

PENGENALAN PEMBICARA UNTUK ANALISA SUARA PADA BARANG BUKTI DIGITAL MENGGUNAKAN METODE LEARNING VECTOR QUANTIZATION (LVQ)

SPEAKER RECOGNITION FOR DIGITAL FORENSIC AUDIO ANALYSIS USING LEARNING VECTOR QUANTIZATION METHOD

¹Danny Bastian Manurung, ²Ir. Burhanuddin Dirgantoro, M.T, ³Casi Setianingsih, S.T., M.T.

^{1,2,3}Prodi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom

dannybastian@student.telkomuniversity.ac.id, burhanuddin@telkomuniversity.ac.id,

setiacasie@telkomuniversity.ac.id

Abstrak

Pada Tugas Akhir ini dibuat sebuah prototype aplikasi yang dapat mengklasifikasikan suara yang didapat dari rekaman telepon yang berbasis teknik pengenalan pembicara (Speaker Recognition) untuk dapat melakukan klasifikasi suara pembicara yang terdapat pada barang bukti berupa rekaman telepon dan suara tertuduh. Tahapan yang digunakan untuk mengenali suara tersebut adalah dengan mengekstraksi ciri suara menggunakan metode Mel-frequency Cepstral Coefficients (MFCC) dan menggunakan metode Jaringan Saraf Tiruan Learning Vector Quantization (JST-LVQ) sebagai metode klasifikasi hasil ekstraksi ciri suara tersebut.

Dengan menggunakan metode LVQ, keakuratan dalam melakukan klasifikasi suara untuk selanjutnya dapat dikenali identitas pemilik suara tersebut cukup baik. Penggunaan metode LVQ menghasilkan akurasi sebesar 73,33% untuk dapat mengenali kalimat yang sama dan 46,67% untuk kalimat yang berbeda. Sehingga hasil yang di harapkan sesuai dengan yang di harapkan

Kata Kunci : *Speaker Recognition, Mel-frequency Cepstral Coefficients, Learning Vector Quantization*

Abstract

In this final project is made a application's prototype that can be used to classify and in that case will be done speaker recognition technique (Speaker Recognition) to be able to classify the speaker's voice in the evidence and the voice of the suspect. The stages used to compare the sound is by extracting the sound features using the Mel-frequency Cepstral Coefficients (MFCC) method and using the Learning Vector Quantization Neural Network (JST-LVQ) method as the classification method of the voice extraction result.

By using LVQ, the accuracy in recognizing the speaker's voice is pretty good. The use of LVQ method produces best accuracy at 73,33% to recognize the speaker that with the same sentence, and 46,67% for different sentence. So the results obtained in accordance with the expected.

Keywords : *Speaker Recognition, Mel-frequency Cepstral Coefficients, Learning Vector Quantization*

1. Pendahuluan

Pada era yang berkembang ini, penggunaan teknologi telah mendominasi dalam setiap bidang, baik dalam pekerjaan, berkomunikasi hingga pengolahan data. Jangkauan teknologi yang sangat luas ini memberikan kemudahan bagi kriminalitas untuk melakukan tindak kriminal, seperti kemudahan melakukan komunikasi.

Kemudahan ini juga pada satu sisi sangat membantu pihak penegak hukum dalam melakukan pengumpulan barang bukti digital yang salah satunya adalah barang bukti berupa rekaman untuk dapat mengenali pelaku kejahatan. Rekaman suara dihasilkan oleh artefak digital, rekaman suara sendiri merupakan kumpulan frekuensi yang mengandung informasi dan ciri [1]. Informasi ini dapat digunakan sebagai petunjuk, identitas individu, lokasi, waktu dan sebagainya.

Untuk mempermudah pada proses pengenalan suara yang ada pada artefak digital, Tugas Akhir ini membuat prototype aplikasi pengenalan penutur/pembicara (*speaker recognition*) dalam barang bukti berupa rekaman suara. Proses pengenalan penutur/pembicara tersebut dilakukan dengan membandingkan rekaman suara yang tidak di ketahui identitas individunya (*unknown*) dengan suara pembanding yang telah di ketahui identitasnya (*known*) [2] menggunakan metode Jaringan Saraf Tiruan *Learning Vector Quantization* (JST-LVQ). Dengan menggunakan metode Jaringan Saraf Tiruan *Learning Vector Quantization* (JST-LVQ), aplikasi ini dapat membantu proses pengenalan pembicara pada barang bukti digital dengan nilai akurasi yang baik karena mampu mengklasifikasikan pola masukan dan pola keluaran serta toleran terhadap kesalahan.

2. Dasar Teori

2.1 Suara

Pada prinsipnya suara terdiri dari beberapa komponen yang akan digunakan untuk mengidentifikasi karakteristik suara seseorang untuk kepentingan speaker recognition. komponen tersebut adalah pitch, formant dan spectogram.

- a. Pitch : frekuensi getar pada pita suara. Masing-masing orang memiliki pitch yang berbeda-beda atau khas yang sangat dipengaruhi oleh aspek fisiologis manusia
- b. Formant : frekuensi-frekuensi resonansi dari filter, yaitu vocal tract yang meneruskan dan memfilter bunyi periodik dari getarnya pita suara menjadi bunyi keluaran
- c. Spectrogram : bentuk visualisasi dari masing-masing nilai formant yang dilengkapi dengan level energi yang bervariasi terhadap waktu

2.2 Speaker Recognition

Speaker Recognition adalah proses pengenalan secara otomatis siapa yang berbicara, proses pengidentifikasian ini dilakukan dengan mengidentifikasi identitas seseorang melalui frekuensi suaranya [5] pada umumnya, pengenalan pembicara terbagi menjadi dua yaitu pengenalan yang bergantung pada teks (Text-Dependent Recognition) dan pengenalan yang tidak bergantung pada teks (text-independent recognition). Untuk pengenalan yang bergantung pada teks, pembicara hanya diizinkan untuk mengatakan beberapa kalimat atau kata-kata tertentu yang diketahui oleh sistem. Sebaliknya, pengenalan yang tidak bergantung pada teks, sistem dapat memproses suara tanpa harus terpaku pada kata-kata atau kalimat tertentu [6].

Teknik pengenalan pembicara dapat diklasifikasikan kedalam verifikasi dan identifikasi. Verifikasi pembicara adalah proses menerima atau menolak klaim identitas atas suatu suara sedangkan identifikasi adalah proses penentuan pembicara mana yang terdaftar pada sistem yang memberikan sebuah ucapan atau suara [6].

2.3 Preprocessing

Preprocessing merupakan tahapan awal yang merupakan tahapan dimana dilakukan pengolahan pada data input sebelum masuk ke tahap ekstraksi ciri pada data. Preprocessing pada data suara yaitu dimulai dengan inialisasi daerah silent pada suara, yang kemudian akan dilakukan proses normalisasi, penghilangan noise serta proses silent removal menggunakan daerah silent yang telah ditentukan.

2.4 Ekstraksi Ciri

Ekstraksi Ciri merupakan sebuah proses untuk mengekstrak ciri-ciri fitur dari suara pembicara yang bertujuan untuk mengkonversi sinyal suara menjadi beberapa jenis representasi parametrik (Pada tingkat informasi yang jauh lebih rendah). Data masukan dipilah-pilah dan dimasukkan dan di proses satu-persatu [7]

2.5 Mel-frequency Cepstral Coefficients (MFCC)

Mel-frequency Cepstral Coefficients merupakan sebuah metode ekstraksi ciri pada sinyal suara yang berdasarkan prinsip karakteristik pendengaran manusia. Kemampuan pendengaran manusia terhadap sinyal suara tidak berskala linear. Filter dengan skala linear di tempatkan pada frekuensi rendah pada sinyal, dan secara logaritmik pada frekuensi tinggi, yang digunakan untuk menangkap karakteristik fonetik yang penting pada sinyal suara. Sinyal suara dinyatakan dalam MEL Scale, dimana filter linear digunakan pada frekuensi dibawah 1000 Hz dan filter logaritmik digunakan pada frekuensi diatas 1000 Hz [8].

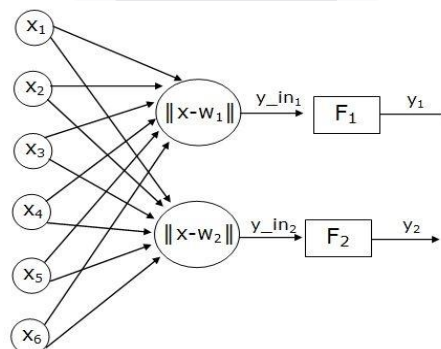
2.6 Jaringan Saraf Tiruan

Jaringan Saraf Tiruan (JST) adalah prosesor yang terdistribusi besar-besaran secara paralel yang dibuat dari unit proses sederhana, yang mempunyai kemampuan untuk menyimpan pengetahuan berupa pengalaman dan dapat digunakan untuk proses lain.

Manusia memiliki 100 miliar sel saraf neuron yang terletak pada bagian otak, kebanyakan sel neuron tersebut tidak diganti lagi ketika sel tersebut mati. Dengan kehilangan neuron tersebut, manusia bisa tetap belajar. Dengan mengacu pada hal tersebut dapat dibentuk suatu Jaringan Saraf Tiruan (JST) atau Artificial Neural Network (ANN) yang dapat meniru cara kerja sel saraf biologis, dan dapat dimodelkan yang kemudian dapat dibentuk dalam bentuk sebuah algoritma pemrograman [11].

2.7 Learning Vector Quantization (LVQ)

Learning Vector Quantization (LVQ) merupakan bentuk khusus dari Competitive Learning Algorithm (Unsupervised Learning), namun LVQ mempunyai target, sehingga LVQ juga termasuk metode yang melakukan pembelajaran pada lapisan kompetitif yang membutuhkan pengawasan (Supervised Learning) . Jaringan LVQ terdiri dari 2 jenis lapisan yaitu lapisan kompetitif dan lapisan linear.



Gambar 1 . Arsitektur JST-LVQ

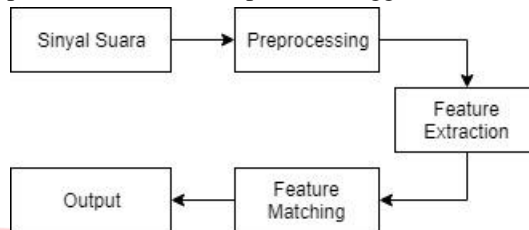
Arsitektur jaringan LVQ dengan n unit pada lapisan masukan, dan l unit (neuron) pada lapisan keluaran. Proses yang terjadi pada setiap neuron adalah mencari jarak antara vektor masukan ke bobot yang menghubungkan ke bobot yang bersangkutan ($w_1, w_2, w_3, \dots, w_m$). w_1 adalah vektor bobot yang menghubungkan setiap neuron pada lapisan

masuk ke neuron pertama pada lapisan keluaran, sedangkan w_2 adalah vektor bobot yang menghubungkan setiap neuron pada lapisan masukan ke neuron kedua pada lapisan keluaran, demikian seterusnya. Fungsi aktivasi F1 akan memetakan y_{in1} ke $y_1=1$ apabila $|x-w_1| < |x-w_2|$, dan $y_1=0$ jika sebaliknya. Demikian juga dengan fungsi aktivasi F2 akan memetakan y_{in2} ke $y_2=1$ apabila $|x-w_2| < |x-w_1|$, dan $y_2=0$ jika sebaliknya dan seterusnya [13].

3. Perancangan Sistem

3.1 Gambaran umum

Pada tugas akhir ini, perancangan sistem secara umum adalah membuat sistem pengenalan pembicara pada barang bukti digital, dan sistem diproses dalam sebuah aplikasi menggunakan bahasa pemrograman python.



Gambar 2 Blok Diagram sistem

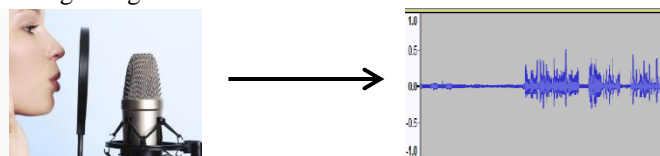
Gambar diatas merupakan blok diagram sistem secara umum, sistem ini dibangun dengan dua tahapan yaitu tahap pelatihan (Training) dan tahap pengujian (testing). pengerjaan dalam tugas akhir ini adalah melakukan perbandingan sinyal suara yang digunakan sebagai data uji dengan sinyal suara yang digunakan sebagai data latih.

3.2 Perancangan Sistem

Pada bagian perancangan sistem akan dijelaskan lebih detail perancangan pada tahap training dan testing. Berikut langkah-langkah dalam proses perancangan pada sistem ini.

3.2.1 Perekaman suara

Pada tahap ini dilakukan proses perekaman suara yang akan digunakan sebagai data train. Proses perekaman suara menggunakan mikrofon Razer Seiren dan aplikasi Audacity. Sample suara diambil dari 5 orang laki-laki sebagai terduga dalam skenario dengan frequency sampling 16000 Hz dengan format file .wav. Terdapat 33 kata yang di ucapkan 5 kali oleh masing-masing orang.

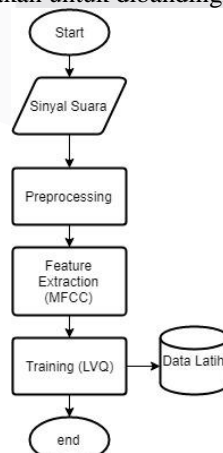


Gambar 3 Proses perekaman suara

Proses perekaman suara menggunakan aplikasi Audacity untuk mempermudah proses pengeditan suara yaitu proses pemotongan suara yang tidak perlu seperti pemotongan sinyal suara silence dan sinyal suara dimana si pembicara melakukan penarikan nafas pada awal proses perekaman suara dan pembuangan nafas pada akhir proses perekaman suara.

3.2.2 Training

Pada tahap ini, sinyal suara yang telah diterima melalui rekaman yang telah dilakukan, akan di olah sebagai data train yang tujuannya adalah untuk melatih sistem mengenali dan memahami parameter-parameter yang ada pada setiap suara pada masing-masing orang. Setelah melalui tahap pelatihan, parameter-parameter keluaran dari pelatihan akan disimpan sebagai data latih yang akan digunakan untuk dibandingkan data uji.



Gambar 4 Diagram alir proses training

Pada diagram diatas dapat dilihat data sinyal suara yang digunakan sebagai inputan akan dilakukan preprocessing menggunakan aplikasi Audacity, setelahnya akan dilakukan proses ekstrasi ciri dari sinyal suara inputan menggunakan metode Mel-Frequency Cepstral Coefficients (MFCC). Setelah mendapatkan nilai koefisien dari MFCC yang

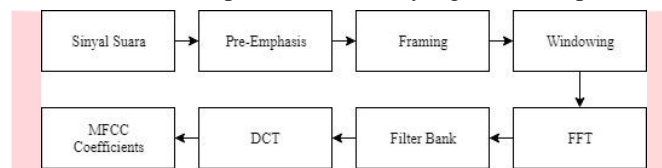
merupakan parameter-parameter suara, akan dilakukan training pada nilai-nilai tersebut menggunakan metode Jaringan Saraf Tiruan Learning Vector Quantization (LVQ), setelah mendapatkan nilai pelatihan, nilai-nilai tersebut akan disimpan sebagai data train yang akan digunakan sebagai nilai pembandingan pada tahap data uji.

a. Preprocessing

Pada tahap preprocessing ini dilakukan beberapa pengolahan pada sinyal suara yang telah direkam. Tahap ini dilakukan untuk mendapatkan sinyal suara yang dibutuhkan. Proses pengolahan pada sinyal suara ini yaitu dengan menghilangkan suara-suara yang tidak perlukan, process ini melakukan remove silence pada sinyal suara, yaitu dengan menggunakan aplikasi Audacity. Setiap sample suara memiliki Silence pada awal sinyal, yaitu kondisi dimana pembicara tidak mengeluarkan suara apapun, sinyal silence tersebut akan digunakan sebagai noise profil yang digunakan saat membersihkan noise dari seluruh sinyal. Proses ini dilakukan untuk mendapatkan nilai akurasi yang tinggi pada sistem. Setelah menghilangkan noise pada sinyal suara, dilakukan proses pemotongan sinyal-sinyal yang tidak di butuhkan pada awal dan akhir sinyal.

b. Feature Extraction (MFCC)

Pada tahap Feature Extraction (Ekstrasi cir) pada sinyal suara menggunakan metode Mel-Frequency Cepstral Coefficients. Metode ini sangat populer digunakan karena proses MFCC ini sangat baik dalam proses pengolahan suara. Proses ini akan memiliki keluaran berupa feature vector yang disebut cepstrum.



Gambar 5 Diagram Proses MFCC

Gambar diatas merupakan tahap-tahap proses MFCC yang terdiri dari :

a. Pre-emphasis

Pre-Emphasis merupakan langkah pertama dalam proses ekstraksi ciri menggunakan MFCC. Pre-emphasis menyaring sinyal suara agar sinyal menjadi lebih jelas ketika masuk ke tahap selanjutnya. Beberapa fungsi Pre-emphasis adalah menyeimbangkan spektrum frekuensi karena biasanya frekuensi yang tinggi memiliki magnitude yang lebih kecil dari pada frekuensi yang lebih rendah, menghindari masalah numerik pada proses transformasi Fourir dan dapat memperabiki Signal-to-Noise Rasio (SNR). Pada tahap ini dilakukan filtering pada sinyal menggunakan FIR filter orde satu untuk meratakan spektral pada sinyal tersebut. Proses ini sudah mencakup proses penambahan energi suara pada frekuensi yang tinggi. Berikut persamaan matematis pada Pre-emphasis :

$$s'_n = s_n - a s_{n-1} \quad (1)$$

Dimana s'_n adalah sinyal hasil dari proses pre-emphasis dan s_n adalah sinyal masukan yang belum dilakukan proses pre-emphasis. Dan a adalah konstanta dengan nilai 0.97 yang menunjukkan sinyal yang diekstrak merupakan sinyal yang memiliki 97% sinyal aslinya. s_n harus dimulai sinyal kedua dari sinyal masukan, karena s_{n-1} merupakan sinyal masukan sebelum sinyal s_n .

b. Framing

Pada tahap ini suara dibagi menjadi beberapa frame dan masing-masing frame terdiri dari N sampel sinyal yang saling berdekatan yang dipisahkan jarak antar sample M, langkah ini dilakukan karena frekuensi dalam sebuah sinyal dapat berubah-ubah setiap waktu, sehingga akan sulit melakukan proses transformasi Fourir melalui seluruh sinyal. Pada umumnya panjang tiap frame sangat singkat, sekitar 20 sampai 40 ms. Nilai $M < N$ sehingga akan ada bagian pada frame yang saling tumpang tindih (Overlap). Keuntungan dari framing adalah memudahkan analisis dan mengurangi alokasi memori.

Pada tugas akhir ini, panjang frame diatur 25ms dan jarak antar frame 10ms sehingga akan memiliki 15ms sinyal yang saling overlap.

c. Windowing

Setelah dilakukan proses framing pada hasil keluaran Pre-emphasis, kemudian diterapkan Windowing pada sinyal ini. Pada saat sinyal suara di dibagi menjadi beberapa frame akan menyebabkan efek diskontinuitas atau ada nya noise pada awal dan akhir sinyal. Tahap ini berfungsi untuk meredam noise yang muncul di kedua ujung frame. Teknik windowing yang digunakan dalam perancangan sistem ini adalah Hamming Window. Berikut rumus dari Hamming Window :

$$H(n) = 0,54 - 0,46 \cos\left(2\pi \frac{n}{N-1}\right) \quad (2)$$

Dimana N adalah panjang sinyal yang ada dalam sebuah frame, dan n adalah posisi sinyal yang akan kemudian akan dihitung.

d. Fast Fourier Transform (FFT)

Pada tahap ini, dapat dilakukan N-point FFT pada setiap frame untuk menghitung nilai spektrum frekuensi, yang di sebut juga Short Time Fourir Transform (STFT), dimana N biasanya bernilai 256 atau 512, pada perancangan sistem

pengenalan suara ini digunakan NFFT bernilai 512, kemudian untuk menghitung spektrum daya (periodogram) dapat menggunakan persamaan berikut [9] :

$$P = \frac{|FFT(x_i)^2|}{N} \tag{3}$$

e. Filter Bank

Performansi dari MFCC juga dipengaruhi oleh salah-satunya adalah jumlah filter yang digunakan, jumlah filter yang terlalu banyak atau yang terlalu sedikit dapat menghasilkan akurasi yang tidak baik [14]. Pada umumnya filter yang digunakan sebanyak 40 filter. Sinyal suara pada domain frekuensi diubah menjadi domain frekuensi mel dimana rumusnya adalah sebagai berikut :

$$Mel(f) = 2595(\ln(1 + \frac{f}{700})) \tag{4}$$

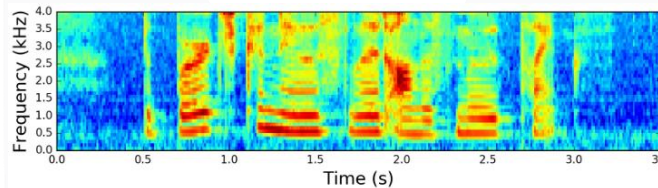
Mel (f) adalah nilai frekuensi mel dari f. Hasil akhir dari tahap ini akan mendapatkan sejumlah mel filter bank. Nilai mel filter bank menunjukkan besar energi pada rentang frekuensi yang ada pada masing-masing filter mel. Rumus untuk kembali ke frekuensi dari skala Mel :

$$f = 700(10^{\frac{m}{2595}} - 1) \tag{5}$$

Setiap filter pada filter bank berbentuk segitiga yang memiliki respon 1 pada frekuensi tengah dan menurun secara linear terhadap 0 hingga mencapai frekuensi tengah dari dua filter yang berdekatan dimana responnya adalah 0. Dapat dimodelkan dengan persamaan berikut [9] :

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k < f(m) \\ 1 & k = f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \tag{6}$$

Setelah menerapkan filter bank ke spektrum sinyal (periodogram), akan di dapatkan spektrogram seperti gambar dibawah ini.



Gambar 6 Spektrogram Sinyal

f. Discrete Cosine Transform (DCT)

DCT berfungsi untuk mengembalikan sinyal suara pada domain frekuensi ke domain waktu sehingga didapatkan koefisien cepstrum.

Berikut rumus dari DCT [7]:

$$c_n = \sum_{k=1}^K (f'_k) \cos \left[n(k - 0,5) \frac{\pi}{k} \right] \tag{7}$$

K adalah jumlah mel frequency filter bank, f'k berasal dari hasil transformasi non linear, n adalah bilangan bulat dari 1 hingga N (jumlah total sampel) sehingga didapatkan N buah koefisien cepstrum. Namun masih terdapat beberapa kelemahan pada cepstrum tersebut, maka akan di lakukan proses cepstral liftering untuk memperhalus cepstrum. Berikut persamaan yang digunakan pada cepstral liftering.

$$c = 1 + ((L - 1)/2) \times \sin(\pi \times n/(L - 1)) \tag{8}$$

L adalah jumlah cepstral coefficient dan yang digunakan pada proses pengenalan suara ini berjumlah 12.

3.2.3 Training LVQ

a. Inialisasi bobot awal

Pada tahap ini kita harus menentukan bobot awal pada LVQ yaitu learning rate (α), nilai Epoch, MaxEpoch, dan bobot awal. Bobot awal dapat diambil dari vektor tiap masing-masing kelas.

b. Input Nilai

Nilai yang diinputkan pada tahap ini merupakan nilai yang didapat dari proses ekstrasi ciri. Setiap nilai sudah memiliki kelas nya masing-masing, dari nilai inputan ini akan diambil satu dari setiap kelas untuk menjadi nilai bobot awal (w_j).

c. Iterasi

Pada tahap ini, sistem mulai melakukan pembelajaran terhadap data-data yang telah di inputkan. Iterasi ini akan terus di lakukan selama nilai Epoch < MaxEpoch dan nilai learning rate (α) > Min learning rate (Min α). Setiap iterasi dilakukan dengan menghitung jarak setiap data dengan nilai bobot awal. Jika nilai data tersebut mendekati nilai bobot maka nilai bobot akan di update dengan menggunakan rumus :

$$w_j = w_j + \alpha(x_i - w_j) \tag{9}$$

Pada setiap iterasi juga nilai learning rate akan di update dengan rumus :

$$\alpha = \alpha * Decc \alpha \quad (10)$$

Setelah seluruh iterasi diselesaikan, maka nilai bobot terakhir akan menjadi nilai bobot yang di gunakan sebagai data latih.

3.2.4 Testing

a. Input Ciri Uji

Pada tahap ini, nilai hasil dari ekstraksi ciri pada data yang akan di uji disimpan. Yang kemudian akan dihitung jaraknya dengan nilai bobot akhir dari setiap kelas yang telah disimpan ke dalam dataset setelah proses pelatihan data latih.

b. Hitung Jarak

Pada tahap ini, akan dihitung jarak Euclidean antara vektor ciri data uji (x) dengan semua nilai bobot dari setiap kelas data latih (Wm) pada lapisan kompetitif. Perhitungan jarak Euclidean bisa dilakukan dengan menggunakan persamaan 3.9.

$$|X - W_m| = \sqrt{\sum_1^n (X_n - W_{m_n})} \quad (11)$$

Neuron dengan jarak paling kecil dengan nilai ciri uji akan memenangkan kompetisi.

4. Pengujian

4.1 Pengujian Dengan Kalimat Sama Dengan Data Latih

Pengujian dilakukan 10 kali dengan nilai Epoch yang berbeda beda dengan interval 10 dengan Learning Rate yang digunakan sebesar 0.3, yang di dapat dari hasil penelitian sebelumnya [15] dengan hasil 88.89%. Hasil pemenangan dari setiap bobot dapat dilihat di lampiran.

Setelah dilakukan 10 pengujian dengan nilai Epoch yang berbeda-beda dengan interval nilai 10 dengan menggunakan kalimat yang sama pada data Uji dan data Latih, didapatkan nilai akurasi setiap Epoch dan Avarage akurasi sebagai berikut :

Tabel 1. Rata-rata Akurasi Pengujian Epoch

Epoch	10	20	30	40	50	60	70	80	90	100	Avg
Akurasi	60%	50%	60%	60%	60%	60%	60%	60%	60%	60%	59%

Dari hasil pengujian tersebut dapat di simpulkan bahwa akurasi yang akan digunakan untuk pengujian Learning Rate adalah Epoch 100 untuk setiap pengujian Learning Rate dan didapatkan hasil pada 10 kali pengujian dengan Avarage sebagai berikut :

Tabel 2. Rata-Rata Akurasi Pengujian Learning Rate

LR	0,1	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01	Avg
Akurasi %	40,00	40,00	40,00	46,67	46,67	46,67	46,67	53,33	53,33	73,33	48,67

4.2 Pengujian Dengan Kalimat Beda Dengan Data Uji

Setelah dilakukan pengujian pada Epoch dan Learning Rate dengan sampel suara yang memiliki kalimat yang sama dengan data latih, selanjutnya akan dilakukan pengujian terhadap sampel suara yang memiliki kalimat yang berbeda dengan data latih, dengan jumlah kata yang terdapat pada kalimat yang diucapkan tersebut berjumlah 10, kalimat tersebut adalah "Bisa pak, bisa, besok akan saya kirimkan sebesar 4 milyar."

Pengujian dilakukan sekali dengan jumlah data uji 3 per orang yang sampel suaranya di ambil dengan melakukan komunikasi dengan 2 telepon dan pembicaraan tersebut di rekam. Pengujian dilakukan dengan Epoch terbaik pada pengujian sebelumnya yaitu 100, dan Learning Rate terbaik yaitu 0,01 dengan hasil sebagai berikut:

Tabel 3. Pengujian Beda Kalimat

Pembicara Suara ke-	Dandi	Deo	Dika	Afif	Teguh
1	Teguh	Dika	Dika	Afif	Teguh
2	Dika	Dika	Dika	Afif	Teguh
3	Dika	Dika	Teguh	Teguh	Teguh
Akurasi	46,67%				

4.3 Pengujian 29 Kata

Pengujian perkata ini dilakukan dengan menguji suara sampel yang telah direkam menggunakan telepon saat sedang melakukan komunikasi, dengan data uji 10 sampel suara dengan kata yang berbeda-beda yaitu "saya", "sudah", "siapkan", "bagian", "untuk", "anda", "dengan", "nilai", "sembilan", "milyar". Pengujian dilakukan dengan menggunakan Epoch 100 dan Learning Rate 0,01 dengan hasil sebagai berikut :

Tabel 4. Akurasi Pengujian 29 Kata

Pembicara	Klasifikasi Kata									
	saya	sudah	siapkan	bagian	untuk	anda	dengan	nilai	sembilan	milyar
Afif	✓	Deo	Deo	Dika	Teguh	✓	Dandi	Dandi	Deo	Deo
Dandi	Dika	Deo	✓	Dika	Teguh	Teguh	Dika	✓	Deo	✓
Dika	Dika	Dika	✓	Teguh	Teguh	Dika	Dandi	Dandi	Dika	✓
Deo	✓	✓	Deo	✓	Deo	Teguh	✓	Dandi	✓	Dandi
Teguh	Dika	Dandi	Deo	Diika	Afif	✓	Dika	Deo	Deo	Dika
Akurasi	26,00%									

4.4 Pengujian 1 Kata

Pengujian ini merupakan pengujian perkata, dimana sampel suara yang digunakan hanya memiliki 1 suku kata saja yaitu kata "Selesai". Sampel suara yang akan menjadi data latih direkam dengan menggunakan Microphone direkam dari 5 orang yang berbeda dengan jumlah sampel suara 4 per orang dengan total 20 sampel suara yang selanjutnya akan dilakukan proses ekstraksi ciri dan akan dilatih dengan menggunakan Epoch 100 dan Learning Rate 0,01.

Pengujian dilakukan dengan menggunakan sampel data uji yang direkam dari proses komunikasi antar dua orang melalui telepon. sampel suara uji direkam 3 sampel per orang yang kemudian akan di uji dengan hasil seperti berikut :

Tabel 5. Akurasi Pengujian 1 Kata

Pembicara Suara ke-	Dandi	Deo	Dika	Afif	Teguh
1	Dika	Deo	Dika	Afif	Dika
2	Dika	Dika	Dika	Afif	Dandi
3	Dika	Afif	Dika	Afif	Dika
Akurasi	46,67%				

4.5 Analisis Performansi

Setelah dilakukan beberapa pengujian terhadap sistem pengenalan pembicara yang telah dirancang dengan menggunakan metode MFCC dan LVQ dengan akurasi sebagai berikut :

Tabel 6. Tabel Akurasi Seluruh Pengujian

Pengujian	Akurasi
Kalimat Sama	73,33%
Kalimat Beda	46,67%
29 Kata	26,00%
1 Kata	46,67%

Kecil nya angka akurasi proses pengenalan pembicara yang bertujuan untuk mengenali suara pembicara yang direkam menggunakan telepon saat sedang melakukan komunikasi ini dikarenakan perbandingan kualitas suara antara data latih dan data uji yang sangat signifikan. Sampel suara data latih di rekam dengan menggunakan Microphone Reizer Seiren yang cukup baik untuk melakukan perekaman suara berkualitas tinggi, sedangkan data uji direkam dengan menggunakan Microphone pada Smartphone dan saat proses komunikasi sedang berlangsung, yang tentunya menyebabkan suara dikompres demi kelancaran komunikasi yang menyebabkan turunnya kualitas suara.

Pada tabel 4.18 dapat dilihat akurasi setiap pengujian yang semakin menurun ketika dilakukan pergantian isi kalimat yang di gunakan pada sampel suara data uji. Maka pada tugas akhir ini, sistem pengenalan pembicara untuk analisis suara melalui telepon hanya mampu memproses dengan baik data-data masukan yang memiliki kata-kata yang sama dengan data pembanding.

5. Kesimpulan

Berdasarkan dari hasil Tugas Akhir ini, dapat ditarik beberapa kesimpulan yaitu :

1. Speaker Recognition dengan menggunakan metode Learning Vector Quantization untuk menganalisis audio forensik dapat dilakukan dengan keakuratan sistem sebesar 73,33%.
2. Nilai Learning Rate dan Epoch yang digunakan sangat mempengaruhi keakuratan sistem. Akurasi terbaik pada sistem menggunakan nilai Learning Rate sebesar 0,01 dan Epoch 100. Banyaknya data latih juga sangat berpengaruh terhadap akurasi identifikasi.
3. Akurasi tertinggi untuk identifikasi didapat sebesar 73,33% pada pengujian per kalimat dibanding dengan per kata, hal ini disebabkan oleh metode ekstraksi ciri Mel-frequency Cepstral Coefficient yang mendekati sistem pendengaran manusia, sehingga semakin panjang durasi data masukan yang diberikan, akan lebih baik hasil yang di dapat.

6. Referensi

1. Huizen, Roy Rudolf, Jayanti, Ni ketut Dewi Ari, Hostiadi, Dandy Pramana. *Model Acquisisi Rekaman Suara Di Audio Forensik*. Bali : STIKOM BALI.
2. Mohammed Algabri, Hassan Mathkour, Mohamed A. Bencherif, Mansour Alsulaiman, and Mohamed A. Mekhtiche. *Automatic Speaker Recognition for Mobile Forensic Applications*. Saudi Arabia : King Saud University.
3. Prayudi, Yudi. 2013. *Teknik Forensika Audio Untuk Analisa Suara Pada Barang Bukti Digital*. Yogyakarta : ResearchGate.
4. Dewi, Irma Amelia. 2012. *DIGITAL FORENSIC VOICE*. Bandung : Institut Teknologi Bandung.
5. Putra, Doanda Khabi. 2017. *Simulasi Dan Analisis Speaker Recognition Menggunakan Metode Mel Frequency Cepstrum Coefficient (MFCC) Dan Gaussian Mixture Model (GMM)*. Bandung : Universitas Telkom.
6. Sharma, Varun. 2013. *A Review On Speaker Recognition Approaches And Challenges*. International Journal of Engineering Research and Technology (IJERT).
7. Fawziah, Siti Khodijah Fathonatun Nurul. 2013. *Pemodelan Speech Recognition Speech-to-Text Dalam Bahasa Indonesia Menggunakan Mel Frequency Cepstral Coefficients (MFCC) dan Hidden Markov Model (HMM)*. Bandung : Tugas Akhir Institut Teknologi Telkom.
8. De Lara, José Ramón Calvo. 2005. *A Method of Automatic Speaker Recognition Using Cepstral Features and Vectorial Quantization*. Springer-Verlag Berlin Heidelberg.
9. Fayek, Haytham. 2016. *Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between* [Online] Available at: haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html#fn:1 [Access: Juny 14, 2018].
10. Muhammad, Hariz Zakka. 2018. *Perancangan Dan Implementasi Prototipe Speech To Text Pada Alat Penerjemah Bahasa Inggris Ke Bahasa Indonesia Menggunakan Klasifikasi Hidden Markov Model*. Bandung : Tugas Akhir Institut Teknologi Telkom.
11. Putra, Roba Laba. 2011. *Aplikasi Pengenalan Suara Untuk Request Lagu Menggunakan Jaringan Syaraf Tiruan Back Propagation (JST-BP)*. Bandung, Telkom University.
12. Siang, J.J. 2009. *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan MATLAB*. Yogyakarta: ANDI.
13. Aprillia, Bahagia. 2011. *Desain Dan Implementasi Pengenalan Individu Melalui Sinyal Suara Menggunakan Metode Learning Vector Quantization*. Bandung, Telkom University
14. Tiwari, Vibha. 2010. *MFCC And Its Application In Speaker Recognition*. International Journal on Emerging Technologies.
15. Widodo, Sukoreno Mukti. 2016. *Penerapan Metode Mel Frequency Cepstral Coefficient Dan Learning Vector Quantization Untuk Text-Dependent Speaker Identification*. Bandung, Institut Teknologi Harapan Bangsa.