

BAB II

TINJAUAN PUSTAKA

A. Hasil Penelitian Terdahulu

Pada penelitian yang dilakukan oleh Ayu Mika Sherila, Bambang Hidayat, Inung Wijayanto yang berjudul *Sistem Identifikasi Dialek Suku Bangsa Di Indonesia Menggunakan Metode Mel-Frequency Cepstral Coefficient dan Self Organizing Map* (2012) dalam penelitiannya menggunakan metode *Mel-Frequency Cepstral Coefficient* dan metode klasifikasi *Self Organizing Map*. Hasil akurasi yang didapatkan pada proses *non real time* yaitu 92,5% dan pada proses *real time* akurasi yang didapatkan sebesar 80% (Sherila, et al., 2012).

Pada penelitian yang dilakukan oleh Erwin Lopian, Andrew Briand Osmond, Randy Erfa Saputra dengan judul *Recurrent Neural Network Untuk Pengenalan Ucapan Pada Dialek Manado* (2018) menggunakan metode *Recurrent Neural Network* untuk mengklasifikasikan jenis dialek yang ada di daerah Manado. Nilai akurasi yang didapat adalah sebesar 87,8 % (Lopian, et al., 2018).

Charalambos Themistocleous (2019) dalam penelitiannya yang berjudul *Dialect Classification From a Single Sonorant Sound Using Deep Neural Networks* menggunakan metode *Deep Neural Networks*. Sistem ini membedakan dua dialek yang menggunakan kombinasi momen spektral, informasi temporal,

dan informasi frekuensi formant dengan akurasi klasifikasi 81% (Themistocleous, 2019).

Penelitian yang dilakukan oleh Kristiawan Nugroho (2019) yang berjudul *Javanese Gender Speech Recognition Based on Machine Learning Using Random Forest and Neural Network* menggunakan metode *Mel-Frequency Cepstral Coefficient* dan membandingkan tingkat akurasi antara *Random Forest* dan *Neural Network*. Sistem ini digunakan untuk mengklasifikasikan suara dari gender masyarakat Jawa dengan tingkat akurasi 91,3% menggunakan *Random Forest* dan 92,2% menggunakan Jaringan Saraf Tiruan (Nugroho, 2019).

Penelitian yang dilakukan oleh Mada Sanjaya W.S dan Zabidin Salleh dengan judul *Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuro-Fuzzy Inference System (ANFIS) Sebagai Kontrol Lampu Otomatis* (2014) menggunakan metode MFCC dan ANFIS untuk mengklasifikasi pola suara untuk mengendalikan lampu otomatis. Hasil pengujian sistem tersebut menunjukkan tingkat keberhasilan untuk data suara terlatih sebesar 98,57% dan data tak terlatih sebesar 95,90% (Sanjaya & Salleh, 2014).

Penelitian yang dilakukan oleh Mada Sanjaya dan Dyah Anggraeni (2016) dengan judul *Sistem Kontrol Robot Arm 5 DOF Berbasis Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuro-Fuzzy Inference System (ANFIS)* menggunakan metode MFCC dan ANFIS untuk mengklasifikasikan pola suara untuk mengendalikan robot *arm*

5 DOF. Sistem tersebut memiliki tingkat keberhasilan sebesar 87,77% untuk data terlatih dan 78,53% untuk data tak terlatih (Sanjaya & Anggraeni, 2016).

B. Landasan Teori

1. Speech Recognition

Speech Recognition atau yang kita kenal sebagai *Automatic Speech Recognition* (ASR) adalah kemampuan mesin atau program untuk mengidentifikasi kata dan frasa dalam bahasa lisan dan mengkonversikannya ke format yang dapat dibaca oleh mesin (Adi, et al., 2019). Masukan sistem yang berupa kata atau kalimat yang diucapkan manusia akan diidentifikasi untuk menghasilkan teks yang sesuai. Sinyal ucapan pertama kali akan dilewatkan pada bagian penganalisis ucapan untuk mendapatkan besaran atau ciri yang berbeda.

Teknologi ini memungkinkan suatu perangkat dapat mengenali kata-kata dengan cara menganalisis spesifikasi kata-kata yang disebutkan, lalu mendigitalisasi kata dan mencocokkan dengan sinyal digital tersebut dengan pola tertentu 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 sampel yang akan dijadikan sebagai data latih (Xiong, et al., 2014).

Adapun dua modul utama yang dibutuhkan dalam perancangan *Speech Recognition*, yaitu :

a. Ekstraksi 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.

b. Pencocokan Ciri (*Pattern Matching*)

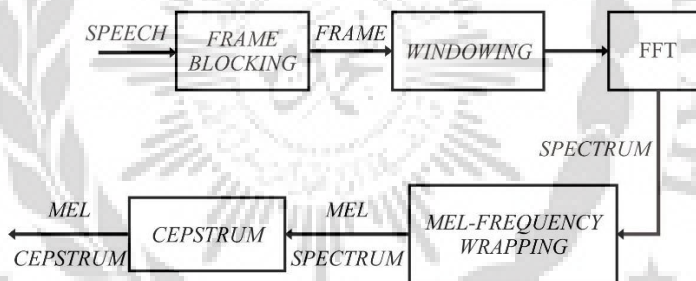
Dalam pencocokan ciri akan di lakukan perbandingan atau mencocokkan data dari sinyal masukan dengan data latih yang telah ada di dalam *database*. Hasil dari pencocokan ciri ini akan menjadi keluaran sistem.

2. Ekstraksi Ciri

Tujuan dari ekstraksi ciri adalah untuk mengkonversi gelombang suara menggunakan alat Pemrosesan Sinyal Digital (PSD) menjadi set tertentu 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. Metode yang paling sering digunakan untuk mengkarakterisasi sinyal suara dalam *Speech Recognition* yaitu *Mel Frequency Cepstral Coefficients* (MFCC).

3. Mel-Frequency Cepstral Coefficient (MFCC)

Mel-Frequency Cepstral Coefficients (MFCC) adalah metode ekstraksi fitur yang mendekati sistem pendengaran manusia. MFCC sering digunakan untuk ekstraksi fitur dalam pengolahan ucapan yang bertujuan untuk mengekstrak karakteristik penting dari sinyal ucapan yang unik dan untuk membedakan antara serangkaian kata yang berbeda. MFCC dianggap sebagai metode standar untuk ekstraksi fitur dalam pengenalan suara dan sering digunakan untuk pemrosesan suara (Helmiyah, et al., 2018). Proses ekstraksi fiturnya dapat dilihat pada Gambar 2.1.



Gambar 2.1 Diagram Blok Metode MFCC (Sanjaya & Salleh, 2014)

a. Pre-Emphasis

Proses ini bertujuan untuk mem-filter sinyal ucapan untuk mengurangi nilai frekuensi sinyal suara masukan dan dapat mengurangi *noise* pada masukan, sehingga hanya sinyal bicara saja yang dapat ditangkap oleh sistem. Persamaannya dapat dilihat pada rumus nomor 1.

$$p(n) = s(n) - xs(n - 1) \quad (1)$$

dimana x adalah konstanta filter *pre-emphasis*, biasanya bernilai antara $0,9 < x < 1,0$.

b. *Framing*

Pada proses ini *input* suara dipotong menjadi *frame-frame* dengan durasi yang lebih pendek sebanyak matriks (M) yang disimpan di matriks Y dengan ukuran $M \times W$. Sinyal suara dilakukan segmentasi menjadi beberapa *frame* dengan cara tumpang tindih (*overlap*) agar tidak ada sinyal yang hilang. Proses ini terus berlanjut sampai seluruh sinyal masuk ke dalam satu *frame* atau bahkan lebih dari satu *frame*.

c. *Windowing*

Proses ini bertujuan untuk memperoleh sampel sinyal yang tepat dalam waktu interval yang sangat singkat. Proses ini menghasilkan *window* $X(t)$ dimana $t = 1,2,3,\dots,T$ yang disebut *frame*. Teknik *windowing* yang digunakan adalah teknik *Hamming Windowing*, persamaannya dapat dilihat pada rumus nomor 2.

$$w(n) = 0,54 + 0,46 \cos \frac{2\pi n}{N-1}, 0 \leq n \leq N - 1 \quad (2)$$

dimana n adalah jumlah sampel dan N adalah jumlah *frame*.

d. *Fast Fourier Transform (FFT)*

Fast Fourier Transform (FFT) adalah salah satu metode pengkonversian dari sinyal suara menjadi sinyal frekuensi. Proses ini dilakukan terhadap semua frame dari sinyal yang sudah di *windowing*. FFT adalah salah satu algoritma yang cepat untuk menerapkan *Discrete Fourier Transform* (DCT) yang beroperasi pada sinyal diskrit. Sinyal diskrit terdiri dari N sampel, persamaannya dapat dilihat pada rumus nomor 3.

$$f(n) = \sum_{k=0}^{N-1} w_k e^{-2\pi jkn/N}, 0 \leq n \leq N - 1 \quad (3)$$

dimana w adalah *windowing* n adalah jumlah sampel dan N adalah jumlah *frame*.

e. *Mel Filterbank*

Proses ini sama dengan *triangular* dari *filterbank*, yang membedakan adalah dari *range* frekuensi linier hasil FFT kemudian dikonversi ke skala *mel frequency* untuk mendapatkan batas-batas *filterbank*, rumus *mel-filterbank* dapat dilihat pada rumus nomor 4.

$$B(f) = 1.125 \times \ln\left(1 + \frac{f}{700}\right) \quad (4)$$

Tahapan ini bertujuan untuk menentukan batas atas dan batas bawah dari filter. Tahapan pertama yang dilakukan adalah menentukan batas atas dan batas bawah dari filter, jadi nilai yang berada diluar batas tidak dimasukkan kedalam filter. Kemudian bagi *range* batas atas dan

bawah sesuai dengan jumlah filter yang dibuat untuk setiap *filterbank* dalam skala mel, selanjutnya batas tersebut dikonversi ke skala frekuensi linier. Langkah terakhir adalah membuat filter *triangular* berdasarkan batas-batas yang sudah ditentukan. Hasil dari FFT pada tahap sebelumnya kemudian dikalikan dengan *mel-filterbank*.

f. *Discrete Cosine Transform*

Proses ini adalah proses perubahan konversi dari domain frekuensi ke domain waktu yang dilakukan untuk mendapatkan nilai koefisien dari hasil perkalian *mel-filterbank* yang sudah dikonversikan. Proses ini akan menghasilkan log dari perkalian DCT yang sudah diubah ke domain waktu yang menghasilkan *Mel-Frequency Cepstrum Coefficient (MFCC)*, persamaannya dapat dilihat pada rumus nomor 5.

$$C_j = \sum_{i=1}^M X_i \cos\left(\frac{j(i-1)}{2\frac{\pi}{M}}\right) \quad (5)$$

dimana $j = 1, 2, 3 \dots K$ adalah koefisien, dan M adalah jumlah filter.

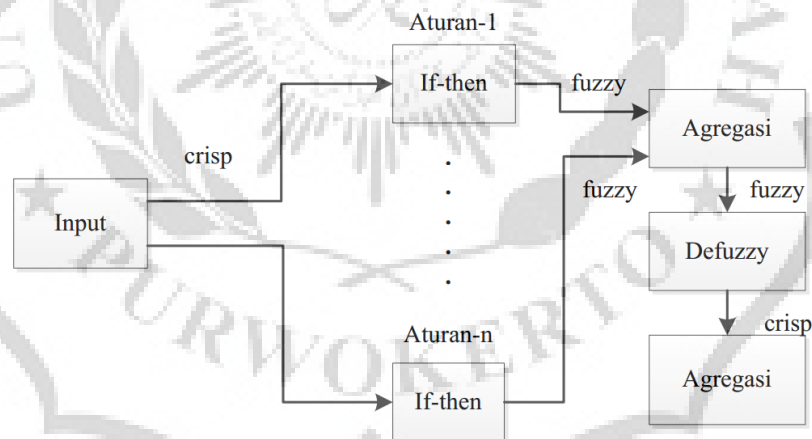
4. Logika Fuzzy

Pada tahun 1965, Zadeh memodifikasi teori himpunan dimana anggotanya memiliki derajat keanggotaan yang bernilai kontinu antara 0 sampai 1. Himpunan ini disebut dengan Himpunan kabur (*Fuzzy Set*). Logika *Fuzzy* merupakan pengembangan dari teori himpunan *fuzzy* yang diprakarsai

oleh Prof. Lofti Zadeh dari University California USA pada tahun 1965. Logika *Fuzzy* berbeda dengan logika digital biasa, dimana logika digital biasa hanya mengenal dua keadaan sedangkan logika *fuzzy* meniru cara berpikir manusia dengan menggunakan konsep sifat kesamaran suatu nilai (Pratama, 2018).

5. FIS (*Fuzzy Inference System*)

Sistem Inferensi *Fuzzy* (*Fuzzy Inference System* atau FIS) merupakan suatu kerangka komputasi yang didasarkan pada teori himpunan *fuzzy*, aturan *fuzzy* berbentuk *if – then*, dan penalaran *fuzzy*. Secara garis besar, diagram blok proses *Inference Fuzzy* terlihat pada Gambar 2.2.



Gambar 2.2 Diagram Blok Sistem *Inferensi Fuzzy* (Pratama, 2018)

FIS menerima *input crisp* yang kemudian dikirim ke basis pengetahuan berisi aturan *fuzzy* dalam bentuk *if – then*. *Fire Strength* (bobot)

akan dicari pada setiap aturan. Apabila jumlah aturan lebih dari satu, maka akan dilakukan agregasi dari semua aturan. Selanjutnya akan dilakukan *defuzzy* untuk mendapatkan nilai *crisp* sebagai keluaran sistem (Pratama, 2018). FIS memiliki dua metode, yaitu :

a. Metode Tsukamoto

Sistem inferensi *fuzzy* didasarkan pada konsep penalaran monoton, dimana nilai *crisp* pada daerah konsekuen dapat diperoleh secara langsung berdasarkan *fire strength* pada antesedennya. Syarat yang harus dipenuhi adalah himpunan *fuzzy* pada konsekuennya harus bersifat monoton (baik monoton naik maupun monoton turun).

b. Metode Takagi-Sugeno-Kang

Sistem inferensi *fuzzy* menggunakan metode Sugeno memiliki karakteristik yaitu konsekuen tidak merupakan himpunan *fuzzy*, namun merupakan suatu persamaan linear dengan variabel yang sesuai dengan variabel masukannya. Metode ini diperkenalkan oleh Takagi-Sugeno Kang pada tahun 1985. Ada dua model untuk sistem inferensi *fuzzy* dengan menggunakan metode TSK, yaitu model TSK orde-0 dan model TSK orde-

1.

1) Model TSK Orde-0

Secara umum bentuk model *Fuzzy* Sugeno orde-0 adalah : IF $(x_1 \text{ is } A_1) \circ (x_2 \text{ is } A_2) \circ (x_3 \text{ is } A_3) \circ \dots \circ (x_N \text{ is } A_N) \circ$ THEN $z = k$. Dengan A_i adalah himpunan *fuzzy* ke- i sebagai anteseden, \circ adalah operator *fuzzy* (seperti *AND* atau *OR*), dan k adalah suatu konstanta (tegas) sebagai konsekuen.

2) Model TSK Orde-1

Secara umum bentuk model *Fuzzy* Sugeno orde-1 adalah : IF $(x_1 \text{ is } A_1) \circ \dots \circ (x_N \text{ is } A_N) \circ$ THEN $z = p_1 * x_1 + \dots + p_n * x_n + q$. Dengan A_i adalah himpunan ke- i sebagai anteseden, \circ adalah operator *fuzzy* (seperti *AND* atau *OR*), p_i adalah suatu konstanta ke- i dan q juga merupakan konstanta dalam konsekuen.

6. ANFIS (*Adaptive Neuro Fuzzy Inference System*)

Model *Fuzzy* dapat digunakan sebagai pengganti dari persepsi dengan banyak lapisan. Sistem dapat dibagi menjadi dua grup, yaitu grup pertama berupa jaringan syaraf dengan bobot-bobot *fuzzy* dan fungsi aktivasi *fuzzy*, sedangkan grup kedua berupa jaringan syaraf dengan *input* yang difuzzy-kan pada lapisan pertama atau kedua, namun bobot-bobot pada jaringan syaraf tersebut tidak difuzzy-kan (Pratama, 2018).

Neuro Fuzzy adalah gabungan dari dua sistem, yaitu sistem logika *fuzzy* dan jaringan syaraf tiruan. Sistem *neuro fuzzy* berdasar pada sistem inferensi *fuzzy* yang dilatih menggunakan algoritma pembelajaran yang diturunkan dari sistem jaringan syaraf tiruan (Sanjaya & Salleh, 2014).

Adaptive Neuro-Fuzzy Inference System (ANFIS) adalah penggabungan *Fuzzy Inference System* yang digambarkan dalam arsitektur jaringan syaraf. ANFIS adalah arsitektur yang secara fungsional sama dengan *Fuzzy Rule Base* model Sugeno. Sistem inferensi *fuzzy* yang digunakan adalah sistem inferensi *fuzzy* model Tagaki-Sugeno-Kang (TSK) orde satu dengan pertimbangan kesederhanaan dan kemudahan komputasi. Dalam menentukan nilai akurasi dari sistem menggunakan perhitungan dengan rumus nomer 6.

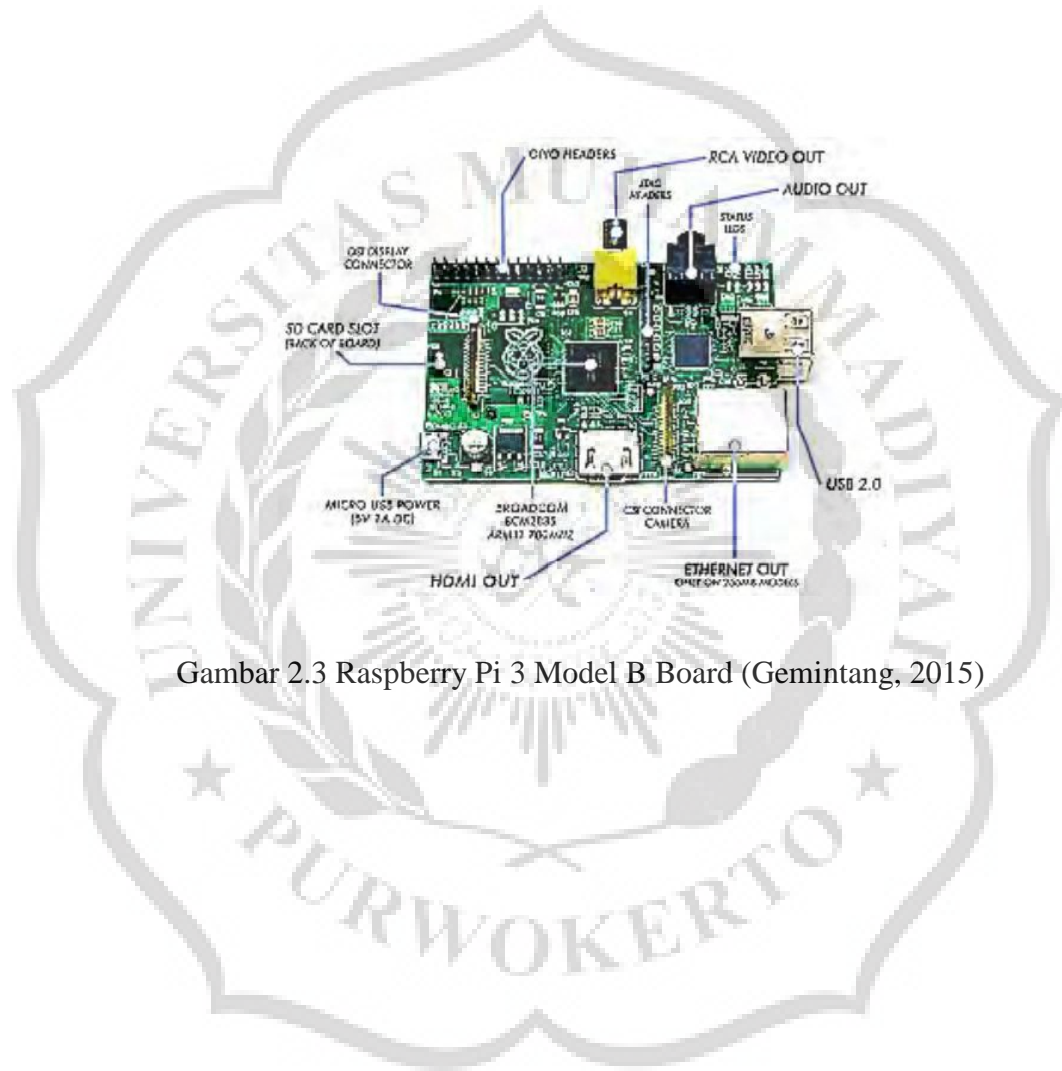
$$\text{tingkat keakuratan} = \frac{\text{jumlah data benar}}{\text{jumlah data keseluruhan}} \times 100\% \quad (6)$$

7. Raspberry Pi

Raspberry Pi® (juga dikenal sebagai RasPi) adalah sebuah SBC (*Single Board Computer*) seukuran kartu kredit. Raspberry Pi® menggunakan *System on a Chip* (SoC) dari *Broadcom* BCM2835 yang sudah termasuk prosesor ARM1176JZF-S 700 MHz, GPU VideoCore IV dan RAM sebesar 256 MB (untuk Rev. B), bentuk raspberry dapat dilihat pada Gambar 2.3.

Berbeda dengan komputer pada umumnya, mini komputer ini menggunakan SD Card untuk proses *booting* dan penyimpanan pada proses yang lama. Perangkat ini menggunakan *Operating System* (OS) yang berbeda

dengan komputer pada umumnya. OS yang digunakan berbasis Linux dengan berbagai jenis, beberapa diantaranya yang sudah sering dipakai di masyarakat umum adalah Raspbian dan Fedora (Gemintang, 2015).



Gambar 2.3 Raspberry Pi 3 Model B Board (Gemintang, 2015)