

DEEP NEURAL NETWORK UNTUK PENGENALAN UCAPAN PADA BAHASA SUNDA DIALEK TENGAH TIMUR (MAJALENGKA)

DEEP NEURAL NETWORK FOR SPEECH RECOGNITION ON SUNDANESE LANGUAGE OF THE MIDDLE EAST DIALECT (MAJALENGKA)

¹Dede Nur Fathurrahman, ²Andrew Brian Osmond, ³Randy Erfa Saputra

¹²³Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom
¹dedenur@student.telkomuniversity.ac.id, ²abosmond@staff.telkomuniversity.ac.id,
³resaputra@staff.telkomuniversity.ac.id

Abstrak

Indonesia merupakan negara besar dengan memiliki banya keberagaman budaya dan suku sehingga Indonesia memiliki banyak bahasa atau pun dialek yang berbeda beda satu daerah dengan daerah yang lainnya, dalam hal ini penulis membuat suatu program atau aplikasi *speech recognition* dengan metode *Deep Neural Network* untuk pengenalan ucapan bahasa sunda dialek Majalengka. Berbagai penelitian dalam pengolahan sinyal suara telah banyak di kembangkan salah satu penelitian yang menarik untuk dikembangkan adalah identifikasi dialek. Identifikasi dialek dilakukan untuk mendapat informasi lebih lengkap dari seseorang melalui logat suara dari berbagai daerah di Indonesia. Dengan menggunakan *Deep Learning* yang berfokus pada menemukan representasi fitur bertingkat yang dimana pada tingkatan fitur yang lebih tinggi merepresentasikan lebih banyak aspek abstrak dari data. Oleh karena itu, penulis akan membuat suatu aplikasi Speech Recognition dengan metode *Deep Neurel Network* (DNN) untuk mengenali bahasa sunda dialek Majalengka.

Keywords: *Deep Neural Networks, Speech Recognition, Dialek, Neural Networks, Deep Learning.*

Abstract

Indonesia is a big country with a lot of cultural diversity and ethnic groups so that Indonesia has many different languages or dialects different one region with other regions, in this case the authors make a program or application *speech recognition* with *Deep Neural Network* method for *speech recogniciton* of Sundanese language of Majalengka dialect. Various studies in the processing of voice signals have been widely developed one of the interesting research to develop is the identification of dialect. Identification of dialect is done to get more information from someone through voice accent from various regions in Indonesia. By using *Deep Learning* that focuses on finding representation of multilevel features which at a higher level feature represent more abstract aspects of the data. Therefore, the author will make an application of Speech Recognition with *Deep Neurel Network* (DNN) method to recognize on Sundanese language of Majalengka dialect.

Keywords : *Deep Neural Networks, Speech Recgnition, Dialek, Neural Networks, Deep Learning.*

1. Pendahuluan

1.1 Latar Belakang

Bahasa sunda adalah bahasa daerah yang dipakai di Jawa Barat, sebagian Jawa Tengah sebelah barat, dan sebagian Daerah Ibu Kota Jakarta. Bahasa Sunda memiliki variasi Bahasa, atau undak-unsuk, yang terbentuk karena hubungan antara pembicara dan lawan bicara. Undak-usuk dalam bahasa Sunda (undak-usuk; tingkatan bahasa; unggah-ungguh; tata krama) ada 3 tingkatan yaitu basa kasar, basa loma (akrab), basa hormat/lemes (santun).

Dialek adalah varian dari sebuah bahasa menurut pemakai. Berbeda dengan ragam bahasa yaitu varian ini masih terdiri dari satu bahasa yang sama. Variasi ini berbeda satu sama lain, tetapi masih banyak menunjukkan kemiripan sehingga belum pantas disebut bahasa yang berbeda, pada bahasa sunda ini memiliki 6 dialek yaitu :

Barat (Bahasa Banten), Dialek Utara, Dialek Selatan (Priangan), Dialek Tengah Timur, Dialek Timur Laut (termasuk Bahasa Sunda Cirebon), Dialek Tenggara.

Walaupun bahasa sunda digunakan oleh banyak orang, riset Automatic Speech Recognition (ASR) untuk mengenali bahasa Sunda masih terbilang sedikit dibanding dengan ASR untuk mengenali bahasa lainnya. Pengenalan ASR bahasa Sunda memiliki beberapa kendala diantaranya : kurangnya data suara, banyaknya dialek bahasa Sunda, dan lainnya. Dalam hal ini, bagaimana suatu aplikasi ASR dapat memahami dan mengenal bahasa Sunda dialek tengah timur menggunakan metode deep neural network.

Pada umumnya, orang Indonesia yang berbeda daerah terkendala saat berkomunikasi dikarenakan ketidaktahuan dari bahasa daerah tersebut, sehingga dapat menimbulkan kesalahpahaman saat berkomunikasi. Dalam tugas akhir ini, metode deep neural network akan diterapkan pada ASR dengan menggunakan data ucapan bahasa sunda untuk memudahkan komunikasi antar daerah di Indonesia khususnya masyarakat Jawa Barat. Dengan demikian, sistem yang dibangun diharapkan dapat memiliki performansi yang baik.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, dapat dirumuskan permasalahan yang dihadapi adalah bagaimana membuat aplikasi Automatic Speech Recognition untuk pengenalan Bahasa Sunda dialek Tengah Timur dengan metode Deep Neural Network dan memproses suatu masukan berupa suara ke suatu aplikasi agar dimengerti oleh aplikasi dan dapat mengklasifikasikan masukan kedalam Bahasa Sunda dialek Majalengka serta memberikan bobot nilai untuk masukan tersebut.

1.3 Tujuan

Tujuan yang ingin dicapai dari tugas akhir ini adalah :

1. Mengimplementasikan metode deep neural network pada Automatic Speech Recognition bahasa Sunda dialek tengah timur (Majalengka).
2. Mengukur akurasi kinerja deep neural network.
3. Mengenali dialek Majalengka dengan aplikasi Automatic Speech Recognition.

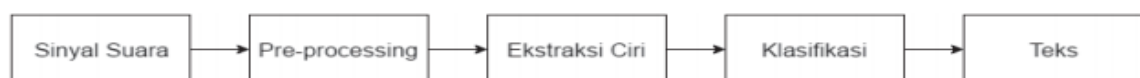
1.4 Identifikasi Masalah

Dengan merujuk pada rumusan masalah yang telah dijelaskan, beberapa identifikasi masalah yang akan dibahas dalam Tugas Akhir ini adalah bagaimana membuat sistem Speech Recognition untuk pengenalan Bahasa Sunda dialek Majalengka 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

2.1 Speech Recognition

Speech Recognition merupakan teknologi yang dapat mengubah ucapan yang berupa sinyal suara menjadi tulisan. Selain untuk mengenali ucapan, ASR dapat digunakan untuk belajar bahasa baru . Hal ini dikarenakan ASR dapat mengetahui letak kesalahan dari ucapan pengucap serta memberikan skor berdasarkan ucapannya. Dalam tugas akhir ini, sistem ASR dibangun dengan tujuan tersebut [1], yaitu untuk membantu pengguna belajar Bahasa Sunda.



Gambar 2. 1 Proses Umum Automatic Speech Recognition

Pada gambar 2.1, masukan sinyal suara akan di pre-processing terlebih dulu untuk mempersiapkan dan mengolah data awal sehingga data yang digunakan merupakan data yang sudah siap pakai dan dapat mempermudah proses-proses dalam tahapan berikutnya, lalu data yang sudah di pre-processing ini di ekstraksi ciri nya dan dipisahkan, setelah itu, sistem akan membandingkan hasil ekstraksi yang sudah dilakukan dengan database

yang tersedia. Parameter yang dibandingkan ialah tingkat penekanan suara yang kemudian akan dicocokkan dengan template database yang tersedia.

Parameter yang dibandingkan ialah tingkat penekanan suara yang kemudian akan dicocokkan dengan template database yang tersedia. Pada zaman sekarang, *Speech Recognition* sudah banyak digunakan, namun sayangnya aplikasi yang menggunakan teknologi *Speech Recognition* lebih banyak dirancang untuk ter *built-in* pada *mobile phone* tertentu, contoh : (Siri pada IOS), dan pada umumnya, teknologi Automatic Speech Recognition digunakan untuk salah satu alat bantu menjalankan perintah – perintah singkat, seperti mengetik teks, pencarian kontak, dll [1].

2.2 Deep Learning Algorithm

Deep Learning Algorithm merupakan cabang dari machine learning yang fokus ke arah *Artificial Intelligence* dimana algoritma ini memiliki satu set algoritma yang mencoba model abstraksi *high-level* pada suatu data dengan menggunakan lapisan pengolahan dengan struktur yang kompleks atau sebaliknya dan terdiri dari beberapa transformasi *non-linear*. Algoritma ini berdasarkan representasi dari *machine learning*. Yang menyebabkan *deep learning* berbeda dengan *normal neural network* yaitu karena *deep learning* memiliki *hidden layer*. *Deep learning* dapat melakukan *training* dengan diawasi maupun tanpa diawasi. Arsitektur *deep learning* ini memiliki beberapa macam variasi, seperti *deep neural networks*, *convolutional deep neural networks*, *recurrent neural networks*, dan *deep belief networks* [2].

Dalam kasus klasifikasi *speech recognition* ini akan digunakan *deep neural network* (DNN). Deep neural network merupakan jaringan saraf tiruan dengan beberapa hidden layer yang terdapat diantara *input layer* dan *output layer*. DNN memiliki layer yang banyak dan juga memiliki koneksi dan bobot yang banyak juga [3]. DNN memiliki hidden layer yang dapat mempelajari level yang lebih tinggi [4]. DNN ini lebih baik untuk pemodelan akustik jika menggunakan lapisan convolutional dalam menurunkan bobot frekuensi di dekatnya [5].

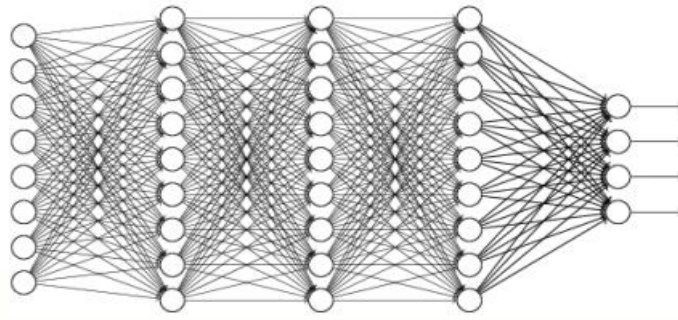
2.3 Neural Network

Neural Network merupakan kategori ilmu *Soft Computing*. *Neural Network* sebenarnya mengadopsi dari kemampuan otak manusia yang mampu memberikan stimulasi/rangsangan, melakukan proses, dan memberikan *output*. *Output* diperoleh dari variasi stimulasi dan proses yang terjadi di dalam otak manusia. Kemampuan manusia dalam memproses informasi merupakan hasil kompleksitas proses di dalam otak. Misalnya, yang terjadi pada anak-anak, mereka mampu belajar untuk melakukan pengenalan meskipun mereka tidak mengetahui algoritma apa yang digunakan. Kekuatan komputasi yang luar biasa dari otak manusia ini merupakan sebuah keunggulan di dalam kajian ilmu pengetahuan [6]. Fungsi dari Neural Network diantaranya adalah:

- Pengklasifikasian pola
- Memetakan pola yang didapat dari input ke dalam pola baru pada output
- Penyimpan pola yang akan dipanggil kembali
- Memetakan pola-pola yang sejenis
- Pengoptimasi permasalahan
- Prediksi

2.4 Deep Neural Network

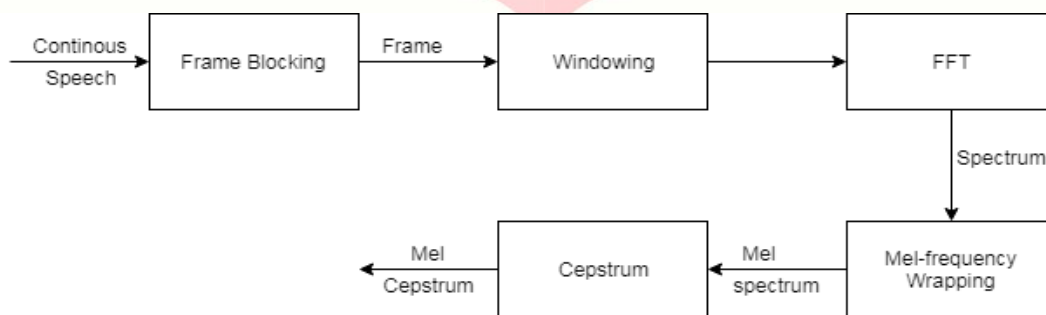
Deep Neural Network (DNN) merupakan salah satu tipe dari *neural network*. DNN ini terdiri dari beberapa *hidden units* dengan koneksi antar *layer* tetapi tidak ada koneksi antar *units* pada setiap *layer*-nya. Metode ini memiliki arsitektur yang serupa dengan arsitektur pada *Artificial Neural Network* (ANN), dengan *supervised training* [7]. Dengan mengidentifikasi masukan dan mencocokkannya dengan pola yang sudah ada. Adapun kelebihan *Deep Learning 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 [5].



Gambar 2.2 Deep Neural Network Layer

2.5 Mel-Frequency Cepstral Coefficient (MFCC)

Mel Frequency Cepstrum Coefficients (MFCC) merupakan satu metode yang banyak dipakai dalam bidang *speech recognition*. MFCC adalah salah satu metode yang digunakan untuk melakukan *feature extraction* (ekstraksi ciri), sebuah proses yang mengkonversikan sinyal suara menjadi beberapa parameter [8]. Analisis suara pada *mel-frequency* didasarkan pada persepsi pendengaran manusia, karena telinga manusia telah diamati dapat berfungsi sebagai filter pada frekuensi tertentu. Filter ini digunakan untuk menangkap karakteristik fonetis penting dari sebuah ucapan. MFCC digambarkan dalam skala mel-frekuensi yang merupakan frekuensi linier dibawah 1000 Hz dan logaritmik di atas 1000 Hz [9]. Dibawah ini adalah gambar diagram blok untuk proses MFCC.



Gambar 2.3 Diagram Blok MFCC

2.5.1 Frame Blocking

Frame blocking merupakan proses segmentasi sinyal suara yang masuk akan dipotong dan dibagi ke dalam beberapa *frame*, agar didapatkan karakteristik sinyal suara yang lebih stabil.

2.5.2 Windowing

Proses *windowing* dilakukan pada setiap *frame* bertujuan agar meminimumkan terjadi hilangnya informasi pada sinyal suara. Hamming window digunakan karena mempunyai hasil yang baik dalam menyaring sinyal yang akan dianalisis. Karena pada proses *frame blocking* dapat menyebabkan sinyal menjadi diskontinuitas. Oleh karena itu proses *windowing* pada setiap *frame* dilakukan

2.5.3 Fast Fourier Transform

Sinyal suara yang ada masih dalam domain waktu sehingga perlu dilakukan konversi dari domain waktu ke domain frekuensi. Untuk mendapatkan sinyal dalam domain frekuensi dari sebuah sinyal diskrit, salah satu metode yang digunakan adalah *Fast Fourier Transform* (FFT). FFT dilakukan terhadap masing-masing *frame* dari sinyal yang telah di *windowing*.

2.5.4 Mel-Frequency Wrapping

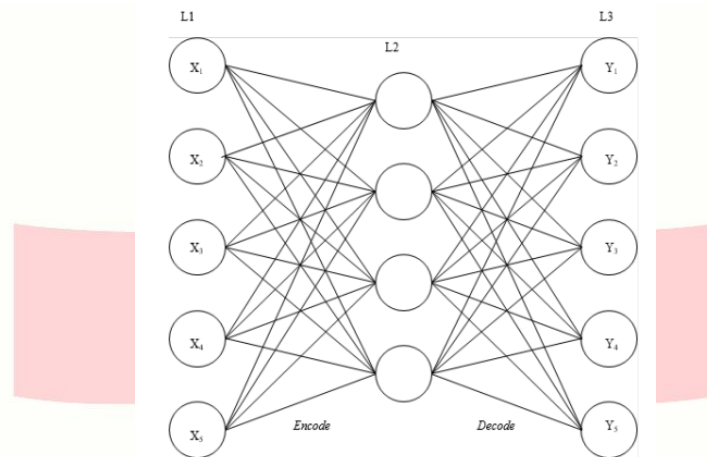
Pada proses ini dilakukan pengubahan skala frekuensi menjadi skala Mel. Penerimaan sinyal suara untuk frekuensi rendah (<1000) bersifat linear, sedangkan untuk frekuensi tinggi (>1000) bersifat logaritmik. diatas 1KHz.

2.5.5 Cepstrum

Pada tahap ini sinyal suara dari domain frekuensi (*mel-freequency*) akan dikembalikan ke domain waktu menggunakan *Discrete Cosine Transform* (DCT).

2.6 Autoencoder

Autoencoder adalah *neural network* yang digunakan untuk mempelajari pengkodean data yang efisien dengan cara yang tidak diawasi. Tujuan dari *autoencoder* adalah untuk mempelajari representasi (*encoding*) dari satu set data, biasanya untuk tujuan pengurangan dimensi. Baru-baru ini, konsep *autoencoder* telah menjadi lebih banyak digunakan untuk belajar model data generative [10]. Beberapa AI pada tahun 2010 telah melibatkan *autoencoder* yang ditumpuk di dalam *deep neural network*. *Autoencoder* belajar untuk mengompres data dari lapisan *input* ke *short code*, dan kemudian mengompres kode itu menjadi sesuatu yang sangat cocok dengan data asli.

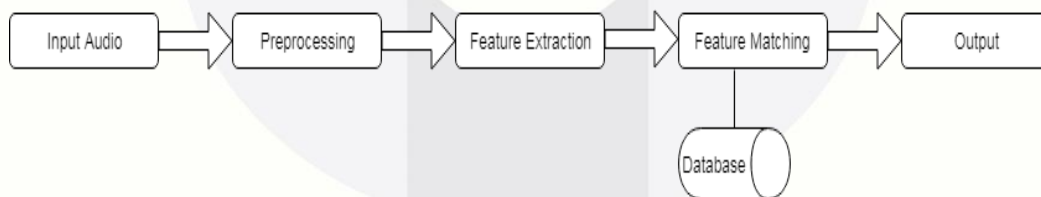


Gambar 2.3 Autoencoder Layer

Secara arsitektur, bentuk paling sederhana dari *autoencoder* adalah *feed forward, non-recurrent neural network* yang sangat mirip dengan banyak *perceptron layer* tunggal yang membuat Multi Layer Perceptron (MLP), memiliki *input layer* dan *output layer* dengan *hidden layer* sebagai penghubungnya. Jumlah *node* di *layer output* memiliki jumlah yang sama dengan jumlah *node* di *layer input* dengan tujuan merekonstruksi inputnya sendiri. Oleh karena itu, *autoencoder* adalah model pembelajaran yang tidak diawasi.

3. Perancangan

Pada tugas akhir ini, dirancang sistem klasifikasi dialek Bahasa sunda dengan menggunakan metode Deep Neural Network (DNN). Berikut adalah gambaran umum perancangan sistem automatic speech recognition.



Gambar 3.1 Gambaran Umum Sistem

Secara umum, sistem Automatic Speech Recognition mengenali dan memahami kata-kata yang diucapkan dengan cara digitalisasi kata dan mencocokkan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan di database sistem. Ada[un tahapan dalam prosesnya yaitu:

1. Tahap penerimaan masukan : masukan berupa sampel audio rekaman yang diucapkan berasal dari narasumber.
2. Tahap preprocessing: mempersiapkan dan mengolah data awal sehingga data yang digunakan merupakan data yang sudah siap pakai dan dapat mempermudah proses-proses dalam tahapan berikutnya.
3. Tahap ekstraksi ciri : Pada Tahap ini dilakukan ekstraksi ciri menggunakan MFCC agar mendapat nicali ciri dari masing-masing data dan menyimpan data yang berupa suara sekaligus pembuatan basis data sebagai pola.
4. Tahap perbandingan : Pada tahap ini sistem akan mencocokkan data baru dengan data latih yang ada di dalam database yang tersedia. Dimulai dengan proses konversi sinyal digital hasil proses ekstraksi

kedalam bentuk spektrum suara yang akan dianalisa dengan membandungkannya dengan pola suara basis data. Sebelumnya data suara masukan dipisahkan dan diproses satu per satu berdasarkan urutan

5. Tahap Validasi : Pada tahap ini sistem akan mengambil keputusan apakah masukan dapat dikenali atau tidak.

3.1 Pengambilan Data

Pengambilan data audio dilakukan melalui hasil perekaman suara narasumber dengan sample kalimat seperti tabel di bawah

Tabel 3.1 Tabel Kata/Kalimat Data Uji dan Data Latih

	Kata 1	Kata 2	Kata 3	Kata 4	Kata 5	Kata 6	Kata 7	Kata 8	Kata 9
Loma	Tempat	Dahar	deukeut	Didieu	Belah	Ndi			
Lemes	Tempat	Tuang	cakeut	Dieu	Palih	Mana			
Indonesia	Tempat	Makan	dekat	Sini	sebelah	Mana			
Loma	Ngajaran	Budak	leutik	Lempang					
Lemes	Ngajaran	Murang	kalih	Mapah					
Indonesia	Mengajari	Anak	Kecil	Berjalan					
Loma	Ngadengekeun	Lagu	Bari	Nyetiran	Teh	Picilakaeun	komo	Bari	Nelepon
Lemes	Ngadangukeun	Lagu	Bari	Nyetiran	Teh	Picilakaeun	komo	Bari	Nelepon
Indonesia	Mendengarkan	Lagu	sambil	Menyetir	Bikin	Celaka	apalagi	Sambil	Telepon
Loma	Lamun	Budak	gering	Panas	Make	Obat	Naen		
Lemes	Lamun	Murangalih	gering	Panas	Pake	Obat	Naon		
Indonesia	Kalo	Anak	Sakit	Panas	pakai	Obat	Apa		
Loma	Kumaha	Cara	Dahar	Anu	Sae				
Lemes	Kumaha	Cara	Neda	Anu	Sae				
Indonesia	bagaimana	Cara	makan	Yang	Bagus				
Loma	Keur	Meuli	Lauk	Di pasar					
Lemes	Nuju	Meser	Lauk	Di pasar					
Indonesia	Lagi	Beli	Ikan	Di pasar					

Kalimat yang dipakai hanya kalimat 'loma' saja, karena kalimat loma adalah kalimat yang biasa digunakan oleh orang Majalengka. Pengambilan data dibagi menjadi 3 narasumber yaitu:

1. Orang Sunda Majalengka, yaitu hasil dari perekaman narasumber yang berasal dari daerah Majalengka.
2. Orang Sunda Garut, yaitu hasil dari perekaman narasumber yang berasal dari daerah Garut.
3. Orang Sunda Bogor, yaitu hasil dari perekaman narasumber yang berasal dari daerah Bogor.

4. Pengujian dan Analisis

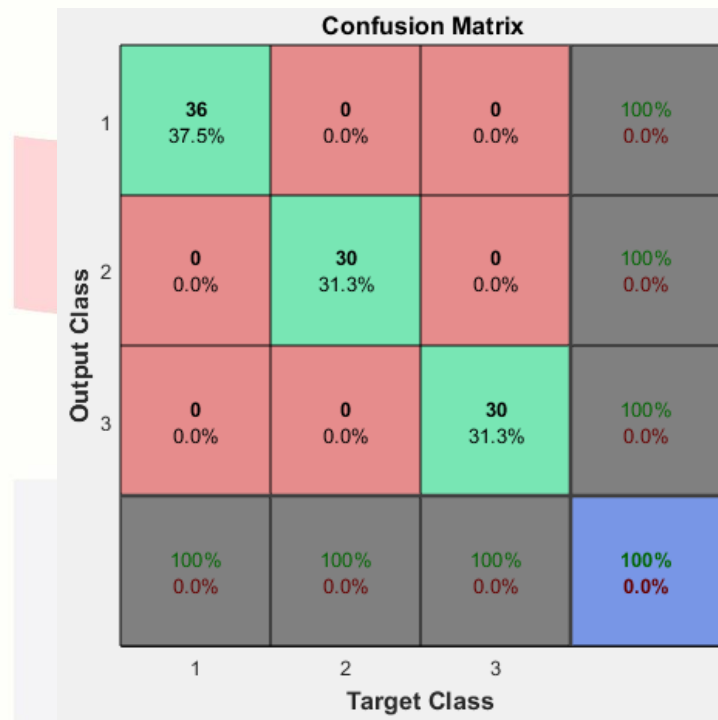
Pengujian yang dilakukan untuk diketahui apakah sistem yang telah dibuat sudah berjalan sesuai dengan yang diinginkan atau belum, maka diperlukan pengujian pada sistem untuk mengetahui tingkat keberhasilan dengan menganalisis beberapa parameter yang menjadi tolak ukur keberhasilan sistem. Pada penelitian ini dilakukan beberapa pengujian untuk menguji keakurasian pada metode yang telah digunakan pada sistem dengan cara merubah parameter yang terdapat dalam sistem agar mendapatkan akurasi terbaik. Tujuan dari pengujian terhadap sistem ini adalah sebagai berikut :

1. Mengetahui apakah metode dari *Deep Neural Network* cocok pada pengklasifikasian pengenalan suara serta mengetahui tingkat akurasi yang dihasilkan sistem dari hasil implementasi metode.
2. Mengetahui pengaruh *L2WeightRegulazitation*, *sparsityRegulazitation*, *hiddensize*, dan *epoch* terhadap tingkat akurasi sistem.

Table 4.1 Pengujian Validasi Parameter

Parameter	Nilai
<i>Hiddensize 1</i>	300
<i>Hiddensize 2</i>	300
<i>L2WeightRegulazitation H1</i>	0,0002
<i>sparsityRegulazitation H1</i>	5
<i>L2WeightRegulazitation H2</i>	0.0002
<i>sparsityRegulazitation H2</i>	5
<i>Epoch</i>	500

Hasil dari pengujian dengan parameter pada tabel 4.1 ditunjukkan pada gambar 4.1



Gambar 4.1 Hasil Klasifikasi Pengujian

4.1 Ringkasan Pengujian

Setelah dilakukan pengujian sistem dengan merubah setiap parameter yang ada untuk mendapatkan akurasi maksimal, dilakukan analisis untuk mengetahui sebab naik atau turunnya tingkat akurasi.

1. Jika nilai dari L2WeightRegularization dan SparsityRegularization terlalu besar atau terlalu kecil dapat menyebabkan Underfitting. Underfitting adalah keadaan ketika kinerja model bernilai buruk, baik pada training maupun testing data. Underfitting terjadi akibat model yang terlalu tidak fleksibel (memiliki kemampuan yang rendah untuk mengestimasi variasi fungsi).
2. L2WeightRegulazitation mempengaruhi tingkat akurasi, karena kegunaannya dari L2WeightRegulazitation sendiri yaitu memodifikasi fungsi kesalahan dengan memberikan nilai bobot untuk setiap jaringan. Semakin kecil nilai L2WeightRegulazitation semakin besar tingkat akurasi yang dihasilkan.
3. SparsityRegularization adalah parameter untuk pembobotan dari nilai fraksi yang terlalu besar disetiap neuron pada setiap layer. Jika nilai dari sparityregularization yang diberikan kecil (mendekati 0) tingkat akurasi yang dihasilkan akan semakin besar (tergantung dari nilai neuron sebelumnya).
4. Untuk parameter dari Hiddensize cukup mempengaruhi tingkat akurasi, karena didalam autoencoder ketika jumlah neuron dilapisan tersembunyi kurang dari ukuran input makan autoencoder akan belajar untuk merepresentasikan merubah input ke dalam bentuk dimensi lebih kecil terkompresi dari input untuk iterasi selanjutnya. Ide yang bagus untuk membuat Hiddensize lebih kecil dari inputsize.

Hiddensize juga mempengaruhi lamanya kinerja sistem, semakin besar nilai dari hiddensize semakin lama kinerja sistem.

5. Parameter epoch mempengaruhi tingkat akurasi, karena semakin banyak epoch maka semakin banyak pula pembelajaran yang dilakukan pada setiap training dan juga pencocokan ciri. Jika pembelajaran yang sistem lakukan semakin banyak maka sistem akan lebih mengenali pola/ciri dari data yang masuk untuk dijadikan database. Epoch maksimal pada sistem ini yaitu 1000 epoch pada setiap layernya.

Setelah mendapatkan nilai parameter ideal untuk menghasilkan tingkat akurasi yang maksimal, pengujian selanjutnya yaitu pengujian terhadap performa sistem berdasarkan banyak data. Untuk data training dan data testing dibuat bervariasi untuk melihat perbedaan akurasi dari setiap variasi data tersebut.

5. Kesimpulan

Berdasarkan dari hasil Tugas Akhir ini, dapat disimpulkan bahwa :

1. Pada proses pengujian validasi sistem didapatkan bahwa tingkat akurasi ditentukan dari nilai yang diberikan untuk setiap parameter.
2. Untuk mendapatkan akurasi yang baik 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 untuk epoch bernilai besar agar pembelajaran semakin banyak.
3. Pada proses pengujian peforma sistem perbedaan kata bisa menyebabkan menurunnya tingkat akurasi.

Referensi

- [1] N. Aini, "Pengertian Speech Recognition (ASR)", (<https://plus.google.com/113146895960339762491/posts/QY3cJAPMTgf>), diakses tanggal 26 Juli 2018.
- [2] F. B. Ahmad, "Panduan Pemula dalam Memahami Machine Learning, Deep Learning, dan AI", (<https://id.techinasia.com/memahami-istilah-terkait-ai-dan-perkembangannya>) diakses tanggal 26 Juli 2018.
- [3] A. Ebru, Sainath T. N., K. Brian, R. Bhuvana, "Deep Neural Network Language Models," Association for Computational Linguistics, pp. 20-21, 2012.
- [4] Ciresan D. C., G. Alessandro, Gambardella L. M., S. Jorgen, "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Image," Advances in Neural Information Processing Systems 25 (NIPS 2012), 2012.
- [5] D. Li, H. Geoffrey, K. Brian, "New Types of Deep Neural Network Learning for Speech Recognition and Related Applications: an Overview," IEEE International Conference on Acoustics, Speech and Signal Processing, 2013.
- [6] D. Suhartono, "Dasar Pemahaman Neural Network", (<https://socs.binus.ac.id/2012/07/26/konsep-neural-network/>), diakses tanggal 26 Juli 2018.
- [7] N. L. Roux, Y. Bengio, "Representational Power of Restricted Boltzmann Machines and Deep Belief Networks," Neural Computation, vol. 20, pp. 1631-1649, 2008.
- [8] Manunggal, HS, "Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara dengan Menggunakan Analisa MFCC Feature Extraction. Surabaya", Universitas Kristen Petr, 2005
- [9] M. N. Rabbani, "Implementation of Voice Recognition Based Key Using Mel Frequency Cepstral Coefficient (MFCC)", e-Proceeding of Engineering : Vol.3, No.3 December 2016
- [10] D. P. Kingma, W. Max, "Auto-Encoding Variational Bayes", arSiv:1312.6114, 2013