

APLIKASI *SPEECH-TO-TEXT* DENGAN METODE *MEL FREQUENCY CEPSTRAL* *COEFFICIENT (MFCC)* DAN *HIDDEN MARKOV* *MODEL (HMM)* DALAM PENCARIAN KODE ICD-10

Meri Susanti¹, Boko Susilo², Desi Andreswari³

^{1,2,3}Program Studi Informatika, Fakultas Teknik, Universitas Bengkulu.
Jl. WR. Supratman Kandang Limun Bengkulu 38371A INDONESIA
(telp: 0736-341022; fax: 0736-341022)

¹meri.susanti23@yahoo.com

²bokosusilo@unib.ac.id

³desi.andreswari@unib.ac.id

Abstrak: Suara/ucapan adalah cara kita sebagai manusia untuk berkomunikasi dan mengekspresikan diri. Proses komunikasi tidak hanya terjadi antar manusia saja. Proses komunikasi juga terjadi antara manusia dan komputer. Pada penelitian ini akan dibangun sistem pencarian kode dari diagnosis penyakit dengan menggunakan masukan suara dan luaran teks. Tujuan dari penelitian ini adalah untuk merancang dan membangun aplikasi *Speech to text* kode ICD-10 dengan metode MFCC (*Mel Frequency Cepstral Coefficient*) dan HMM (*Hidden Markov Model*) yang mampu memberikan informasi mengenai kode ICD-10 dari nama diagnosis penyakit dengan pencarian menggunakan suara. Berdasarkan penelitian dan pengujian sistem Aplikasi *Speech to text* Kode ICD-10 mampu memberikan informasi mengenai kode diagnosis dari suatu penyakit. Hasil kecocokan data masukan dan luaran dari pengujian yang telah dilakukan menggunakan parameter jumlah data set 3, filter bank 20, iterasi 2 dan state 3 mendapatkan nilai persentase 100%.

Kata kunci : *Speech to text*, *Mel Frequency Cepstral Coefficient (MFCC)*, *Hidden Markov Model HMM*), Kode ICD-10

Abstract: Voice / speech is our way as human beings to communicate and express themselves. Communication process does not just happen between humans only. Communication processes also occur between humans and computers. In this research will be built a code search system of text Application Code ICD-10 able to provide disease diagnosis by using voice input and text output. The purpose of this research is to design and build ICD-10 speech-to-text application with MFCC (*Mel Frequency Cepstral Coefficient*) method and HMM (*Hidden Markov Model*) that can provide information about ICD-10 code from disease diagnosis name with search using sound. Based on the research and testing system *Speech to text* Application Code ICD-10 able to provide information about the diagnosis code of a disease. The result of the matching of input and output data from the test that has been done using

parameters of data set number 3, filter bank 20, iteration 2 and state 3 get 100% percentage value.

Keywords: Speech-to-Text, Mel Frequency Cepstral Coefficient (MFCC), Hidden Markov Model (HMM), ICD-10 Code

I. PENDAHULUAN

Suara/ucapan adalah cara kita sebagai manusia untuk berkomunikasi dan mengekspresikan diri. Namun, seiring dengan perkembangan teknologi, proses komunikasi tersebut tidak hanya terjadi antar manusia saja. Proses komunikasi juga terjadi antara manusia dan perangkat-perangkat seperti komputer dan yang lainnya. Komunikasi yang dilakukan antara manusia dan komputer sering disebut sebagai Interaksi Manusia dan Komputer (IMK). Dalam penelitian ini memanfaatkan konsep Interaksi Manusia dan Komputer (IMK) untuk melakukan pencarian kode diagnosis penyakit yaitu Aplikasi *Speech-to-Text* Kode ICD-10.

ICD-10 merupakan singkatan dari *International Classifical of Diseases and Health Problems* yang merupakan buku pedoman revisi kesepuluh klasifikasi penyakit yang digunakan secara Internasional yang disusun berdasarkan sistem kategori dan dikelompokkan dalam satuan penyakit menurut kriteria yang telah disepakati pakar Internasional. ICD-10 sebagai alat diagnostik standar untuk epidemiologi, manajemen kesehatan dan tujuan klinis [1].

Aplikasi *Speech-to-Text* Kode ICD-10 merupakan suatu sistem yang yang dapat melakukan pencarian terhadap kode dari diagnosis penyakit yang disebut dengan kode ICD-10 dengan menggunakan masukan suara dan

luarannya berupa *text* yaitu nama diagnosis penyakit dan kode ICD-10.

Dalam aplikasi ini menerapkan metode *Speech Recognition* yang dikenal sebagai *Automatic Speech Recognition* (ASR), atau komputer pengenalan suara merupakan proses mengonversikan sinyal suara ke urutan kata-kata, melalui sebuah algoritma diimplementasikan sebagai komputer program[2].

Algoritma yang akan diterapkan pada sistem adalah algoritma MFCC (*Mel Frequency Cepstral Coefficient*) pada ekstraksi ciri (*feature extraction*) dan HMM (*Hidden Markov Model*) pada pencocokan pola.

Algoritma MFCC digunakan untuk melakukan *feature extraction*, sebuah proses yang mengonversikan sinyal suara menjadi beberapa parameter. Keunggulan dari metode ini adalah mampu menangkap karakteristik suara yang sangat penting bagi pengenalan suara atau dengan kata lain mampu menangkap informasi-informasi penting yang terkandung dalam sinyal suara, menghasilkan data seminimal mungkin tanpa menghilangkan informasi-informasi penting yang ada dan mereplikasi organ pendengaran manusia dalam melakukan persepsi sinyal suara.

Sedangkan, algoritma HMM adalah algoritma yang paling efisien dan paling akurat dalam pengenalan suara. Pengenalan suara berbasis HMM menggunakan konsep bahwa ucapan terdiri dari urutan suara dasar. Kata apapun di dalam kamus terdiri dari urutan suara dasar. Masing-masing suara dasar ini mempunyai model statistik. Oleh karena itu, sebuah kata dapat diekspresikan sebagai urutan model statistik. Suara dijadikan sampel oleh sistem pengenalan suara untuk membentuk sebuah urutan vektor fitur suara (parameter numerik). Urutan vektor fitur suara ini

adalah urutan observasi. Kemudian pengenalan menentukan secara probabilistik model yang mana yang kemungkinan besar menghasilkan vektor suara. Mengidentifikasi urutan model menghasilkan identifikasi kata yang diucapkan [3].

Maka dari itu, skripsi ini berjudul “Aplikasi *Speech-to-Text* dengan Metode MFCC (*Mel Frequency Cepstral Coefficient*) dan HMM (*Hidden Markov Model*) dalam Pencarian Kode ICD-10” yang diharapkan mampu memberikan informasi mengenai kode ICD-10 dari nama diagnosis penyakit dengan pencarian menggunakan suara.

II. LANDASAN TEORI

A. Pengertian ICD-10

ICD-10 merupakan singkatan dari *International Classifical of Diseases and Health Problems* yang merupakan buku pedoman revisi kesepuluh klasifikasi penyakit yang digunakan secara Internasional yang disusun berdasarkan sistem kategori dan dikelompokkan dalam satuan penyakit menurut kriteria yang telah disepakati pakar internasional [1].

B. Suara (Wicara)

Suara khususnya wicara merupakan cara yang natural bahkan paling penting dalam melakukan proses komunikasi. Dalam kehidupan sehari-hari, manusia melakukan berbagai jenis komunikasi dengan sesama manusia, misalnya: *body language*, berbicara (*speech*) dan lain-lain. Di antara banyak komunikasi yang dilakukan oleh manusia, berbicara (*speech*) memberikan paling banyak informasi penting dan paling efektif dalam berkomunikasi. Terdapat dua hal yang menentukan kualitas suara, yaitu frekuensi dan amplitudo. Frekuensi adalah jumlah getaran yang

dihasilkan dalam satuan waktu (detik). Rentan frekuensi suara yang dapat didengar manusia berkisar diantara 20 Hz - 20.000 Hz [4].

C. Automatic Speech Recognition (ASR)

Automatic Speech Recognition (ASR), atau komputer pengenalan suara merupakan proses mengkonversi sinyal suara ke urutan kata-kata, melalui sebuah algoritma diimplementasikan sebagai komputer program [2].

D. Mel Frequency Cepstrum Coefficient(MFCC)

Ekstraksi fitur pada ASR (*Automatic Speech Recognition*) merupakan proses perhitungan urutan dari fitur vektor yang mampu merepresentasikan sinyal wicara yang ada secara optimal [5]. Fitur yang biasa digunakan adalah *cepstral coefficient*. MFCC merupakan metode ekstraksi fitur yang menghitung koefisien *Cepstral* yang didasarkan pada variasi dari frekuensi kritis pada telinga manusia. Filter dipetakan secara linear pada frekuensi rendah (< 1 kHz) dan logaritmik pada frekuensi tinggi (> 1kHz) untuk mendapatkan karakteristik suara yang penting. Beberapa keunggulan dari metode ini adalah [6] :

1. Mampu menangkap karakteristik suara yang sangat penting bagi pengenalan suara
2. Menghasilkan data seminimal mungkin, tanpa menghilangkan informasi-informasi penting yang terkandung di dalamnya
3. Mereplikasi organ pendengaran manusia dalam melakukan persepsi terhadap sinyal suara

E. Hidden Markov Model

HMM merupakan model stokastik dimana suatu sistem yang dimodelkan diasumsikan sebagai *Markov* proses dengan kondisi yang tidak terobservasi. Suatu HMM dapat dianggap sebagai jaringan Bayesian dinamis yang sederhana

(Simplest Dynamic Bayesian Network) [7]. Secara umum, HMM terdiri atas elemen- elemen berikut :

1. Himpunan nilai luaran observasi $O = \{O_1, O_2, \dots, O_M\}$, dimana M adalah jumlah simbol observasi.

2. Himpunan state $\Omega = \{1, 2, \dots, N\}$. Dimana N menyatakan jumlah state yang terdapat pada HMM.

3. Himpunan probabilitas transisi antar state. Diasumsikan bahwa state berikutnya tergantung pada state pada saat ini. Asumsi ini menyebabkan proses perhitungan menjadi lebih mudah dan efisien untuk dilakukan. Probabilitas transisi dapat dinyatakan dengan sebuah matriks $A = \{a_{ij}\}$, dimana a_{ij} adalah probabilitas transaksi dari state i ke state j . Sebagai contoh:

$$a_{ij} = P(S_{t-1} = i), \quad 1 \leq i, j \leq N$$

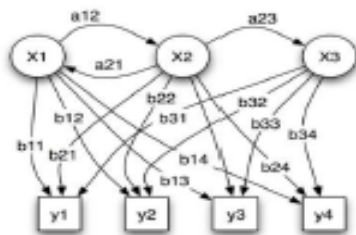
dimana S_t merupakan state pada waktu ke- t .

4. Himpunan probabilitas luaran $B = \{b_i(k)\}$ pada setiap state. Yang juga disebut probabilitas emisi, $b_i(k)$ adalah probabilitas dari simbol luaran o_k pada state i yang didefinisikan sebagai

$$b_i(k) = P(S_t = i) \text{ dimana } v_t \text{ adalah simbol observasi pada waktu ke-} t.$$

5. Himpunan state awal $\pi = \{\pi_i\}$, dimana π_i adalah probabilitas state i menjadi state awal pada urutan state HMM.

Parameter probabilistik pada Hidden Markov Model dapat dilihat pada gambar 2.1.



Gambar 2.1 Parameter Probabilistik pada Hidden Markov Model

Dimana :

x = kondisi

y = observasi yang mungkin

a = kemungkinan keadaan transisi

b = kemungkinan luaran

F. Penyelesaian Masalah Dengan HMM

Dalam penggunaannya terdapat 3 permasalahan dasar pada HMM untuk dapat melakukan pengenalan terhadap suara [8]. Pertama, masalah evaluasi. Dimana diberikan sebuah λ dari HMM dan barisan observasi $O = O_1, O_2, \dots, O_t$ dimana terdapat probabilitas obsevasi yang dihasilkan oleh model $p\{O \mid \lambda\}$. Kedua, masalah decoding diberikan sebuah model λ dan barisan observasi $O = O_1, O_2, \dots, O_t$ dimana kemiripan maksimal barisan state di

model yang menghasilkan observasi. Ketiga adalah masalah pembelajaran dimana diberikan model λ dan barisan pengamatan $O = O_1, O_2, \dots, O_t$ dimana kita harus menyesuaikan paramter $\lambda = (A, B, \pi)$ untuk memaksimalkan $p\{O \mid \lambda\}$.

III. METODOLOGI

Dalam penelitian ini menggunakan data nama diagnosis penyakit yang direkam dari 4 orang karyawan Rumah Sakit DKT Bengkulu yang diulang sebanyak 3 kali dengan nada datar, rendah dan tinggi dengan total keseluruhan data suara pelatihan sebanyak 372 file yang berekstensi .wav. Selain itu juga penulis menggunakan literatur-literatur sebagai berikut dalam penelitian ini :

- 1) Jurnal dan Skripsi

Jurnal dan Skripsi yang digunakan berkaitan dengan kajian mengenai metode dan algoritma yang digunakan dalam penelitian.

- 2) Buku Referensi

Referensi berupa buku yang digunakan berkaitan dengan kajian mengenai metode dan prosedur penelitian, teori dan aplikasi pengolahan

citra pemrograman MATLAB untuk pengenalan suara, dan buku lain yang terkait dalam penelitian.

3) Artikel

Artikel yang digunakan adalah artikel-artikel yang berkaitan dengan topik penelitian yang diperoleh dari internet.

IV. ANALISIS DAN PERANCANGAN

A. Identifikasi Masalah

Suara/ucapan adalah cara manusia untuk berkomunikasi dan mengekspresikan diri. Seiring perkembangan teknologi, kebutuhan akan adanya sistem dan aplikasi yang mampu menganalisa dan mengidentifikasi suatu sinyal suara pun semakin tinggi. Pemanfaatan aplikasi ini telah berkembang sebagai sarana pembelajaran, bidang keamanan dan kesehatan. Dalam bidang kesehatan misalnya, kemampuan sistem untuk mengenali masukan suara

diagnosis penyakit dengan luaran kode ICD-10 sangat membantu perekam medis untuk menyelesaikan tugasnya dalam melakukan pencarian kode diagnosis penyakit yang disebut dengan kode ICD-10.

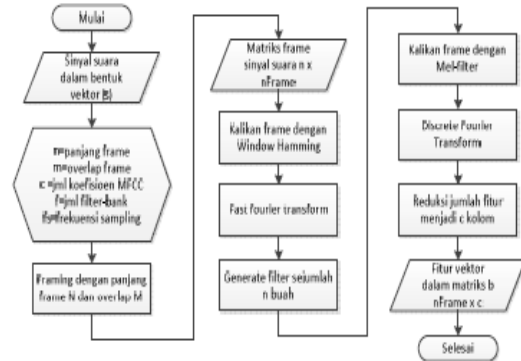
Dalam pembuatan aplikasi ini menerapkan konsep *speech-to-text* dengan algoritma *Mel Frequency Cepstral Coefficient* (MFCC) pada ekstraksi ciri (*feature extraction*) dan *Hidden Markov Model* (HMM) pada pencocokan pola. Dengan diterapkan metode tersebut diharapkan aplikasi dapat mengolah data dengan baik.

B. Alur Kerja Sistem

Analisis alur kerja sistem merupakan analisis yang ditujukan untuk menguraikan cara kerja dari suatu aplikasi yang dibangun secara runtun dimulai dari aktivitas masukan hingga menghasilkan suatu keluaran melalui proses-proses yang dijalankan aplikasi.

a. Melakukan Ekstraksi Fitur Suara

Flowchart proses ekstraksi fitur suara ditunjukkan pada gambar 4.1



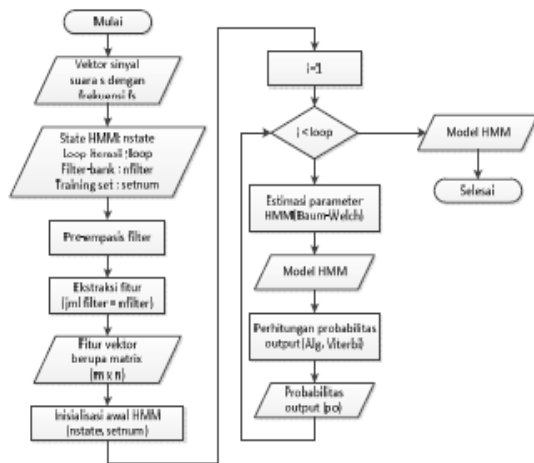
Gambar 4.1 *Flowchart* Proses Ekstraksi Fitur Suara

Tahap dalam ekstraksi fitur suara, sebagai berikut :

1. Vektor sinyal suara dibagi menjadi bagian-bagian kecil dengan panjang *frame n* dan *overlap frame m*.
2. Melakukan proses *framing* dengan panjang *frame n* dan *overlap m*.
3. Melakukan perkalian terhadap masing-masing *frame* dengan *window hamming* sepanjang *n*.
4. Melakukan transformasi *frame* ke dalam domain frekuensi.
5. Memberikan nilai jumlah *filter bank f* yang didapat dari mengkonversi frekuensi biasa ke dalam skala mel.
6. Filter bank yang terbentuk dikalikan dengan hasil FFT. Kemudian, melakukan *Discrete Fourier Transform* terhadap hasil yang didapat dari perhitungan.
7. Mereduksi jumlah fitur sesuai dengan nilai koefisien MFCC (c).
8. Hasil dari proses ini berupa fitur suara matrik dengan ordo $m \times n$.

b. Pelatihan HMM

Flowchart proses pelatihan model HMM ditunjukkan pada gambar 4.2.



Gambar 4.2 Flowchart Proses Pelatihan Model HMM

Tahap dalam pelatihan HMM, sebagai berikut :

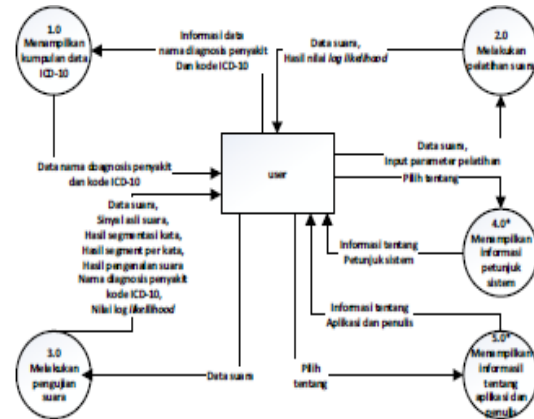
1. Inisialisasi model HMM dengan membuat sebuah *finite state automata* dengan jumlah *state nstate*.
2. Melakukan pelatihan model sebanyak *loop* kali untuk mendapatkan nilai parameter yang optimal sesuai data pelatihan yang digunakan.
3. Melakukan re-estimasi parameter dengan algoritma *Baum-Welch*.
4. Melakukan perhitungan probabilitas output dengan algoritma viterbi.
5. Hasil dari proses ini berupa model kata HMM untuk setiap kata.

C. Perancangan Sistem

Setelah mengidentifikasi masalah dan melakukan analisis berbagai kebutuhan pengguna, tahap pengembangan selanjutnya adalah melakukan perancangan aplikasi. Perancangan aplikasi bertujuan untuk memberikan gambaran mengenai rancang bangun aplikasi yang akan dibangun, sebelum nantinya diimplementasikan dengan kode-kode pemrograman. Untuk merancang aplikasi dibutuhkan adanya tahap konseptualisasi yang salah satunya dikerjakan dengan menggunakan *Data Flow Diagram* (DFD) sebagai salah satu *tool* atau model untuk merancang pengembangan perangkat lunak. Selain

konseptualisasi dengan menggunakan DFD, perancangan aplikasi juga dikerjakan untuk antarmuka pengguna.

DFD level 1 Aplikasi *Speech-to-text* Kode ICD-10 ditunjukkan pada gambar 4.3.



Gambar 4.3 DFD level 1 Aplikasi *Speech-to-text* Kode ICD-10

Pada gambar 4.3 menunjukkan DFD level 1 yang menggambarkan modul-modul yang ada dalam aplikasi yang akan dikembangkan. DFD level 1 merupakan hasil *breakdown* DFD level 0 yang sebelumnya sudah dibuat. Pada DFD level 1 memiliki 5 proses yang terjadi dalam aplikasi. Berikut ini adalah penjelasan dari proses diagram level 1.

- a. Proses 1.0 Menampilkan Kumpulan Kode ICD-10

Pada proses 1.0 pengguna memasukkan data nama diagnosis penyakit beserta kode ICD- 10. Setelah itu, sistem akan menampilkan data nama diagnosis dan kode ICD-10 ke dalam tabel.

- b. Proses 2.0 Melakukan Pelatihan Suara

Pada proses 2.0 pengguna dapat melakukan pelatihan suara. Di dalam tahap ini terjadi dua proses yaitu ekstraksi fitur suara dengan metode MFCC dan pemodelan kata menggunakan metode HMM. Proses ini akan menampilkan hasil nilai *log likelihood* dari setiap kata. Nilai

log likelihood ini berfungsi untuk tahap pengenalan suara.

c. Proses 3.0 Melakukan Pengujian Suara

Tahap pengujian ini dilakukan untuk melakukan pengenalan suara. Pengguna memasukkan data suara yang ingin dikenali yaitu nama diagnosis penyakit. Sistem akan menampilkan hasil sinyal asli dari suara, segmentasi kata dari sinyal asli dan hasil segment per kata. Suara akan dikenali dengan menghitung *log likelihood* dari kata yang diuji dengan setiap model (satu model untuk satu kata) yang ada dalam *codebook*. Nilai *log likelihood* terbesar dari model-model yang ada merupakan kata yang dikenali sistem. Setelah menemukan model yang cocok maka akan tampil luaran berupa nama diagnosis penyakit dan kode ICD-10.

d. Proses 4.0* Menampilkan informasi Petunjuk Sistem

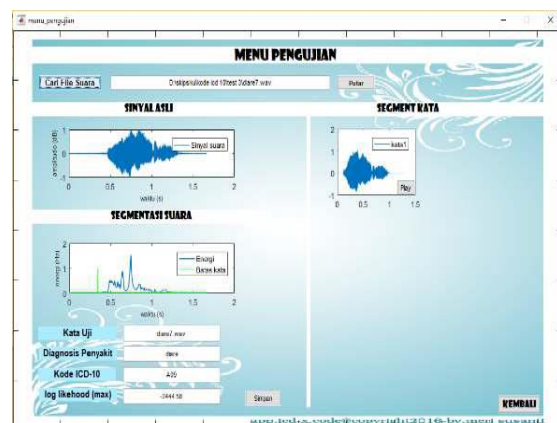
Pada proses 4.0* pengguna dapat memilih menu petunjuk. Sistem akan menampilkan informasi mengenai petunjuk penggunaan sistem. Pada proses ini terdapat tanda * yang artinya dalam proses ini tidak memiliki proses selanjutnya.

e. Proses 5.0* Menampilkan Informasi Tentang Aplikasi Dan Sistem

Pada proses 5.0* pengguna dapat memilih menu tentang. Sistem akan menampilkan informasi tentang aplikasi dan penulis.

data suara untuk melakukan pengujian. Sistem menampilkan tempat penyimpanan data testing dan memutar suara data. Selain itu, sistem juga menampilkan hasil sinyal suara asli dan segmentasi dari kata. Tahap pengenalan ini dilakukan dengan menghitung *log likelihood* dari kata yang diuji dengan setiap model (satu model untuk satu kata) yang ada dalam *codebook*. Nilai *log likelihood* terbesar dari model-model yang ada merupakan kata

yang dikenali sistem. Setelah suara dikenali yaitu suara nama diagnosis dari penyakit akan tampil luaran kode ICD-10 beserta nilai *log likelihood* terbesar dari kata.



Gambar 5.1 Halaman Menu Pengujian

2. Halaman Menu Pelatihan

Pada Gambar 5.2 menunjukkan tampilan dari menu pelatihan. Pada halaman menu pelatihan pengguna melakukan load data *training* untuk memilih data suara yang ingin dilatih. Sistem memberikan informasi mengenai tempat penyimpanan suara. Kemudian, pada halaman ini harus memasukkan parameter nilai dari jumlah *training set*, iterasi *Baum*, jumlah filter bank dan jumlah *state* HMM yang diinginkan. Setelah memasukkan parameter tersebut dapat melakukan ekstraksi suara dan pemodelan kata dengan HMM.

V. PEMBAHASAN

A. Implementasi Antar Muka

1. Halaman Menu Pengujian

Pada Gambar 5.1 menunjukkan tampilan menu pengujian pengguna dapat melakukan pencarian data testing yang merupakan data yang berisi file

1) Dataset adalah jumlah data pelatihan untuk masing-masing kata pada *dataset*. Jumlah *dataset* yang digunakan pada pengujian ini adalah 3 set data. Nilai tersebut digunakan karena jumlah data pada setiap kata pada proses pelatihan ada tiga.

2) Jumlah Filter-bank MFCC

Parameter kedua adalah jumlah filter-bank yang degenerate dalam proses ekstraksi fitur dengan nilai parameter 20 yang didapat dari perhitungan :

$$\begin{aligned} nFilter &= [\text{ceil}(4.6 * \log_{10}(fs))]; \\ &= [\text{ceil}(4.6 * \log_{10}(16000))] \\ &= [\text{ceil}(4.6 * 4.204)] \\ &= [\text{ceil}(19.3389)] \\ &= 20 \end{aligned}$$

3) Jumlah iterasi

Nilai parameter iterasi maksimum pada proses reestimasi parameter model HMM yang digunakan untuk pengujian dalam penelitian ini adalah 2. Nilai tersebut digunakan karena semakin kecil nilai iterasi maka dapat mempercepat proses pelatihan.

4) Jumlah state model HMM

Parameter jumlah *state* pada sebuah model kata yang akan diujikan antara lain adalah 3 buah *state*. Nilai tersebut digunakan karena jenis suara terdiri dari tiga yaitu datar, sedang dan tinggi.



Gambar 5.2 Halaman Menu Pelatihan

B. Hasil Pengujian Sistem

Pengujian sistem dilakukan dengan memasukkan nilai parameter yang digunakan pada penelitian ini. Himpunan parameter uji yang digunakan pada tahap pengujian ini terdiri atas :

1. Jumlah *dataset* per kata

Dataset adalah jumlah data pelatihan untuk masing-masing kata pada *dataset*. Jumlah *dataset* yang digunakan pada pengujian ini adalah 3 set data. Nilai tersebut digunakan karena jumlah data pada setiap kata pada proses pelatihan ada tiga.

2. Jumlah Filter-bank MFCC

Parameter kedua adalah jumlah filter-bank yang degenerate dalam proses ekstraksi fitur dengan nilai parameter 20 yang didapat dari perhitungan :

$$\begin{aligned} nFilter &= [\text{ceil}(4.6 * \log_{10}(fs))]; \\ &= [\text{ceil}(4.6 * \log_{10}(16000))] \\ &= [\text{ceil}(4.6 * 4.204)] \\ &= [\text{ceil}(19.3389)] \\ &= 20 \end{aligned}$$

3. Jumlah iterasi

Nilai parameter iterasi maksimum pada proses reestimasi parameter model HMM yang digunakan untuk pengujian dalam penelitian ini adalah 2. Nilai tersebut digunakan karena semakin kecil nilai iterasi maka dapat mempercepat proses pelatihan.

4. Jumlah state model HMM

Parameter jumlah *state* pada sebuah model kata yang akan diujikan antara lain adalah 3 buah *state*. Nilai tersebut digunakan karena jenis suara terdiri dari tiga yaitu datar, sedang dan tinggi.

Tabel 5. 1 Data Hasil Pengujian Sistem

No.	kata uji 1	kata terdeteksi	kode icd10	log likelihood
1.	anemia.wav	Anemia	I50	-2093.99
2.	anemia.wav	Anemia	I50	-2107.79
3.	anemia.wav	Anemia	I50	-1998.79
4.	anemia.wav	Anemia	I50	-2045.7
5.	asma.wav	Asma	J45	-1974.24
6.	asma.wav	Asma	J45	-1924.3
7.	asma.wav	Asma	J45	-1945.9
8.	asma.wav	Asma	J45	-2026.42
9.	buta.wav	Buta	H54	-1879.2
10.	buta.wav	Buta	H54	-1879.2
11.	buta.wav	Buta	H54	-2181.81
12.	buta.wav	Buta	H54	-1878.55
13.	diare.wav	Diare	A09	-1771.03
14.	diare.wav	Diare	A09	-1815.02
15.	diare.wav	Diare	A09	-1871.52
16.	diare.wav	Diare	A09	-1781.23
17.	epilepsi.wav	Epilepsi	G40.9	-2108.73
18.	epilepsi.wav	Epilepsi	G40.9	-2105.95
19.	epilepsi.wav	Epilepsi	G40.9	-2108.73
20.	epilepsi.wav	Epilepsi	G40.9	-2108.73
21.	filariasis.wav	Filariasis	B74	-2395.2
22.	filariasis.wav	Filariasis	B74	-2384.12
23.	filariasis.wav	Filariasis	B74	-2294.57
24.	filariasis.wav	Filariasis	B74	-2315.9
25.	gastritis.wav	Gastritis	K29	-2140.34
26.	gastritis.wav	Gastritis	K29	-2040.93
27.	gastritis.wav	Gastritis	K29	-2083.81
28.	gastritis.wav	Gastritis	K29	-2146.43
29.	hematoma.wav	Hematoma	T14.0	-1977.96
30.	hematoma.wav	Hematoma	T14.0	-1947.58
31.	hematoma.wav	Hematoma	T14.0	-1972.98
32.	influenza.wav	Influenza	J10	-3224.97
33.	influenza.wav	Influenza	J10	-3215.11
34.	influenza.wav	Influenza	J10	-3062.52
35.	kolera.wav	Kolera	A00	-3902.5
36.	kolera.wav	Kolera	A00	-4026.8
37.	kolera.wav	Kolera	A00	-3901.08

persentase kesalahan kecocokan

$$= \frac{\sum \text{data yang cocok}}{\sum \text{data input}} \times 100\%$$

$$= \frac{37}{37} \times 100\%$$

$$= 100\%$$

Dari hasil perhitungan persentase di atas, dapat disimpulkan bahwa semua data yang dimasukkan dalam pengujian memperoleh hasil yang sama. Dengan kata lain, kata diagnosis yang diuji dapat mengenali suara dan mampu mencocokkan suara uji.

C. Proses Ekstraksi Suara dengan Metode MFCC

a. Implementasi proses ekstraksi suara dengan metode MFCC

Untuk melakukan pengolahan sinyal suara pada aplikasi ini menggunakan masukan:

F_s = frekuensi sampling 16000 Hz

N = panjang frame

M = panjang overlap tiap frame

$nFilter$ = jumlah filter dalam melfilterbank

set_num = jumlah training set

$loop$ = jumlah iterasi

$nstate$ = jumlah state pada arsitektur model

HMM

1) Menghitung nilai N :

Panjang frame yang digunakan untuk pemrosesan sinyal antara 15-20 ms. Dalam penelitian ini menggunakan 20 ms.

N = frame size = ukuran frame

Rumus :

$N = \text{sample rate} * T_s (\text{detik})$

Dimana :

sample rate : banyaknya jumlah sample (f) yang diambil dalam satuan waktu

$N = \text{sample rate} * T_s (\text{detik})$

$$= 16000 * 20 \text{ ms}$$

$$= 16000 * 0.02 \text{ s}$$

$$= 320$$

2) Menghitung nilai M

$$M = N/2 = 320/2 = 160$$

3) Menghitung nilai $nFilter$

$$nFilter = \lceil \log_{10}(4.6 * \log_{10}(fs)) \rceil;$$

$$= \lceil \log_{10}(4.6 * \log_{10}(16000)) \rceil$$

$$= \lceil \log_{10}(4.6 * 4.204) \rceil$$

$$= \lceil \log_{10}(19.3389) \rceil$$

$$= 20$$

1. Vektor sinyal suara dibagi menjadi bagian-bagian kecil dengan panjang *frame* n dan *overlap frame* m .
2. Melakukan proses *framing* dengan panjang *frame* n dan *overlap* m .
3. Melakukan perkalian terhadap masing-masing *frame* dengan *window hamming* sepanjang n .
4. Melakukan transformasi *frame* ke dalam domain frekuensi.

5. Memberikan nilai jumlah *filter bank* f yang didapat dari mengkonversi frekuensi biasa ke dalam skala mel.
6. Filter bank yang terbentuk dikalikan dengan hasil FFT. Kemudian, melakukan *Discrete Fourier Transform* terhadap hasil yang didapat dari perhitungan.
7. Mereduksi jumlah fitur sesuai dengan nilai koefisien MFCC (c).
8. Hasil dari proses ini berupa fitur suara matrik dengan ordo $m \times n$.

4) Proses Pemodelan dengan Metode HMM

1. State

- a. Hidden state terdiri dari 2 state yaitu high (H) dan low (L).
- b. Visible state merupakan urutan observasi yang ditentukan secara acak.

Contoh : $Q = \{ABBA\}$

2. Inisialisasi probabilitas

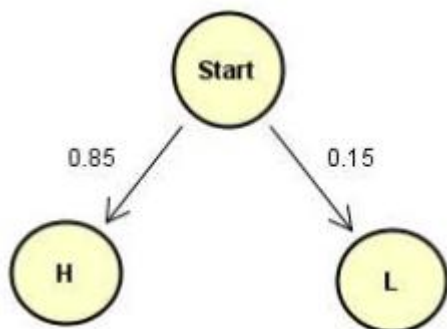
Inisialisasi probabilitas ditentukan secara sembarang. Tetapi, jumlah probabilitas untuk semua state sama dengan 1.

$$H \rightarrow 0.85$$

$$L \rightarrow 0.15$$

Probabilitas = probabilitas H + probabilitas

$$L = 0.85 + 0.15 = 1$$

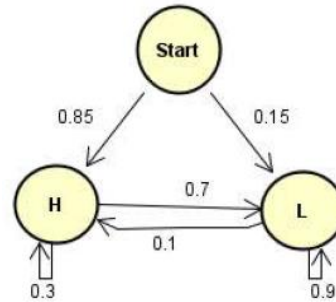


3. Probabilitas transisi

Probabilitas transisi adalah probabilitas antar *state*.

$$H \rightarrow H = 0.3 \quad L \rightarrow L = 0.9$$

$$H \rightarrow L = 0.7 \quad L \rightarrow H = 0.1$$



4. Probabilitas emisi

Probabilitas emisi adalah probabilitas pada *visible state*.

a. Pada state H :

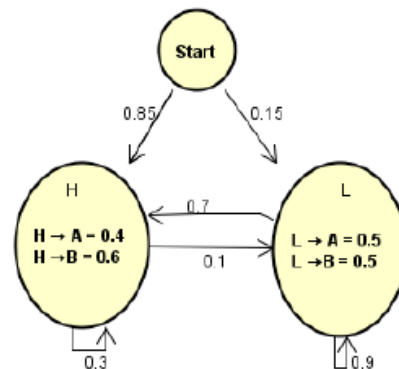
$$H \rightarrow A = 0.4$$

$$H \rightarrow B = 0.6$$

b. Pada state L

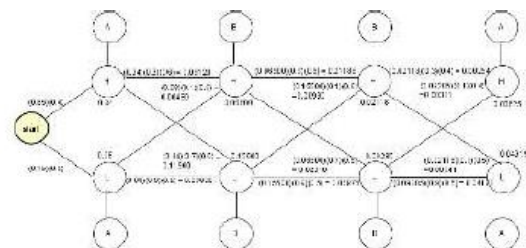
$$L \rightarrow A = 0.5$$

$$L \rightarrow B = 0.5$$



Perhitungan peluang dengan prosedur forward

Untuk $Q = \{ABBA\}$



$$\text{Probabilitas untuk } Q = 0.00625 + 0.04919 = 0.05544$$

$$\text{Nilai log likelihood dari } Q = \log 0.05544 = -2.8925$$

VI. PENUTUP

A. Kesimpulan

Dari penelitian yang telah dilakukan, didapatkan beberapa kesimpulan yaitu :

1. Aplikasi *speech-to-text* kode ICD-10 dengan metode MFCC (*Mel Frequency Cepstral Coefficient*) dan HMM (*Hidden Markov Model*) mampu memberikan informasi mengenai kode ICD-10 dari nama diagnosis penyakit yang dimasukkan dalam proses pengujian.
2. Hasil kecocokan data masukan dan luaran dari pengujian yang telah dilakukan menggunakan parameter jumlah data set 3, filter bank 20, iterasi 2 dan state 3 mendapatkan nilai persentase 100%.

B. Saran

Saran yang dapat digunakan untuk pengembangan sistem lebih lanjut antara lain adalah :

1. Pengembangan dari segi kebahasaan agar dapat mengenali kata dengan urutan kata yang sesuai dengan kaidah bahasa Indonesia sehingga tidak adanya perulangan kata pada saat pengenalan kata.
2. Sebaiknya, pengembangan selanjutnya penyimpanan data ekstraksi suara dan pemodelan kata di simpan dalam *database* mysql.

- [4] Nilsson, M., Einarson, M. (March 2002). "Speech Recognition using Hidden Markov Model : performance evaluation ins noisy enviroenment. *Departement of Telecommunication and Signal Processing*. Ronneby.
- [5] Dave, Namrata. (2013). Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition. *International Journal for Advance Research in Engineering and Technology, (Volume 1, Issue VI, July 2013)*.
- [6] Manunggal, HS. 2005. Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara dengan Menggunakan Analisa MFCC *Feature Extraction*. Surabaya : Universitas Kristen Petra.
- [7] Prasetyo, B., & Jannah, L. M. (2012). *Metode Penelitian Kuantitatif Teori dan Aplikasi*. Jakarta: Rajawali Pers.
- [8] Uchat (2009), N. S. A Seminar *Ressport on Hidden Markov Model and Speech Recognition*. *Department of Computer Science and Engineering Indian Institute of Technology*, Bombay Mumbai.

REFERENSI

- [1] WHO (2004). *International Statistical Classification of Diseases and Related Health Problems Tenth Revision Volume 1, 2 dan 3*. Geneva: WHO.
- [2] Anusuya, A & Katti, K. 2009. Speech Recognition by Machine. Volume 6, No. 3. Department of Computer Science and Engineering. India.
- [3] Chandra, D. (2007). *Speech Recognition CoProcessor*. North Carolina: issertation Doctor of Philosophy North Carolina State University.