

## RECURRENT NEURAL NETWORK UNTUK PENGENALAN UCAPAN PADA DIALEK MANADO

### *Recurrent Neural Network For Speech Recognition On Manado Dialect*

<sup>1</sup>Erwin Lapijan, <sup>2</sup>Andrew Briand Osmond, <sup>3</sup>Randy Erfa Saputra

<sup>123</sup>Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom  
<sup>1</sup>erlapijan@gmail.com, <sup>2</sup>abosmond@telkomuniversity.ac.id, <sup>3</sup>resaputra@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, karena perbedaan ini penulis akan membuat sebuah sistem yang akan membantu masyarakat agar tidak terjadi salah paham dalam mengartikan sebuah bahasa daerah.

*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. dalam hal ini akan membuat suatu sistem atau program *speech recognition* dengan metode *Recurrent Neural Network*.

Dari hasil penelitian yang dilakukan, dari nilai parameter tertentu didapatkan akurasi sebesar 88%. Setelah mendapatkan parameter ideal maka dilakukan pengujian keseluruhan sistem sebanyak sepuluh kali dan akurasi yang didapatkan rata-rata sebesar 87%. Dari pengujian tersebut dapat disimpulkan bahwa, semakin banyak epoch dan data latihan yang banyak akan meningkatkan tingkat akurasi.

**Keywords:** *Recurrent Neural Networks, Speech Processing, Dialek, Neural Networks.*

#### Abstract

Indonesia is a big country with a diversity of cultures and tribes so that Indonesia has many languages or dialects different from one region to another because of this difference the author will create an application that will help the community to avoid misunderstanding in interpreting a regional language.

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. in this case will create a system or program speech recognition with Recurrent Neural Network method.

From the results of research conducted, from the value of certain parameters obtained an accuracy of 88%. After obtaining the ideal parameters then do the overall test of the system as much as ten times and the average accuracy obtained by 87%. From these tests it can be concluded that, the more epoch and training data that many will increase the accuracy level.

**Keywords :** *Recurrent Neural Networks, Speech Processing, Dialek, Neural Networks.*

## 1. Pendahuluan

### 1.1 Latar Belakang

Bahasa merupakan sebuah alat komunikasi yang sangat penting dalam kehidupan manusia. Bahasa juga mencerminkan sebuah identitas daerah atau negara. Indonesia memiliki kebudayaan yang beragam. Perbedaan tersebut berdampak pada variasi penggunaan bahasa oleh masyarakat. Dan dialek merupakan variasi bahasa yang digunakan oleh sekelompok orang masyarakat pada suatu tempat tertentu.

Teknologi *Speech Recognition* merupakan teknologi yang memanfaatkan sinyal suara manusia sebagai masukan untuk dikenali oleh sistem, teknologi speech recognition tersebut yang kemudian dimanfaatkan dalam Tugas Akhir ini yaitu dalam mengenali suara dialek bahasa Manado.

Sebagian besar kata-kata dalam bahasa Manado sama seperti kata-kata dalam bahasa Indonesia. Karena bahasa Manado hanya digunakan untuk komunikasi lisan, tidak ada standar ortografi/tulisan yang pernah disahkan. Bahasa manado berhubungan dekat dengan bahasa indonesia. Perbedaannya yang paling mendasar adalah dengan

adanya kata-kata serapan dari bahasa Belanda dan Portugis, serta penggunaan kata “kita” sebagai kata ganti orang pertama tunggal.

Pada Tugas akhir ini dilakukan bertujuan untuk membangun sebuah sistem atau program dengan menggunakan metode *Recurrent Neural Network* pada sistem pengenalan dialek.

Dengan adanya teknologi yang baru ini penulis berharap akan memudahkan komunikasi antar daerah di Indonesia khususnya masyarakat kota Manado dengan tingkat akurasi berkisar di angka 70%.

## 1.2 Rumusan Masalah

Rumusan masalah dalam tugas akhir ini yaitu bagaimana membuat membuat sistem *Speech Recognition* untuk pengenalan Bahasa Sunda dialek Utara dengan menerapkan metode *Recurrent 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.

## 1.3 Tujuan

Tujuan dari Tugas Akhir ini adalah sebagai berikut:

1. Mengimplementasikan metode *Recurrent Neural Network* pada *Speech Recognition* dengan menggunakan dialek Manado
2. Mengukur ketepatan performansi kinerja *Recurrent neural Network*

## 1.4 Identifikasi Masalah

Beberapa identifikasi masalah yang akan dibahas dalam Tugas Akhir ini adalah sebagai berikut :

Bagaimana membuat membuat sistem *Speech Recognition* untuk pengenalan Bahasa Manado dialek Utara dengan menerapkan metode *Recurrent 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 *Recurrent neural network* untuk klasifikasi Bahasa Manado. Adapun teori-teori yang digunakan adalah sebagai berikut.

### 2.1 *Speech Recognition*

*Speech recognition* adalah sistem yang berfungsi untuk mengubah bahasa lisan menjadi tulisan atau dari suara menjadi sebuah text. Inputan sistem berupa suara manusia, selanjutnya sistem akan mengidentifikasi kata atau kalimat yang diucapkan dan menghasilkan teks yang sesuai dengan apa yang diucapkan. [8]

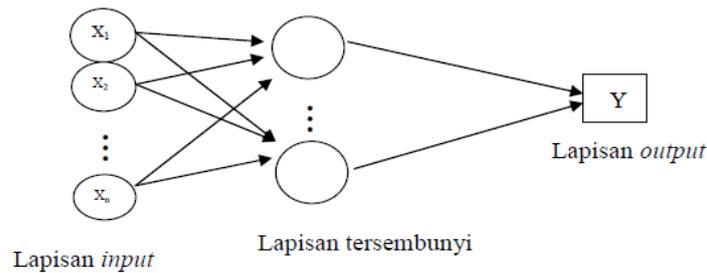
Teknologi ini memungkinkan suatu perangkat untuk mengenali dan memahami kata-kata yang diucapkan dengan cara digitalisasi kata dan mencocokkan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan dalam suatu perangkat. Kata-kata yang diucapkan diubah bentuknya menjadi sinyal digital dengan cara mengubah gelombang suara menjadi sekumpulan angka yang kemudian disesuaikan dengan kode-kode tertentu untuk mengidentifikasi 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. [8]

*Speech Recognition* berhubungan erat dengan bahasa sehari-hari. Bahasa merupakan bagian yang tidak terpisahkan dari kehidupan umat manusia. Bahasa dapat berupa catatan dari pengetahuan yang didapat dari kehidupan manusia dari generasi ke generasi berikutnya, sedangkan dalam bentuk lisan merupakan sarana komunikasi antar individu dalam kelompok masyarakat. Bahasa alami atau natural language adalah suatu bahasa yang diucapkan, ditulis atau diisyaratkan oleh manusia untuk berkomunikasi secara umum. [8]

## 2.2 Neural Network

Neural Network atau sering kita dengar dengan sebutan Jaringan Syaraf Tiruan (JST) merupakan kategori ilmu Soft Computing. Neural Network sebenarnya mengadopsi dari kemampuan otak manusia yang mampu memberikan stimulasi/rangsangan, melakukan proses, dan memberikan output.

Model jaringan syaraf ditunjukkan dengan kemampuannya dalam emulasi, analisis, prediksi dan asosiasi. Kemampuan yang dimiliki jaringan syaraf tiruan dapat digunakan untuk belajar dan menghasilkan aturan atau operasi dari beberapa contoh atau input yang dimasukkan dan membuat prediksi tentang kemungkinan output yang akan muncul atau menyimpan karakteristik input yang diberikan kepada jaringan syaraf tiruan.



Gambar 2.1 Jaringan Neural Network

Secara umum, lapisan pada JST dibagi menjadi tiga bagian:

1. Lapisan Masukan (input layer)

Lapisan masukan merupakan lapisan yang terdiri dari beberapa neuron yang akan menerima sinyal dari luar dan kemudian meneruskan ke neuron-neuron lain dalam jaringan. Lapisan ini diilhami berdasarkan ciri-ciri dan cara kerja sel-sel saraf sensori pada jaringan saraf biologi.

2. Lapisan tersembunyi (hidden layer)

Lapisan tersembunyi merupakan tiruan dari sel-sel syaraf konektor pada jaringan saraf biologis. Lapisan tersembunyi berfungsi meningkatkan kemampuan jaringan dalam memecahkan masalah. Konsekuensi dari adanya lapisan ini adalah pelatihan menjadi makin sulit atau lama.

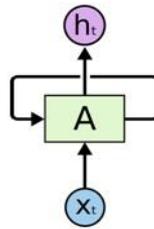
3. Lapisan keluaran (output layer)

Lapisan keluaran berfungsi menyalurkan sinyal-sinyal keluaran hasil pemrosesan jaringan. Lapisan ini juga terdiri dari sejumlah neuron. Lapisan keluaran merupakan tiruan dari sel saraf motor pada jaringan saraf biologis..

## 2.3 Recurrent Neural Network

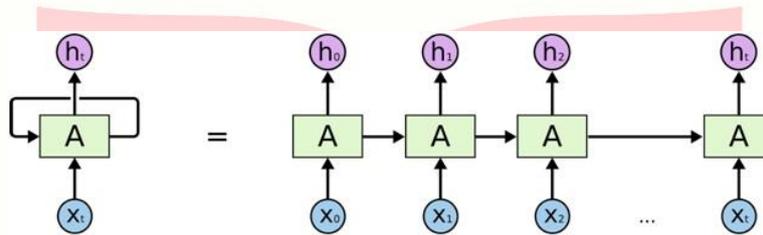
Pada umumnya, manusia tidak membuat keputusan secara tunggal setiap saat. Kita akan selalu memperhitungkan masa lalu dalam membuat sebuah keputusan. Cara berpikir seperti ini adalah dasar dari pengembangan Recurrent Neural Network. Sama seperti analogi tersebut, RNN tidak membuang begitu saja informasi dari masa lalu dalam proses pembelajarannya. Hal inilah yang membedakan RNN dari *Artificial Neural Network* biasa.[10]

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.[10]



Gambar 2.2 Loop RNN tertutup

Gambar 1 adalah visualisasi contoh potongan dari sebuah RNN A. RNN tersebut mendapat input  $x_t$  dan menghasilkan output  $h_t$ . Dan alur loop tersebut memungkinkan informasi untuk dapat dilempar dari satu step menuju step selanjutnya.[10].



Gambar 2.3 Loop RNN terbuka

Looping dari RNN sebenarnya akan memproses input dari skala waktu 0 sampai t. RNN akan memproses data input satu per satu secara sekuensial, *hidden layer* pun akan melempar data menuju ke *hidden layer* pada skala waktu selanjutnya. Begitu seterusnya secara sekuensial.[10]

### 3 Perancangan

Dalam penelitian ini sistem yang dibuat merupakan sistem yang dapat membedakan kata dengan melihat hasil akurasi. Sistem ini memiliki 2 bagian besar yaitu bagian training untuk membuat model yang akan digunakan dalam pengenalan kata dan bagian testing untuk melakukan scenario pengujian, untuk lebih jelasnya bias dilihat pada gambar..



Gambar 3.1 flowchart sistem

Metode yang digunakan dalam penelitian ini adalah Recurrent neural network yang berguna untuk membentuk pola yang akan mempresentasikan pengklasifikasian. Langkah utama yang dilakukan adalah pencarian sampel suara yang akan digunakan pada proses input sampel berupa audio rekaman dari narasumber yang berbicara bahasa manado untuk dijadikan data latihan..

Setelah data terkumpul maka dilakukan proses preprocessing dimana data masukan yang diolah agar dapat menghasilkan bentuk dan nilai setelah itu dilakukan ekstraksi ciri dengan menggunakan *Mel-Frequency Cepstral Coefficient* (MFCC) agar mendapatkan nilai ciri dari masing-masing data masukan. Setelah itu masuk kedalam neural network dimana menggunakan Recurrent neural network.

### 3.1 Pengambilan Data

Pengambilan data pada penelitian ini penulis melakukan dengan cara sebagai berikut :

1. Mencari narasumber dengan bahasa keseharian menggunakan dialek Manado.
2. Merekam suara dengan aplikasi Recording audio bawaan dari handphone iphone.
3. Pengambilan data audio melalui hasil rekaman suara narasumber menggunakan kalimat seperti tabel di ini :

Tabel 3.1 Data kalimat

No	Kalimat
1	Ngana deng sapa
2	Kita da deng tamang-tamang ini
3	Ngoni mo kamana
4	Torang mo ka rumah
5	Bae-bae disana
6	Ngana so makang
7	Co cari akang
8	Kita da makang ikang
9	Kypa ngana ini
10	Jadi torang mo kaluar
11	Jemput pa kita
12	Kita pe puru saki
13	Ngana jang ba nakal
14	For apa le ini

4. Pengambilan data audio dibagi menjadi 3 narasumber :
  - a. Orang Manado.
  - b. Orang luar Manado yang menggunakan bahasa manado.
  - c. Bukan orang Manado.

#### 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 epoch

Tabel 4.1 Tabel Pengujian Epoch

Pengujian	Epoch	MiniBatchSize	Akurasi
Pengujian Pertama	10	27	59 %
Pengujian Kedua	20	27	64 %
Pengujian Ketiga	30	27	67 %
Pengujian Keempat	40	27	68 %
Pengujian Kelima	50	27	70 %
Pengujian Keenam	60	27	78 %
Pengujian Ketujuh	70	27	79 %
Pengujian Kedelapan	80	27	81 %
Pengujian Kesembilan	90	27	83 %
Pengujian Kesepuluh	100	27	85 %
Pengujian Kesebelas	150	27	88 %

Dari Nilai akurasi tertinggi yaitu 88 % dengan nilai epoch 150 yang didapatkan dari hasil pengujian akan diuji kembali dengan menggunakan parameter MiniBatchSize yang berbeda untuk melihat apakah ada perubahan nilai pada akurasi.

##### 4.2 Pengujian MiniBatchSize

Tabel 4.2 Pengujian MiniBatchSize

Pengujian	Epoch	MiniBatchSize	Akurasi
Pengujian Pertama	150	10	87 %
Pengujian Kedua	150	20	89 %
Pengujian Ketiga	150	30	87 %
Pengujian Keempat	150	40	85 %
Pengujian Kelima	150	50	79 %
Pengujian Keenam	150	60	85 %
Pengujian Ketujuh	150	70	83 %
Pengujian Kedelapan	150	80	81 %
Pengujian Kesembilan	150	90	80 %

Pengujian Kesepuluh	150	100	79 %
Pengujian Kesebelas	150	150	69 %

Dari nilai akurasi tertinggi yaitu 89 % dengan nilai MiniBatchSize 20 yang didapatkan dari hasil pengujian akan diuji kembali dengan menggunakan nilai epoch yang berbeda untuk melihat apakah ada perubahan nilai pada akurasi.

### 4.3 Pengujian sistem

Tabel 4.3 Pengujian sistem

Pengujian	Epoch	MiniBatchSize	Akurasi
Pengujian 1	150	20	91 %
Pengujian 2	150	20	85 %
Pengujian 3	150	20	89 %
Pengujian 4	150	20	86 %
Pengujian 5	150	20	84 %
Pengujian 6	150	20	91 %
Pengujian 7	150	20	91 %
Pengujian 8	150	20	88 %
Pengujian 9	150	20	84 %
Pengujian 10	150	20	89 %
RATA-RATA AKURASI			87.8 %

Pada tabel 4.3 pengujian sistem ini diambil nilai epoch dan MiniBatchSize dengan hasil akurasi terbaik untuk dilakukan pengujian sebanyak sepuluh kali dengan mendapatkan rata-rata akurasi dengan nilai 87.8 %.

## 5. Kesimpulan

Berdasarkan hasil pengujian dan analisa yang telah dilakukan dari Tugas Akhir ini, maka dapat ditarik kesimpulan sebagai berikut :

1. Semakin banyak data training maka semakin baik hasil akurasi yang dihasilkan.
2. Semakin banyak data training maka akan semakin baik hasil akurasi yang dihasilkan.
3. Ukuran epoch dan BatchSize sangat mempengaruhi hasil akurasi.

**DAFTAR PUSTAKA**

- [1] Haykin, Simon. 1994. *Neural Networks*, Prentice Hall.
- [2] Hermawan, Arief. 2006. *Jaringan Saraf Tiruan, Teori dan Aplikasi*. Andi Publisher.
- [3] Diyah Puspitaningrum, Diyah. 2006. *Pengantar Jaringan Saraf Tiruan*, Andi Publisher.
- [4] Linggard, R. Myers, D.J. Nightingale, C. 1992. *Neural Networks for vision, speech, and natural language*.
- [5] Laurene V, Fausett. 1994. *Fundamental of neural networks*.
- [6] Haykin, Simon. 2008. *Neural Networks and Learning Machines 3rd Edition*.
- [7] Diamantaras, K. I. Kung, S. Y. 1996. *Principal Component Neural Networks Theory And Applications*.
- [8] Fu, LiMin. 1994. *Neural Networks in Computer Intelligence*.
- [9] Wikipedia, "Pengenalan Ucapan", Wikipedia, 2017. [Online]. Available: [https://id.wikipedia.org/wiki/Pengenalan\\_ucapan](https://id.wikipedia.org/wiki/Pengenalan_ucapan). [Diakses 04 Oktober 2017].
- [10] Bina Nusantara, "Recurrent Neural Network (RNN) dan Gated Recurrent Unit (GRU)", Aryo Pradipta Gema, Derwin Suhartono, S.Kom., M.T.I., 2017. [Online]. Available: <https://socs.binus.ac.id/2017/02/13/rnn-dan-gru/>. [Diakses 28 September 2017].
- [11] Hasim Sak, Andrew Senior, Fracoise Beaufays. *Long Short-Term Memory Recurrent Neural Network Architecture for Large Scale Acoustic Modeling*
- [12] LeCun, Yan. Yosua Bengio. Geoffrey Hinton. 2015. *Deep Learning*. Jepang