Identifikasi Dialek Suku Bangsa Menggunakan Metode *Mel-Frequncy Cepstral Coefficient* Dan *Zero Crossing Rate* Dengan *Deep Neural Network Classifier*

1st Gesha Faithul Ajrin
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
geshafaithul@telkomuniversity.ac.id

2nd Rita Magdalena
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
ritamagdalena@telkomuniversity.ac.id

3rd Bambang Hidayat
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
bhidayat@telkomuniversity.ac.id

Abstrak*—Indonesia memiliki berbagai macam bahasa. Bahasa merupakan cara manusia untuk bersosialisasi dan dijadikan sebagai penanda stratifikasi sosial dalam menjalin hubungan di suatu masyarakat. Penggunaan bahasa oleh masyarakat terjadi karena faktor lingkungan, suku dan budaya di suatu daerah tertentu. Dengan adanya perbedaan tersebut, dialek masing-masing daerah memiliki karakteristik yang unik. Tugas akhir ini dirancang menggunakan MFCC (Mel Frequency Cepstral Coefficient) dan ZCR (Zero Crossing Rate) sebagai metode untuk melakukan ekstraksi ciri. Suara dari beberapa suku dengan dialek Batak, Serawai, dan Makassar akan direkam oleh Handphone dan diproses menggunakan aplikasi Matlab. Suara yang telah diekstraksi ciri akan dibuat kelas dan diklasifikasikan berdasarkan suku. Klasifikasi tersebut dilakukan menggunakan metode DNN (Deep Neural Network). Dalam penelitian kali ini terdapat dua skenario pengujian yaitu pengujian validasi dan pengujian akurasi. Data yang digunakan adalah data suara primer yang telah direcord oleh narasumber sebanyak 300 (75%) data latih dan 75 (25%) data uji. Pengujian terbaik dari penelitian ini adalah dengan nilai parameter Hidden size 300, L2 Weight Regulazitation 0.001, Sparsity Regulazitation 2, dan Epoch 100 yang menghasilkan tingkat validasi 100% dan akurasi 86%.

Kata kunci*—dialek, mel frequency cepstral coefficient (mfcc), zero crossing rate (zcr), deep neural network.

I. PENDAHULUAN

Dialek merupakan variasi bahasa yang membedakan suatu daerah, kelompok dan kurun waktu tertentu. Dialek terbagi menjadi dua jenis yaitu regional dan sosial. Dialek regional adalah dialek yang berdasarkan geografi yang digunakan oleh daerah tertentu sedangkan dialek sosial adalah dialek yang dipakai oleh kelompok sosial tertentu lingkungan masyarakat[1].

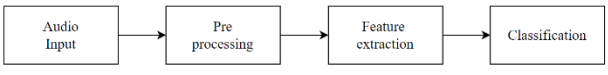
Pengenalan suara (speech recognition) adalah pengolahan sinyal (signal processing) yang merupakan bagian dari ilmu komunikasi. Speech recognition ini memanfaatkan sinyal suara masukan dan menjadikan ciri dari suara sebagai informasi untuk diolah kemudian dilakukan identifikasi terhadap sinyal suara masukan tersebut.

Tugas akhir ini dilakukan menggunakan metode Deep Neural Network (DNN) sebagai classifier dengan ekstraksi ciri menggunakan Mel-Frequency Cepstral Coefficient (MFCC) dan Zero Crossing Rate (ZCR). Penelitian ini dilakukan untuk mengklasifikasikan dialek Batak, Serawai, dan Makassar.

II. KAJIAN TEORI

A. *Speech Recognition*

Speech recognition adalah salah satu kemampuan mesin dalam mengidentifikasi sinyal suara berupa kata yang kemudian diolah kedalam format yang berbeda agar dapat dibaca oleh mesin. Istilah *speech recognition* sering disebut sebagai *Automatic Speech Recognition* (ASR) [4]. Pemanfaatan speech recognition memungkinkan perangkat dapat menganalisis kata yang kemudian didigitalisasi dengan cara pencocokan sinyal digital yang akan menghasilkan akurasi pengenalan suara yang tinggi [2]. Salah satu implementasi speech recognition adalah perintah suara dalam menjalankan aplikasi komputer [5]. Berikut merupakan Gambar 2.1 yang menunjukkan proses dari *speech recognition*:



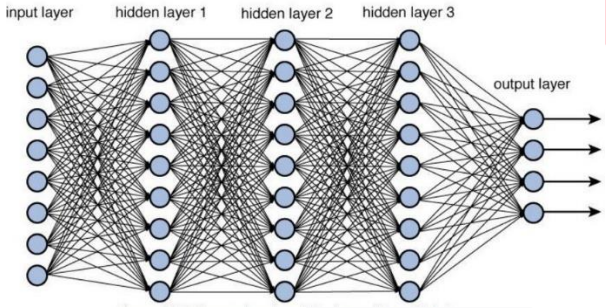
GAMBAR 2. 1
SPEECH RECOGNITION

Sinyal yang ditangkap akan diproses ditahap pertama yaitu dilakukan *pre-processing* dimana pada tahapan ini sinyal tersebut akan disiapkan dan diolah untuk memudahkan tahapan selanjutnya. Kemudian hasil data tersebut diekstraksi dan dipisahkan sesuai dengan cirinya masing-masing. Pada penelitian ini akan dilakukan ekstraksi menggunakan MFCC dan ZCR. Hasil dari ekstraksi ini akan dibandingkan melalui

database yang dimiliki. Kemudian data tersebut akan diklasifikasi berdasarkan jenisnya.

B. *Deep Neural Network*

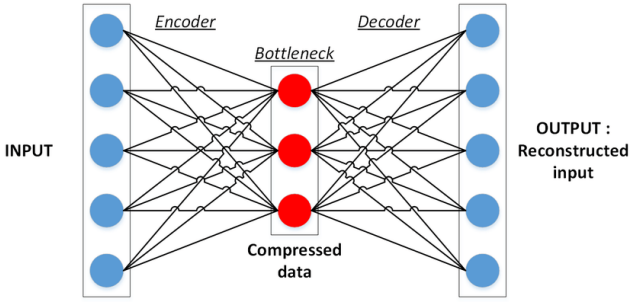
DNN adalah salah satu algoritma *deep learning* yang memiliki cara kerja seperti otak manusia, sehingga algoritma ini dapat mengenali suatu pola dan berfungsi untuk mengelompokkan data. DNN terinspirasi dari jaringan saraf yang bekerja untuk pengambilan keputusan. Metode ini memungkinkan data yang kompleks akan menjadi lebih mudah untuk dimodelkan [11]. DNN memiliki kemampuan dalam mengolah data dengan cara *Supervised Training*. Kemampuan ini memungkinkan untuk memprediksi input yang diberikan terhadap target. Untuk dapat melakukan proses tersebut data akan dilatih dan dijadikan acuan sebagai *database*. Metode ini memiliki kelebihan dalam *Speech Recognition*, yaitu lebih cepat dalam memahami berbagai macam dialek suku, arsitektur jaringan yang lebih baik dan dapat mengoptimalkan berbagai macam parameter [3]. Berikut Gambar 2.2 Arsitektur Deep Neural Network Layer.



GAMBAR 2. 2
DEEP NEURAL NETWORK

C. *Autoencoder*

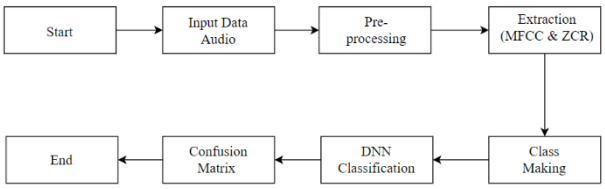
Autoencoder adalah metode yang terdapat pada *hidden layer* dalam arsitektur DNN. Metode ini memungkinkan untuk menghasilkan *output* yang sama dengan *input*. *Autoencoder* berfungsi untuk merekonstruksi satu set data *input* yang kemudian dilatih agar dapat merepresentasikan (*encoding*) data input tersebut. Untuk membangun *autoencoder* dapat dilakukan dengan pembuatan jaringan *feed forward* yang dimodifikasi beberapa pengaturannya. Pengaturan yang dilakukan yaitu pengaturan ukuran layer tersembunyi pada *autoencoder*. Akhir-akhir ini, *Autoencoder* banyak digunakan untuk belajar model data *generative* [12]. Berikut Gambar 2.3 struktur *Autoencoder*.



GAMBAR 2. 3
AUTOENCODER

III. METODE

A. Desain Sistem



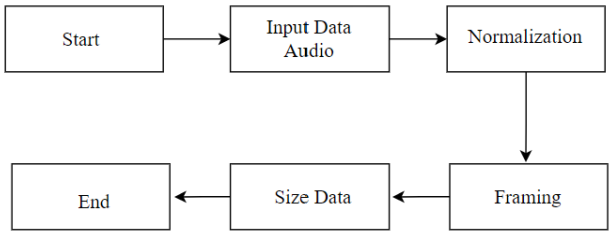
GAMBAR 3. 1
DESAIN SISTEM

1. *Pengambilan Data Audio*

Pengambilan data audio merupakan tahap awal pada perancangan sistem. Pada proses pengambilan data akan dilakukan perekaman terhadap 15 orang dari seluruh suku menggunakan kalimat yang sama, tetapi pengucapannya menggunakan bahasa daerah masing-masing. Setiap suku terdapat lima orang. Setiap orang akan mengucapkan lima kalimat dan diulang sebanyak 5 kali, sehingga total data yang dihasilkan dari setiap orang sebanyak 25 atau 125 setiap suku. Data tersebut jika dijumlahkan adalah 375. Pada penelitian ini data training sebanyak 300 (75%), data testing sebanyak 75 (25%).

2. *Pre-processing*

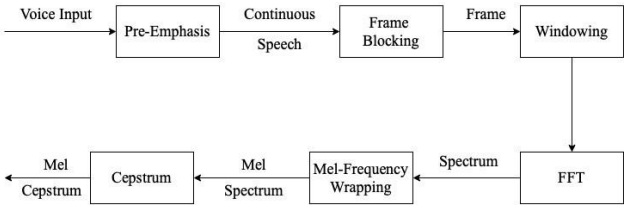
Pada tahap ini terdapat tiga proses dari pre-processing, yaitu:



GAMBAR 3. 2
PRE-PROCESSING

3. *Extraction* (MFCC)

Pada tahap ini terdapat dua proses yaitu pembuatan database sebagai template dan ektraksi ciri masukan data uji. Ekstraksi ciri suara memiliki tujuan untuk mengubah gelombang suara menjadi beberapa tipe representasi parametrik. Metode MFCC adalah salah satu metode ekstraksi ciri yang dapat merepresentasikan suara (audio) secara parametris agar dapat diproses dan dikembangkan secara lebih lanjut.

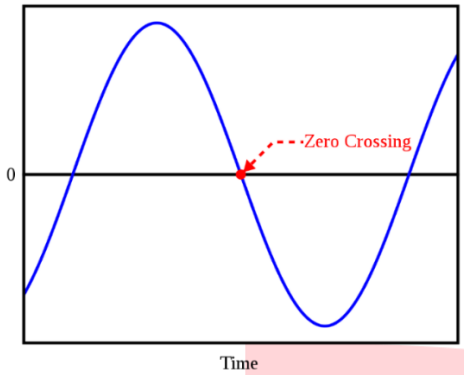


GAMBAR 3. 3
MFCC

4. *Extraction* (ZCR)

Salah satu metode yang bekerja pada domain waktu adalah ZCR. Zero crossing dapat terjadi jika sample pada

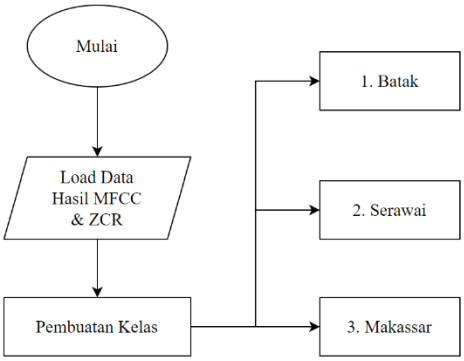
konteks pewaktu sinyal diskrit memiliki tanda aljabar dengan sample paling aktual. Berikut merupakan Gambar 3.4 ZCR:



GAMBAR 3. 4
ZCR

5. Pembuatan Kelas

Pada tahap ini dilakukan pembuatan kelas untuk setiap nilai klasifikasi. Terdapat tiga kelas yaitu suku Batak, Serawai dan Makassar. Berikut Gambar 3.5 merupakan proses dari pembuatan kelas.



GAMBAR 3. 5
PEMBUATAN KELAS

6. DNN *Validation Test*

Pada tahap ini sistem mengestimasi parameter dari sinyal masukan yang telah diektraksi ciri. Model dari hasil akan disimpan dalam database sistem. Pelatihan dilakukan sebanyak sembilan kali agar mendapatkan akurasi yang maksimal. Pada pelatihan ini menggunakan tiga hidden layer. Pada layer pertama dan kedua menerapkan autoencoder untuk pelatihannya, jumlah neuron dari setiap hidden layer pada autoencoder dibuat lebih kecil dari input yang masuk, pada penelitian ini jumlah neuron yang digunakan untuk setiap hidden layer pertama dan kedua sebanyak 100 neuron sedangkan untuk hidden layer ketiga sebanyak 3 neuron.

7. DNN *Accuracy Test*

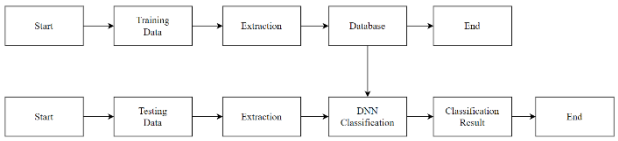
Pada tahapan DNN Test, sistem akan melakukan pengujian dengan input suara yang baru, dan akan dibandingkan terhadap database. Pada tahap ini sistem akan melakukan proses yang sama dengan training. Hasil dari accuracy test akan menampilkan confusion matrix.

8. *Confusion Matrix*

Confusion Matriks akan menampilkan tabel berupa hasil klasifikasi dari hasil pengujian. Dari tabel tersebut akan menampilkan kolom dari tiga kelas yang sudah ditentukan berdasarkan suku.

9. Fase *Training* dan *Testing*

Fase training dan testing akan dilakukan untuk menghasilkan nilai akurasi yang maksimal. Pada tahap ini akan dilakukan trial dan error agar mendapatkan nilai yang terbaik. Untuk membuat sistem speech recognition, dibutuhkan dua buah subsistem (subsistem pelatihan dan subsistem pengujian. Berikut ini Gambar 3.6 merupakan proses dari fase training dan testing:



GAMBAR 3. 6
FASE *TRAINING* DAN *TESTING*

B. Pengujian Validasi

TABEL 3. 1
PENGUJIAN VALIDASI

Pengujian	*Hidden size*	*L2 Weight Regularization*	*Sparsity Regulazitation*	*Epoch*	Hasil Klarifikasi
1	100	0.1	1	100	67.5%
2	150	0.1	1	100	74.7%
3	200	0.1	2	150	76%
4	250	0.1	2	200	81.7%
5	300	0.1	2	100	85.7%
6	300	0.01	5	200	91.3%
7	300	0.01	4	100	87%
8	300	0.001	3	150	97.7%
9	300	0.0001	2	100	100%

Berdasarkan hasil pengujian validasi nilai terbaik yaitu pada pengujian 9 dan didapatkan bahwa rata-rata pengujian terbaik dari seluruh pengujian adalah pada epoch 100.

C. Pengujian Akurasi MFCC

TABEL 3. 2
PENGUJIAN AKURASI MFCC

Pengujian	*Hidden size*	*L2 Weight Regularization*	*Sparsity Regularization*	*Epoch*	Akurasi Batak	Akurasi Serawai	Akurasi Makassar	Rata-rata Akurasi
1	300	0.0001	2	100	80%	40 %	100 %	73.3 %
2	300	0.0001	2	100	80%	60 %	80 %	73.3 %
3	300	0.0001	2	100	80%	100%	60 %	80%
4	300	0.0001	2	100	80%	100%	40 %	73.3 %
5	300	0.0001	2	100	100%	60 %	100 %	86.7 %
Rata-rata Total Akurasi					84%	72 %	76 %	77.32%

D. Pengujian Akurasi ZCR

TABEL 3. 3
PENGUJIAN AKURASI ZCR

Pengujian	*Hidden size*	*L2 Weight Regularization*	*Sparsity Regularization*	*Epoch*	Akurasi Batak	Akurasi Serawai	Akurasi Makassar	Rata-rata Akurasi
1	300	0.0001	2	100	40%	60 %	80 %	60%
2	300	0.0001	2	100	80%	40 %	80 %	66.7 %
3	300	0.0001	2	100	100 %	60 %	40 %	66.7 %
4	300	0.0001	2	100	60%	60 %	60 %	60%
5	300	0.0001	2	100	80%	60 %	100%	80%
Rata-rata Total Akurasi					72%	56 %	72 %	66.68%

IV. HASIL DAN PEMBAHASAN

Dari hasil pengujian validasi dan juga akurasi, terdapat beberapa analisis yang bisa ditemukan sebagai berikut:

A. Nilai L2 *Weight Regularization* yang lebih rendah memiliki akurasi yang lebih baik. Apabila nilai yang diberikan terhadap terlalu besar atau terlalu kecil maka akan menyebabkan *Underfitting*. *Underfitting* adalah keadaan dimana data yang di*training* dan *testing* hasilnya tidak stabil.

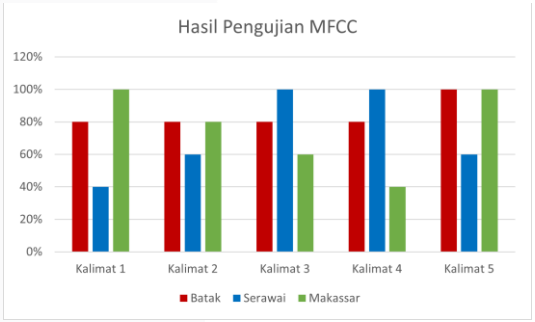
B. *Sparsity Regularization* adalah parameter untuk pembobotan dari nilai fraksi yang terlalu besar disetiap *neuron*. Pada penelitian ini, nilai ideal untuk *Sparsity Regularization* adalah dua.

C. Parameter *hiddensize* mempengaruhi tingkat akurasi. Pengujian *hidden size* juga mempengaruhi lamanya kinerja sistem, semakin besar nilai dari *hidden size* semakin lama kinerja sistem.

D. Parameter *epoch* sangat mempengaruhi tingkat akurasi. Nilai *epoch* yang semakin besar akan memiliki akurasi yang lebih tinggi karena didalamnya terdapat pembelajaran yang lebih banyak terhadap data *training* dan *testing* untuk dijadikan sebagai *database*.

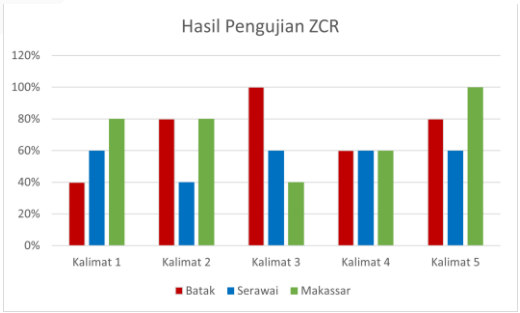
E. Pada pengujian sistem terdapat nilai parameter terbaik yaitu pada saat *hidden size* 300, L2 *Weight Regularization* 0.0001, *Epoch* 100, dan *Sparsity Regularization* 2.

F. Parameter terbaik pada saat pengujian validasi akan dijadikan sebagai acuan untuk pengujian akurasi. Pada pengujian akurasi akan dilakukan uji data suara yang baru.



GAMBAR 4. 1
HASIL PENGUJIAN MFCC

Berdasarkan grafik diatas dapat diketahui bahwa kalimat 5 memiliki rata-rata akurasi terbaik diatas 80%.



GAMBAR 4. 2
HASIL PENGUJIAN ZCR

Berdasarkan grafik diatas dapat diketahui bahwa kalimat 5 memiliki rata-rata akurasi terbaik yaitu 80%.

V. KESIMPULAN

Berdasarkan dari hasil penelitian ini, dapat disimpulkan bahwa metode *Deep Neural Network* (DNN) yang telah dibangun pada penelitian kali ini dapat mengklasifikasi dialek suku yang terdiri dari suku Batak, Serawai, dan Makassar. Dari metode ekstraksi ciri yang terdapat dalam penelitian ini, *Mel-Frequency Cepstral Coefficient* (MFCC) memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan *Zero Crossing Rate* (ZCR) yaitu 86% sedangkan pada metode ZCR memiliki tingkat akurasi terbaik yaitu 80%. Adapun parameter yang mempengaruhi sistem dalam mengklasifikasi suku adalah *Hidden size*, *L2 Weight Regularization*, *Sparsity Regularization*, dan *Epoch*.

Hasil penelitian yang telah dilakukan dapat mencapai nilai terbaik pada pengujian validasi yaitu mencapai 100% dan pengujian akurasi mencapai 86%. Penemuan yang didapatkan dari pengujian ciri suara dengan data training sebesar 75% dan data testing sebesar 25% adalah dengan menggunakan *Hidden size* 300, *L2 Weight Regularization* 0.0001, *Sparsity Regularization* 0.15, dan *Epoch* 100.

REFERENSI

[1] Kompas.com, “Dialek: Pengertian, Asal-Usul, dan Ragamnya.” Internet: www.kompas.com/skola/read/2020/01/29/080000469/dialek.html, 2020.

[2] G. Arwandani, A. B. Osmond, and R. A. Nugrahaeni, “Deep Neural Network Untuk Pengenalan Ucapan Pada Bahasa Sunda Dialek Utara,” vol. 5, no. 3, pp. 6081–6088, 2018.

[3] M. R. Adi, A. B. Osmond, and P. Anggunmeka Luhur, “Penentuan Dialek Jawa Menggunakan Metode Recurrent Neural Network Determination of Jawa Dialek Using Recurrent Neural Network Method,” e-Proceeding Eng., vol. 6, no. 2, pp. 5637–5647, 2019.

[4] Sharma, Mridusmita & Sarma, Kandarpa, “Soft-Computational Techniques and Spectro-Temporal Features for Telephonic Speech Recognition: An Overview and Review of Current State of the Art”. 10.4018/978-1-4666-9474-3.ch006.

[5] Andriana; Olly V; Riyanto S; Ganjar T; Zulkarnain, “Speech Recognition Sebagai Fungsi Mouse Untuk Membantu Pengguna Komputer Dengan Keterbatasan Khusus,” Semin. Nas. Sains dan Teknol. 2016, no. November, pp. 1–7, 2016, [Online]. Available: https://jurnal.umj.ac.id/index.php/semnastek/article/download/778/706.

[6] C. Wu, P. Karanasou, M. J. F. Gales, and K. C. Sim, “Stimulated deep neural network for speech recognition,” Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH, vol. 08-12-September-2016, pp. 400–404, 2016, doi: 10.21437/Interspeech.2016-580.

[7] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” 2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc., no. Ml, pp. 1–14, 2014.