke format yang dapat dibaca oleh mesin.[3] Masukan sistem adalah ucapan manusia, selanjutnya sistem akan mengidentifikasikan kata atau kalimat yang diucapkan dan menghasilkan teks yang sesuai dengan apa yang diucapkan. Sinyal ucapan pertama kali akan dilewatkan pada bagian penganalisis ucapan untuk mendapatkan besaran-besaran atau ciri-ciri yang mudah diolah pada tahap berikutnya. Untuk setiap ucapan yang berbeda akan dihasilkan pola ciri yang berbeda.Teknologi ini memungkinkan suatu perangkat dapat mengenali kata-kata dengan cara menganalisis spesifikasi kata yang disebutkan lalu mendigitalisasi 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 [8,9].

Adapun dua modul utama yang dibutuhkan dalam perancangan *speech recognition*, yaitu:

1. Ektrasi Ciri (*feature extraction)*

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

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 Ekstrasi Ciri

Tujuan dari ekstrasi ciri adalah untuk mengkonversi gelombang suara, menggunakan alat Pemrosesan Sinyal Digital (PSD), menjadi set tertentu (pada tingkat informasi yang jauh lebih rendah) untuk analisis lebih lanjut. Karena sinyal ucapan memiliki karakteristik yang sama dalam interval waktu yang singkat, maka *short-time spectral analysis* merupakan cara paling umum untuk mengkarakterisasi sinyal suara. Saat ini metode yang paling sering digunakan untuk mengkarakterisasi sinyal suara dalam *speech recognition*yaitu*Mel Frequency Cepstral Coefficients* (MFCC). MFCC merupakan metode yang akan digunakan dalam penelitian tugas akhir ini.

2.3 *Mel-Frequency Cepstral Coefficient (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* [5]. Metode ini digunakanuntuk melakukan *feature extraction*, sebuah proses yang mengkonversikan sinyalsuara 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.

2.4 *Neural Network*

Neural Network atau yang biasa disebut dengan *Artificial Neural Network* (ANN) merupakan representasai dari otak manusia . Dimana sebuah model yang terdiri dari elemen-elemen pengolah serta beberapa *neuron* atau *node* yang berfungsi berdasarkan cara kerja *neuron* manusia [6].

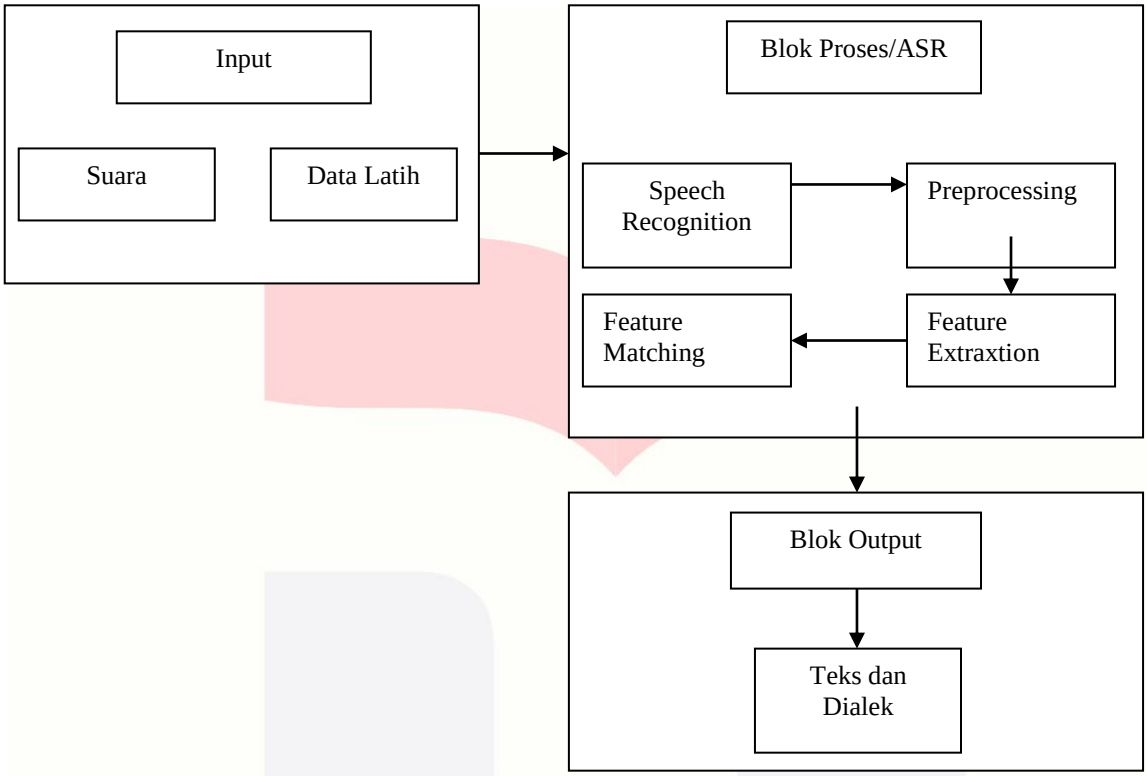
2.5 Deep Neural Network (DNN)

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. Pendekatan ini memungkinkan data yang komplek menjadi lebih mudah dimodelkan [7]. Metode ini memiliki arsitektur yang serupa dengan arsitektur pada *Artificial Neural Networks* (ANNs), dengan *Supervised Training*. *Supervised Training* adalah sebuah pendekatan dimana sudah terdapat data yang dilatih, dan terdapat variable yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengkelompokan suatu data ke data yang sudah ada. *Unsupervised training* tidak memiliki data latih, sehingga dari data yang ada, kita mengelompokan data tersebut menjadi 2 bagian atau 3 bagian dan seterusnya.dapat mengidentifikasi masukan serta mencocokannya dengan pola yang sudah ada [12]. Adapun kelebihan *Deep Lerning methods* untuk *Speech Recognition*, yaitu arsitektur jaringan lebih baik, Dapat mengoptimalkan banyak parameter, DNN cukup baik untuk *Speech Recognition*, DNN lebih cepat dalam memahami banyak Bahasa/Dialek [10].

3. Perancangan

3.1 Gambaran Umum Sistem

Pada perancangan sistem yang akan dibuat merupakan sistem yang akan mengkonversi masukan atau *input*berupa sinyal suara menjadi teks bahasa latin dari masukan tadi, sistem dirancang dengan klasifikasi dialek Bahasa Jawa dengan menggunakan metode *Deep Neural Network* (DNN). Berikut adalah skema umum perancangan sistem *speech recognition.*



Gambar 3.1 Gambaran Umum Sistem

Secara umum, sistem automatic *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 penerima masukan : Pada tahap ini sistem akan mendapat masukan berupa sampel audio suara. Suara berasal dari sampel rekaman yang diucapkan atau tangkapan mikrofon.
2. Tahap *preprocessing* : Pada tahap ini dilakukan persiapan dan mengolah data awal sehingga data yang sudah digunakan merupakan data yang sudah siap dan matang sehingga dapat untuk mempermudah proses-proses dalam tahapan ASR berikutnya.
3. Tahap ekstrasi ciri : Pada Tahap ini dilakukan penyimpanaan masukan yang berupa suara sekaligus pembuatan basis data sebagai pola berpedoman pada MFCC sehingga data proses masukan diproses satu per satu berdasarkan urutannya.
4. Tahap perbandingan : Pada tahap ini sistem akan membandingkan atau mencocokkan data baru dengan dengan data latih yang sudah tersedia dalam database.
5. Tahap validasi : Pada tahap ini sistem akan mengambil keputusan terhadap masukan atau*input* apakah masukan atau *input* dapat dikenali oleh sistem atau tidak

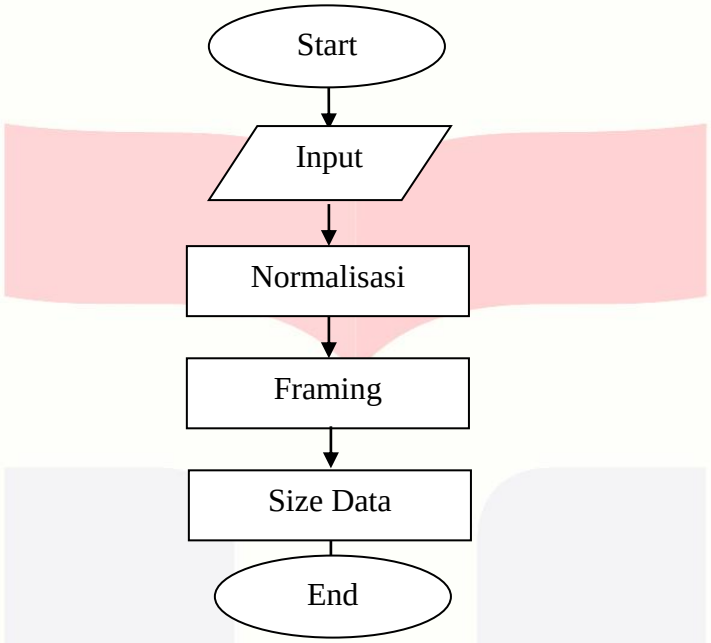
3.2 *Input* Data

Data latih (data keseluruhan) dan data uji (data yang akan diklasifikasi) dimasukan ke dalam program untuk diproses. Data yang dimasukan berupa file suara dari narasumber yang direkam sebelumnya dan dimasukkan ke program dengan format *.wav.

3.3 Preprocessing

Preprocessing adalah mempersiapkan dan mengolah data awal sehingga data yang digunakan adalah data yang sudah siap pakai.

Pada *preprocessing* terdapat 3 prosses, yaitu:



Gambar 3.2 Menunjukkan tahap-tahap *preprocessing*.

Tahap-tahap *preprocessing* adalah sebagai berikut :

1. Normalisasi

   Prosses normalisasi bertujuan agar data tidak berpengaruh pada besar kecilnya amplitudo sinyal 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. *Framing*

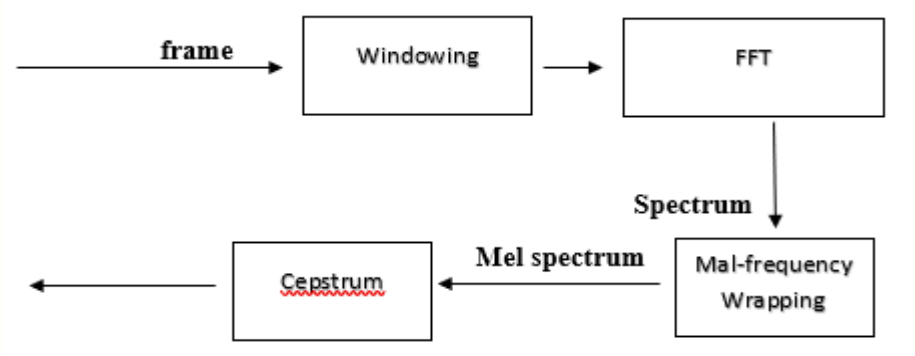
   Sinyal masukan dipotong menjadi ukuran yang lebih kecil (dibuat menjadi *frame-frame*). Untuk pemotongan sinyal (*framing)* dilakukan setiap 1 detik. Didalam pemotongan sinyal *framing* ini terdapat 4 (empat) kombinasi dari campuran logat Jawa.

3. *Size* Data

   Dari hasil perekaman data, diketahui bahwa kalimat terpanjang menggunakan 7 kata (dapat dilihat di gambar table pengambilan data) dan kalimat yang lain hanya memiliki 6 kata, 5 kata, 4 kata, 3 kata dan 2 kata Untuk size data diambil dari kalimat yang mempunyai kata terbanyak dan untuk setiap kalimat yang memiliki kata kurang dari 7 akan disi dengan 0.

3.4 Extraction(MFCC)

Kegunaan ekstraksi ciri adalah untuk mengkonversi gelombang suara, menggunakan alat Pemrosesan Sinyal Digital (PSD), menjadi set tertentu (pada tingkat informasi yang jauh lebih rendah) untuk analisis lebih lanjut. Ekstraksi ciri atau *feature extraction* dilakukan pada dua proses, yaitu ektraksi ciri untuk pembuatan *database* sebagai *template* dan ektraksi ciri masukan data uji. 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 Proses *Extraction*

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.

Tabel 4.1 Tabel parameter awal

Parameter	Nilai
Hiddensize 1	200
Hiddensize 2	200
L2WeightRegulazitation H1	0,004
sparsityRegulazitation H1	5
L2WeightRegulazitation H2	0,004
sparsityRegulazitation H2	5
Epoch	100

4.2 Pengujian Performa Sistem

Pengujian performa sistem dilakukan dengan acuan parameter terbaik dari perhitungan hasil validasi, sehingga untuk parameter terbaik didapatkan sebagai berikut :

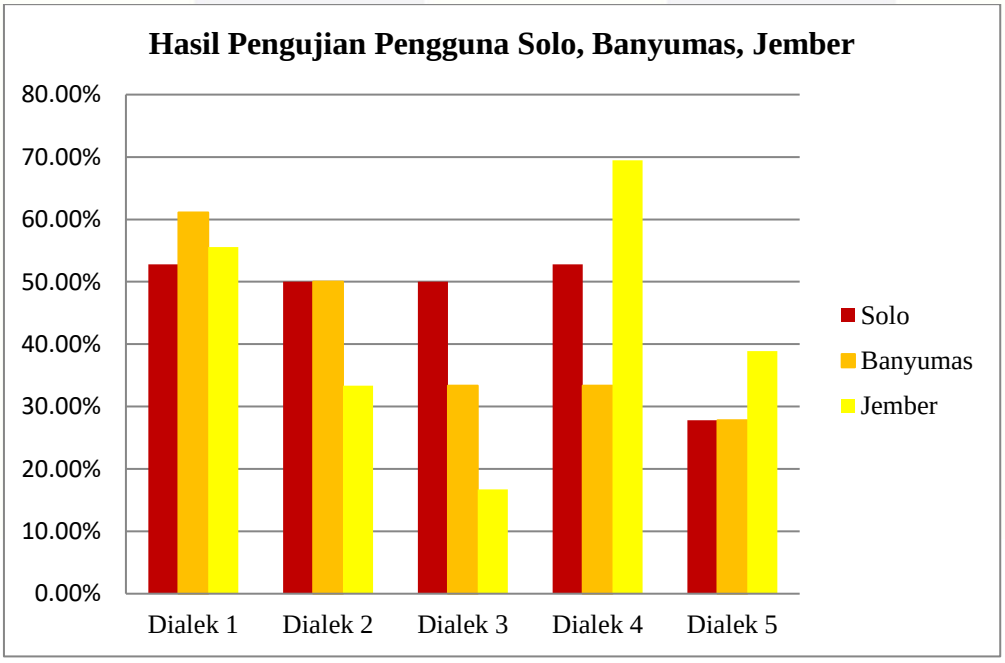
Tabel 4.2 Parameter Acuan Performa

Parameter	Nilai
Hiddensize 1	300
Hiddensize 2	400
L2WeightRegulazitation H1	0,0003
sparsityRegulazitation H1	2
L2WeightRegulazitation H2	0,0003
sparsityRegulazitation H2	2
Epoch	700

Untuk pengujian performa sistem dilakukan dengan kombinasi dari dialek Jawa Solo, Jawa Banyumas, Jawa Malang dan Jawa Jember. Adapun pembagiannya sebagai berikut :

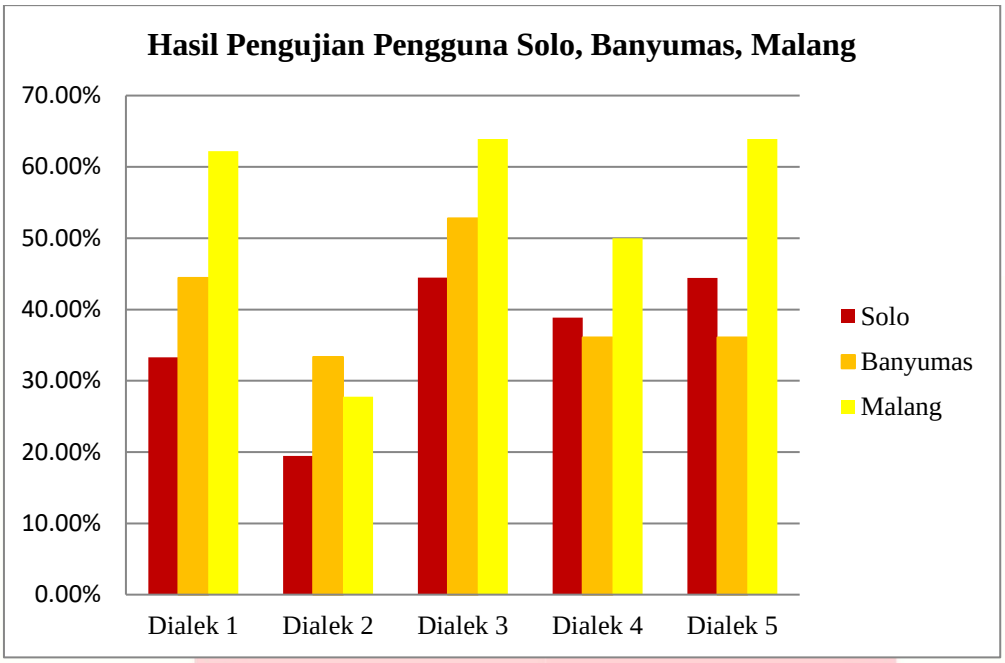
a. Dialek Jawa Solo, Jawa Banyumas dan Jawa Malang
b. Dialek Jawa Solo, Jawa Banyumas dan Jawa Jember
c. Dialek Jawa Solo, Jawa Malang dan Jawa Jember
d. Dialek Jawa Banyumas, Jawa Malang dan Jawa Jember

Hasil total dan rata-rata akurasi pengujian pengguna dari semua pengujian sebagai berikut :



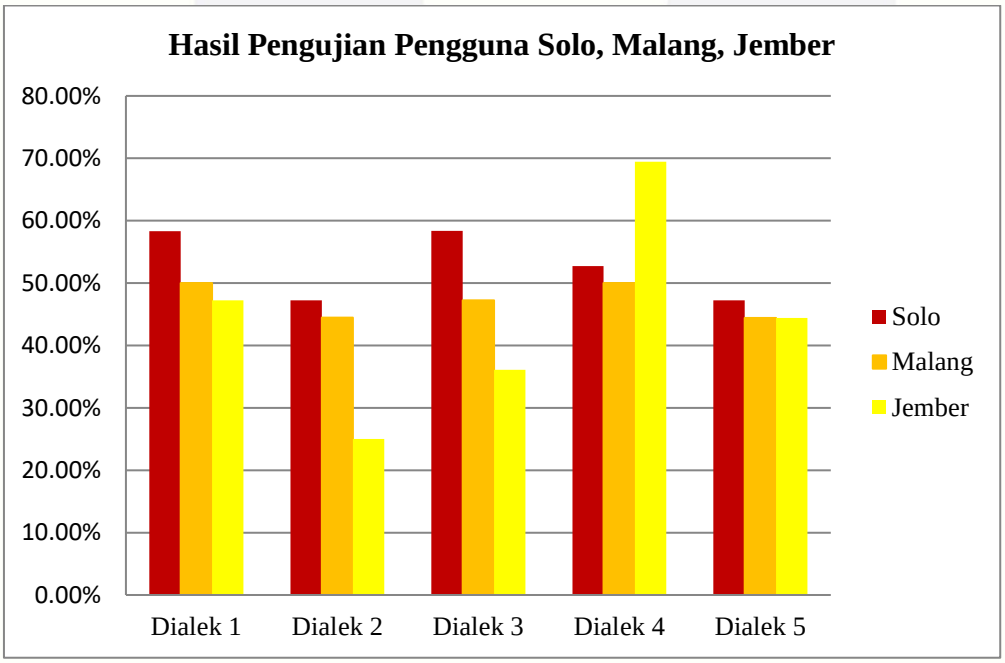
Gambar 4.1 Grafik Rata-rata dari Pengujian Dialek Solo, Banyumas, Jember

Dari gambar 4.1 menunujukkan bahwa dialek 1 merupakan pengujian terbaik dikarenakan hasil nilai rata-rata dialek Solo, Banyumas, Jember diatas 50%



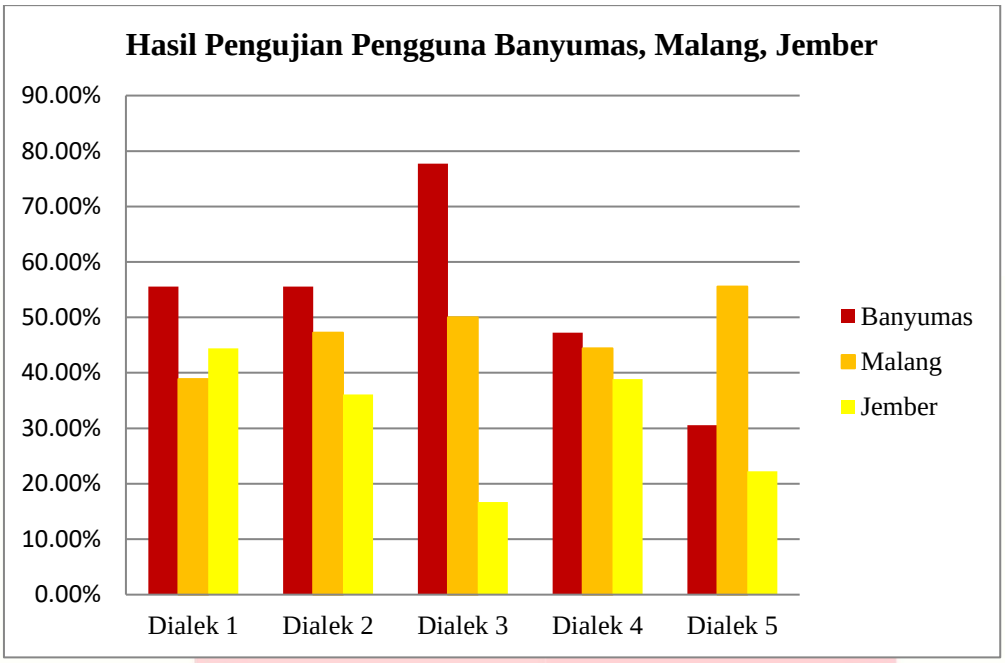
Gambar 4.2 Grafik Rata-rata dari Pengujian Dialek Solo, Banyumas, Malang

Dari gambar 4.2 menunujukkan bahwa dialek 3 merupakan pengujian terbaik dikarenakan terdapat dua dialek Banyumas dan Malang yang berada diatas 50% walaupun terdapat dialek Solo yang rata-ratanya masih dibawah 50%



Gambar 4.3 Grafik Rata-rata dari Pengujian Dialek Solo, Malang, Jember

Dari gambar 4.3 menunujukkan bahwa dialek 4 merupakan pengujian terbaik dikarenakan hasil nilai rata-ratanya dari ketiga dialek Solo, Malang, Jember berada diatas 50%



Gambar 4.4 Grafik Rata-rata dari Pengujian Dialek Banyumas, Malang, Jember

Dari gambar 4.4 menunujukkan bahwa dialek 3 merupakan pengujian terbaik dikarenakan terdapat dua dialek Banyumas dan Malang yang berada diatas 50% walaupun terdapat dialek Jember yang rata-ratanya masih dibawah 50%

5. Kesimpulan

1. Pada proses pengujian validasi sistem didapatkan bahwa tingkat akurasi ditentukan dari nilai yang diberikan untuk setiap parameter.
2. Agar mendapatkan nilai akurasi yang baik maka nilai dari parameter *L2WeightRegulazitation* dibuat kecil, semakin kecil nilai dari *L2WeightRegularization* semakin baik tingkat akurasi yang dihasilkan oleh sistem. Sedangkan *SparsityRegularization* digunakan untuk mengontrol bobot neuron untuk setiap *layer*.Dalam *autoencoder* parameter *hiddensize* akan menghasilkan akurasi yang baik saat *hidden layer* dibuat lebih kecil dari *inputsize*nya.Dalam *Deep Learning* semakin banyak pembelajaran yang dilakukan akan semakin baik hasil akurasi yang diberikan, oleh karena itu semakin besar nilai dari *epoch* maka semakin besar juga tingkat akurasi yang dihasilkan tetapi akan semakin lama proses untuk pembelajarannya
3. Pada proses pengujian peforma sistem berdasarkan besar data didapatkan seharusnya hasil akurasi yang maksimal didapatkan 100% saat data latih 80% dan 90% lebih banyak dibandingkan dari data uji tetapi dalam sistem ini agak berbeda karena jumlah kalimatnya dan susunan bahasa per daerahnya yang berbeda sehingga hasilnya kurang maksimal.
4. Dalam pengujian performa dibuat susunan kombinasi dari susunan dialek Jawa Solo, Jawa Banyumas, Jawa Malang, dan Jawa Jember sehingga memudahkan saat proses pembelajaranya.

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 17 September 2018)

[2] Drs. Abdul Rachman Patji, "139 bahasa daerah di Indonesia terancam punah," 2016. [Online]. Available: http://lipi.go.id/lipimedia/139-bahasa-daerah-di-indonesia-terancam-punah/15938. [Accessed: 17-Sep-2018].

[3] Margaret Rouse, "Speech Recognition" [online] (http://searchcrm.techtarget.com/definition/speech-recognition diakses tanggal 17 September 2018)

[4] T. T. Manunggal and D. Arifianto, "On development deep neural network speech synthesis using vector quantized acoustical feature for isolated bahasa Indonesia words," *2016 Conf. Orient. Chapter Int. Comm. Coord. Stand. Speech Databases Assess. Tech. O-COCOSDA 2016*, no. October, pp. 105–109, 2017.

[5] K. Li, H. Meng, T. Chinese, H. Kong, and H. K. Sar, "Mispronunciation Detection and Diagnosis in L2 English Speech Using Multi - Distribution Deep Neural Networks," pp. 255–259, 2015.

[6] Hu,Yu Hen dan Jenq-Neng Hwang. "Handbook of Neural Network Signal Processing". Florida: CRC Press LLC. (2002).

[7] 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.

[8] V. Mitra, G. Sivaraman, H. Nam, C. Y. Espy-Wilson, and E. Saltzman, "Articulatory features from deep neural networks and their role in speech recognition," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP 2014*, pp. 3017–3021, 2014.

[9] S. Xiong, W. Guo, and D. Liu, "The Vietnamese speech recognition based on rectified linear units deep neural network and spoken term detection system combination," *Proc. 9th Int. Symp. Chinese Spok. Lang. Process. ISCSLP 2014*, pp. 183–186, 2014.

[10] W. Hu, M. Fu, and W. Pan, "Primi Speech Recognition Based on Deep Neural Network," pp. 667–671, 2016.

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

[12] Andreas Chandra, "Perbedaan *Supervized* dan *Unsupervized Learning*" , 2017. [Online](https://datascience.or.id/article/Perbedaan-Supervised-and-Unsupervised-Learning-5a8fa6e6 diakses tanggal 23 Maret 2019)