

# Perbandingan MFCC dan Filterbank Energies dalam Metode HMM untuk Identifikasi Genre Musik

Irsya Aufa Ambang Ramadhan<sup>1</sup>, Anita Qoiriah<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Universitas Negeri Surabaya

E-mail: [irsya.19019@mhs.unesa.ac.id](mailto:irsya.19019@mhs.unesa.ac.id), [anitaqoiriah@unesa.ac.id](mailto:anitaqoiriah@unesa.ac.id)

**Abstrak**— Musik adalah seni mengaransemen dan menciptakan suara, seringkali dengan tujuan menciptakan respons emosional atau estetika pada pendengarnya yang mencakup kombinasi melodi, harmoni, ritme, dinamika, dan timbre, yang bersatu untuk menciptakan pengalaman pendengaran yang kohesif dan menyenangkan. Genre musik adalah kategori yang digunakan untuk mengklasifikasikan musik berdasarkan karakteristik gaya umum, seperti ritme, melodi, harmoni, instrumentasi, dan pengaruh budaya. Klasifikasi genre musik adalah salah satu metode di bidang pengambilan informasi musik (MIR), dengan aplikasi mulai dari rekomendasi musik berbasis konten hingga pembuatan daftar putar otomatis. Penelitian ini memanfaatkan *Spectral Features* yang digunakan dalam klasifikasi genre musik. *Mel-Frequency Cepstral Coefficients* (MFCC) adalah teknik ekstraksi fitur yang umum digunakan dalam pemrosesan sinyal ucapan dan audio untuk merepresentasikan karakteristik spektral sinyal audio dengan cara yang relevan secara persepsi dan efisien secara komputasi. *Filterbank Energies* menggambarkan distribusi energi di berbagai pita frekuensi dalam sinyal audio. *Hidden Markov Models* (HMMs) adalah model probabilistik yang biasa digunakan dalam tugas pemrosesan ucapan dan audio, termasuk klasifikasi genre musik. Penelitian ini menggunakan data yang diambil dari dataset yang diperoleh dari Kaggle bernama GTZAN Dataset dengan genre musik yang terdiri dari *blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, dan rock*. Hasil pengujian pada rasio 95-5 dari MFCC mendapatkan 70% akurasi dalam waktu 4 jam 11 menit 43 detik, sementara *Filterbank Energies* mendapatkan 56% akurasi dalam waktu 1 jam 43 menit 16 detik.

**Kata Kunci**— (Musik, Genre, *Mel-Frequency Cepstral Coefficients*, *Filterbank Energies*, *Hidden Markov Models*)

## I. PENDAHULUAN

Seiring banyaknya musik yang tercipta setiap harinya, kreatifitas pencipta musik dalam membuat musik memiliki beberapa kategori genre yang bisa dikategorikan. Genre musik adalah kategori yang digunakan untuk mengklasifikasikan musik berdasarkan karakteristik gaya umum, seperti ritme, melodi, harmoni, instrumentasi, dan pengaruh budaya [1]. Klasifikasi genre musik adalah salah satu metode di bidang pengambilan informasi musik (MIR), dengan aplikasi mulai dari rekomendasi musik berbasis konten hingga pembuatan daftar putar otomatis [2]. Berdasarkan penelitian oleh Sebastian Dalin-Volsing [3], penelitian ini memanfaatkan *Spectral Features* yang digunakan dalam klasifikasi genre musik. Menurut penelitian dari Xinru Liu [4], *Mel-Frequency Cepstral Coefficients* (MFCC) adalah teknik ekstraksi fitur yang umum digunakan dalam pemrosesan sinyal ucapan dan

audio. Tujuannya adalah untuk merepresentasikan karakteristik spektral sinyal audio dengan cara yang relevan secara persepsi dan efisien secara komputasi. Sesuai penelitian dari Climent Nadeu [5], *Filterbank Energies* menggambarkan distribusi energi di berbagai pita frekuensi dalam sinyal audio. *Filterbank Energies* sering digunakan sebagai fungsi berbagai tugas pemrosesan sinyal audio, seperti pengenalan suara, klasifikasi genre musik, dan deteksi peristiwa suara. *Hidden Markov Models* (HMMs) adalah model probabilistik yang biasa digunakan dalam tugas pemrosesan ucapan dan audio [6], termasuk klasifikasi genre musik.

Penelitian ini menggunakan data yang diambil dari dataset yang diperoleh dari Kaggle bernama GTZAN Dataset dengan format audio '.wav' berjumlah 1000 file, dengan genre musik yang terdiri dari *blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, dan rock*.

## II. TINJAUAN PUSTAKA

Musik adalah seni mengaransemen dan menciptakan suara, seringkali dengan tujuan menciptakan respons emosional atau estetika pada pendengarnya yang mencakup kombinasi melodi, harmoni, ritme, dinamika, dan timbre, yang bersatu untuk menciptakan pengalaman pendengaran yang kohesif dan menyenangkan [1]. Berdasarkan peneliti Sebastian Dalin-Volsing dan buku oleh Kelefa Sanneh [3], [7], genre musik yang sudah diketahui secara umum, dan digunakan dalam penelitian ini, terdiri dari *blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, dan rock*.

*Equalizer* (EQ) sendiri adalah alat yang digunakan dalam pemrosesan audio untuk menyesuaikan keseimbangan antar komponen frekuensi dalam sinyal audio [8]. EQ membagi spektrum suara menjadi beberapa pita, dan setiap pita berhubungan dengan persepsi manusia terhadap nada berupa bass (20–250 Hz) pada drum dan gitar bass, midrange (250–4000 Hz) pada vokal, gitar, dan banyak instrumen melodi, dan treble (4000–20.000 Hz) pada simbal dan *high syth*.

MFCC adalah teknik ekstraksi fitur yang biasa digunakan dalam pemrosesan sinyal ucapan dan audio untuk mewakili karakteristik spektral sinyal ucapan [9]. MFCC terdiri dari skala mel dan analisis cepstral. Skala mel adalah skala nada yang secara persepsi lebih konsisten dengan pendengaran manusia dibandingkan skala frekuensi linier, dan analisis cepstral adalah teknik matematis yang digunakan untuk mewakili spektrum sinyal dengan cara yang menekankan informasi yang relevan secara kognitif [10]. Proses penghitungan MFCC melibatkan beberapa langkah yang berawal dari *Framing*, kemudian *Windowing*, *Fast Fourier*

Transform (FFT), Mel-Filtering, Logarithm, dan berakhir dengan Discrete Cosine Transform (DCT) [11].

Filterbank Energies, juga dikenal sebagai filterbank outputs atau filterbank coefficients, adalah sekumpulan fitur yang biasa digunakan dalam tugas pemrosesan sinyal, khususnya dalam menganalisis sinyal audio [12]. Proses penghitungan Filterbank Energies melibatkan beberapa langkah yang berawal dari Framing, kemudian Windowing, Fast Fourier Transform (FFT), Mel-Filtering, dan berakhir dengan Logarithm [13].

Hidden Markov Model merupakan sebuah model statistik dari sebuah sistem yang diasumsikan sebuah proses Markov dengan adanya parameter yang tak diketahui, lalu akan ditentukan parameter-parameter tersembunyi (hidden) dari parameter-parameter yang dapat diamati [14]. Setiap kondisi memiliki distribusi kemungkinan disetiap output yang berbeda. HMM adalah metode yang ampuh untuk memodelkan data sekuensial yang struktur dasarnya tidak dapat diamati secara langsung. Komponen dari HMM terdiri dari States, Observations, Transitions Probabilities, dan Emissions Probabilities [15], [16].

Hidden Markov Model dapat didefinisikan sepenuhnya dengan jumlah keadaan tersembunyi  $n$ , matriks probabilitas transisi keadaan statis  $P$ , distribusi probabilitas observasi  $B$ , dan distribusi keadaan awal  $\pi$  [15]. Matriks  $P$ ,  $B$  dan  $\pi$  didefinisikan sebagai:

$$\begin{aligned} P &= \{p_{ij}\} & p_{ij} &= P(X_{t+1} = j | X_t = i) & i, j &\in \{1, 2, \dots, n\} \\ B &= \{b_i(k)\} & b_i(k) &= P(o_t = v_k | X_t = i) & v_k, k &\in \{1, \dots, K\} \\ \pi &= \{\pi_j\} & \pi_i &= P(X_0 = i) \end{aligned} \quad (1)$$

Dimana  $n$  adalah jumlah keadaan dalam model,  $P = \{1, \dots, n\}$  adalah ruang kemungkinan keadaan, dan  $V_k, k \in \{1, \dots, K\}$  adalah ruang kemungkinan hasil. Untuk ruang observasi berkelanjutan, kemungkinan hasil ditentukan oleh distribusi probabilitas yang diberikan sesuai dengan keadaan.

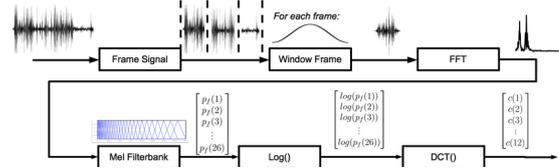
### III. METODE

Data yang digunakan diambil dari dataset yang diperoleh dari Kaggle bernama GTZAN Dataset. GTZAN Dataset merupakan kumpulan data yang banyak digunakan di bidang klasifikasi genre musik dan analisis audio. Data tersebut terdiri dari 1000 file yang terbagi menjadi 10 genre yang terdiri dari blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, dan rock. Tiap genre memiliki 100 musik dengan data berupa format (.wav) dengan panjang durasi 30 detik.

Pre-processing dalam penelitian ini merupakan proses mengubah sinyal audio menjadi sebuah format yang bisa dibaca oleh machine learning, dimana dalam penelitian ini menggunakan Mel-Frequency Cepstral Coefficients (MFCC) dan Filterbank Energies sebagai metode pre-processing yang digunakan.

MFCC merupakan fitur pemrosesan sinyal audio yang menangkap karakteristik spektral sinyal suara dengan cara yang selaras dengan bagaimana telinga manusia merasakan

nada dan frekuensi. Berikut ini adalah diagram alur dari proses MFCC [17]:



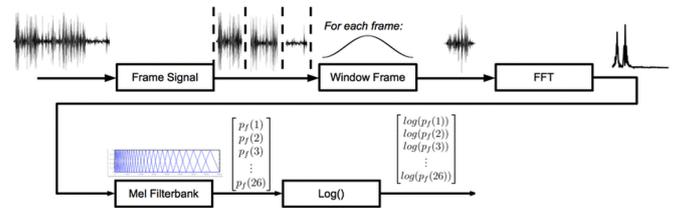
Gbr 1. Diagram alur proses MFCC

Langkah dari proses MFCC terdiri dari: membagi sinyal audio menjadi beberapa frame untuk menangkap segmen pendek audio; menerapkan windowing untuk mengurangi efek tepi di setiap frame; menerapkan Fast Fourier Transform (FFT) untuk mengubah frame ke dalam domain frekuensi; mengaplikasikan mel filterbank ke spektrum daya; mengubah hasil mel ke dalam bentuk logaritma; dan dihitung menggunakan Discrete Cosine Transform (DCT) untuk mendapatkan MFCC.

Dalam program pre-processing yang dijalankan, data audio diambil dari GTZAN dataset dengan parameter yang digunakan seperti sampling rate 22050 Hz, sinyal audio disegmentasi menjadi 25 ms/frames dengan overlap 10 ms, proses windowing menggunakan Hamming window, serta ukuran FFT yang ditetapkan ke 511. Fitur diekstraksi menggunakan filterbank 26 mel untuk kalkulasi energi, dengan 13 koefisien pertama dipertahankan setelah menerapkan Discrete Cosine Transform (DCT) ke energi filterbank berskala logaritma. Semua fitur dinormalisasi ke rata-rata nol dan varians unit untuk memastikan konsistensi di seluruh kumpulan data.

Hasil perhitungan dari MFCC adalah sekumpulan koefisien yang mewakili spektrum daya jangka pendek dari sinyal audio yang bisa digunakan, dimana setiap koefisien memberikan representasi numerik yang ringkas dari pita frekuensi tertentu yang menangkap karakteristik penting suara dalam suatu frame.

Filterbank Energies merupakan fitur yang berasal dari frekuensi sinyal audio yang menggambarkan besaran energi di berbagai pita frekuensi. Berikut ini adalah diagram alur dari proses Filterbank Energies [17]:



Gbr 2. Diagram alur proses Filterbank Energies

Langkah dari proses Filterbank Energies terdiri dari: membagi sinyal audio menjadi beberapa frame untuk menangkap segmen pendek audio; menerapkan windowing untuk mengurangi efek tepi di setiap frame; menerapkan Fast Fourier Transform (FFT) untuk mengubah frame ke dalam domain frekuensi; mengaplikasikan mel filterbank ke spektrum daya; mengubah hasil mel ke dalam bentuk logaritma untuk mendapatkan Filterbank Energies.

Dalam program pre-processing yang dijalankan, data audio diambil dari GTZAN dataset dengan parameter yang digunakan seperti sampling rate 22050 Hz, sinyal audio disegmentasi menjadi 25 ms/frames dengan overlap 10 ms, proses windowing menggunakan Hamming window, serta ukuran FFT yang ditetapkan ke 551. Fitur diekstraksi menggunakan filterbank 26 mel untuk kalkulasi energi, kemudian diubah menjadi bentuk skala logaritma. Semua fitur dinormalisasi ke rata-rata nol dan varians unit untuk memastikan konsistensi di seluruh kumpulan data.

Hasil perhitungan dari Filterbank Energies adalah sekumpulan nilai energi yang mewakili distribusi energi pada pita frekuensi berbeda dari sinyal audio. Nilai energi ini memberikan gambaran terperinci tentang bagaimana suara dipersepsikan dalam hal intensitasnya di berbagai frekuensi, yang sangat mirip dengan cara sistem pendengaran manusia memproses suara.

Data yang diproses oleh MFCC dan *Filterbank Energies* merupakan data musik yang diubah menjadi angka yang bisa dibaca oleh HMM, dimana HMM sendiri mencari pola tersembunyi dari data yang dipelajari.

Dalam melakukan pelatihan Hidden Markov Model (HMM), peneliti menggunakan aplikasi berbasis Python yang akan berjalan sesuai dengan formula HMM yang telah dijelaskan sebelumnya, dengan rincian sebagai berikut [15]:

- $P = \{p_{ij}\} p_{ij} = P(X_{t+1} = j | X_t = i) \quad i, j \in \{1, 2, \dots, n\}$ : Probabilitas transisi dari keadaan  $i$  ke keadaan  $j$ , dimana status  $i$  dan  $j$  dapat mewakili status "tersembunyi" yang berbeda sesuai dengan pola atau fitur musik (misalnya, audio yang termasuk dalam genre tertentu).
- $B = \{b_i(k)\} \quad b_i(k) = P(o_t = v_k | X_t = i) \quad v_k, k \in \{1, \dots, K\}$ : Probabilitas yang mengobservasi  $v_k$  (sebuah *feature vector*, seperti MFCC atau *filterbank energies*) dimana sistem berada dalam keadaan  $i$ , dengan  $O_t$  merupakan *feature vector* (untuk mewakili karakteristik numerik atau simbolik, yang disebut *features*, dari suatu objek dengan cara matematis dan mudah dianalisis) yang diobservasi pada waktu  $t$ , dan  $X_t$  adalah hidden state pada waktu  $t$ . Bagian ini memodelkan kemungkinan fitur audio berdasarkan status pada waktu  $t$ .
- $\pi = \{\pi_i\} \quad \pi_i = P(X_0 = i)$ : Probabilitas awal pada keadaan  $i$ , mendefinisikan distribusi status awal saat model mulai memproses urutan baru.

Ekstraksi fitur (MFCC atau *filterbank energies*) menyediakan data teramati ( $O_t$ ) yang digunakan dalam probabilitas emisi dan HMM memodelkan sifat sekuensial data audio, menangkap ketergantungan temporal antar *feature vector*. Untuk setiap genre, HMM dilatih secara terpisah, dan model dengan kemungkinan tertinggi untuk lagu tertentu menentukan genrenya.

#### IV. HASIL DAN PEMBAHASAN

Data yang digunakan diambil dari dataset bernama GTZAN Dataset yang terdiri dari 1000 file yang terbagi menjadi 10

genre yang terdiri dari *blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, dan rock*, dengan 100 file musik masing-masing dengan datafile berupa format (.wav) berdurasi 30 detik untuk tiap filenya.

Data audio yang sudah diambil dijalankan kedalam sebuah program yang menjalankan modul MFCC dan *Filterbank Energies*. Berikut ini adalah baris kode yang digunakan untuk memproses data audio menggunakan metode MFCC dan *Filterbank Energies*:

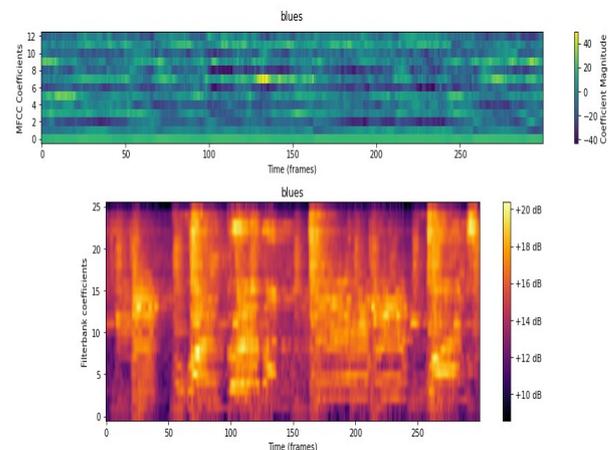
```
sampling_rate, audio_data = wavfile.read("Data/genres_original/blues/blues.00000.wav")
mfcc_features = mfcc(audio_data, samplerate=sampling_rate, nfft=551)
```

Gbr 3. Program metode MFCC

```
sampling_rate, audio_data = wavfile.read("Data/genres_original/blues/blues.00000.wav")
filterbank_features = logfbank(audio_data, samplerate=sampling_rate, nfft=551)
```

Gbr 4. Program metode Filterbank Energies

Kedua gambar diatas merupakan kode program yang menggunakan *python\_speech\_features* sebagai pustaka dari 'mfcc' dan 'logfbank', dengan keluaran dari berupa array dengan panjang koefisien sebesar 13 untuk MFCC dan 26 untuk filterbank.



Gbr 5. Output MFCC (atas) dan Filterbank (bawah) dari genre 'blues'

Gambar diatas merupakan keluaran dari MFCC dan Filterbank, dengan keterangan sebagai berikut:

- Plot MFCC menampilkan 13 koefisien selama 300 *frames*. Sumbu horizontal menunjukkan waktu dalam *frames*, sedangkan sumbu vertikal menunjukkan koefisien MFCC.
- Plot *Filterbank Energies* menggambarkan 26 koefisien selama 300 *frames*. Sumbu horizontal menggambarkan *frames*, dan sumbu vertikal berupa koefisien filterbank.
- Pada detail frekuensi, visualisasi MFCC memampatkan informasi spektral menjadi 13 koefisien, yang menghasilkan representasi audio yang lebih abstrak. Sedangkan visualisasi *Filterbank Energies* menangkap detail frekuensi yang lebih banyak di 26 koefisiennya, menyoroti pola energi rendah dan energi tinggi yang berbeda.
- Untuk klasifikasi, representasi MFCC yang lebih halus menguntungkan untuk model pembelajaran mesin

seperti HMM, karena mengurangi redundansi data dan menyederhanakan pola untuk pengenalan. Sedangkan *Filterbank Energies* memberikan representasi yang lebih rinci, berpotensi membantu model yang mengandalkan informasi spektral mentah untuk membedakan fitur-fitur spesifik genre yang halus.

Setelah proses *Pre-processing*, kedua data tersebut dijalankan ke program pelatihan HMM masing-masing. Proses pelatihan HMM bertujuan untuk mempelajari pola temporal dari vektor fitur (MFCC dan energi filterbank) yang diekstrak dari klip audio. Setiap HMM dilatih untuk mewakili genre musik tertentu.

```
class HMMTrainer(object):
    def __init__(self, model_name='GaussianHMM', n_components=10, cov_type='diag', n_iter=1000):
        self.model_name = model_name
        self.n_components = n_components
        self.cov_type = cov_type
        self.n_iter = n_iter
        self.models = []
        if self.model_name == 'GaussianHMM':
            self.model = hmm.GaussianHMM(n_components=self.n_components, covariance_type=self.cov_type, n_iter=self.n_iter)
        else:
            raise TypeError('Invalid model type')

    def train(self, X):
        np.seterr(all='ignore')
        self.models.append(self.model.fit(X))
        # Run the model on input data
    def get_score(self, input_data):
        return self.model.score(input_data)
```

Gbr 6. Baris kode pelatihan HMM

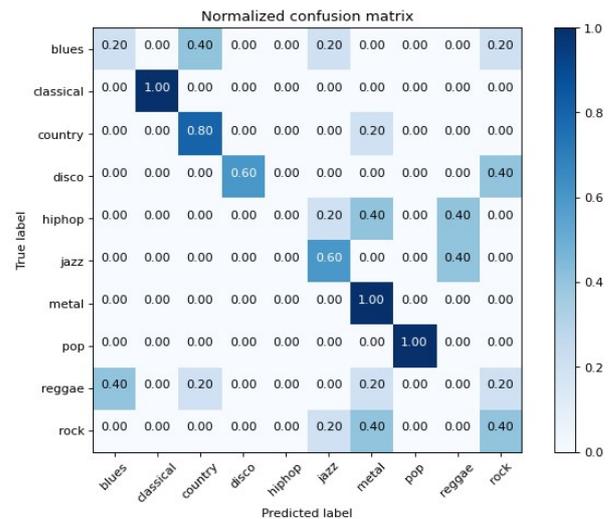
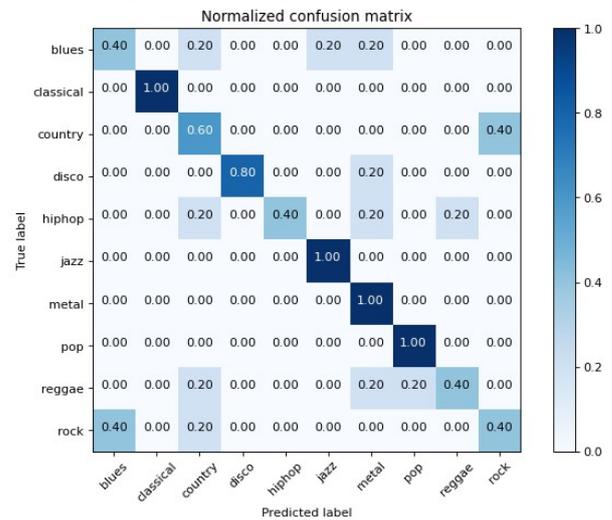
Pelatihan HMM menggunakan model 'GaussianHMM' dengan 10 keadaan tersembunyi, tipe kovarians diagonal, dan menjalankan 1000 iterasi, dengan penjelasan sebagai berikut:

- GaussianHMM merupakan tipe dari HMM yang membuat fitur yang dapat diamati dimodelkan menggunakan distribusi Gaussian. Distribusi Gaussian sendiri merupakan distribusi probabilitas berkelanjutan untuk variabel acak bernilai riil. Hasilnya, setiap keadaan tersembunyi menghasilkan fitur yang mengikuti distribusi normal (kurva lonceng).
- 'n\_components' mengacu pada keadaan tersembunyi di HMM, dimana pada kasus ini terdapat 10 komponen tersembunyi, dan setiap komponen mewakili pola atau karakteristik dasar yang berbeda dalam data audio yang akan coba ditangkap oleh model.
- Kovarians diagonal merupakan jenis matriks kovarians yang digunakan dalam distribusi Gaussian. Menggunakan tipe kovarian ini, berarti matriks kovarians untuk setiap distribusi Gaussian berbentuk diagonal, yang menyiratkan bahwa fitur-fiturnya diasumsikan tidak berkorelasi. Matriks kovarians hanya akan berkorelasi pada diagonalnya.
- 'n\_iter' mengacu pada iterasi untuk pelatihan model yang menyatakan nilai maksimum HMM melakukan pembaharuan parameter, dimana pada kasus ini model akan melakukan 1000 kali iterasi atau berhenti lebih awal jika solusi optimal sudah ditemukan.

Hasil dari pelatihan HMM ini adalah sebuah set parameter model terlatih yang telah dioptimalkan untuk mewakili properti statistik data pelatihan dengan lebih baik. Parameter yang terdiri dari probabilitas keadaan awal, probabilitas

transisi, dan probabilitas emisi, membentuk hasil HMM yang digunakan untuk tugas pemodelan sekuens, seperti klasifikasi atau *recognition*, selama fase pengujian.

Hasil dari MFCC dan *Filterbank Energies* dijalankan kedalam program HMM, dengan hasil keluaran berupa nilai prediksi yang akan dibandingkan dengan nilai asli dan ditampilkan kedalam *normalized confusion matrix*. Perbandingan data training-test yang digunakan terdiri dari rasio 95-5 sampai 50-50 dengan interval 5.



Gbr 7. Output MFCC (atas) dan Filterbank (bawah) dengan perbandingan data training-test '95-5'

TABEL I  
DATA TRAINING-TEST '95-5' DARI DIAGONAL  
HASIL NORMALIZED CONFUSION MATRIX

Genre	MFCC	Filterbank
Blues	0,40	0,20
Classical	1,00	1,00
Country	0,60	0,80
Disco	0,80	0,60
Hiphop	0,40	0,00

Jazz	1,00	0,60
Metal	1,00	1,00
Pop	1,00	1,00
Reggae	0,40	0,00
Rock	0,40	0,40

TABEL III  
METRIK EVALUASI DARI DATA TRAINING-TEST '95-5'

Genre	MFCC			Filterbank		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Blues	0,50	0,40	0,44	0,33	0,20	0,25
Classical	1,00	1,00	1,00	1,00	1,00	1,00
Country	0,43	0,60	0,50	0,57	0,80	0,67
Disco	1,00	0,80	0,89	1,00	0,60	0,75
Hiphop	1,00	0,40	0,57	0,00	0,00	0,00
Jazz	0,83	1,00	0,91	0,50	0,60	0,55
Metal	0,56	1,00	0,71	0,45	1,00	0,62
Pop	0,83	1,00	0,91	1,00	1,00	1,00
Reggae	0,67	0,40	0,50	0,00	0,00	0,00
Rock	0,50	0,40	0,44	0,33	0,40	0,36

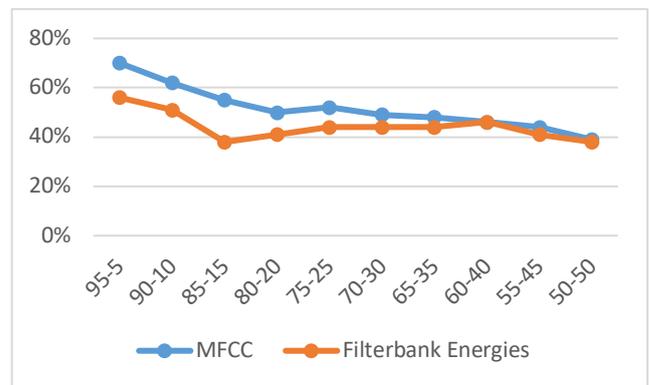
Dari tabel data training-test '95-5' menunjukkan hasil tertinggi dalam mengenali genre musik menggunakan MFCC didapati pada genre classical, jazz, metal, dan pop dengan presentase 100%. Untuk hasil tertinggi dalam mengenali genre musik menggunakan *Filterbank Energies* didapati pada genre classical, metal, dan pop dengan presentase 100%. Sementara hasil terendah dalam mengenali genre musik menggunakan MFCC didapati pada genre blues, hiphop, reggae, dan rock dengan presentase 40%. Untuk hasil terendah dalam mengenali genre musik menggunakan *Filterbank Energies* didapati pada genre hiphop dan reggae dengan presentase 0%. Dari data tersebut, genre classical, jazz, metal, dan pop lebih mudah diidentifikasi, serta genre hiphop dan reggae lebih sulit diidentifikasi berdasarkan penggunaan kedua metode yang digunakan.

Tabel metrik evaluasi didapat dari mengolah data *normalized confusion matrix* training-test '95-5'. Berdasarkan hasil *F1-Score* yang merupakan *harmonic mean* dari *precision* dan *recall*, hasil terbaik didapati pada genre classical dengan nilai 1,00 untuk MFCC, dan genre classical serta pop dengan nilai 1,00 untuk *Filterbank Energies*. Sementara hasil terburuk didapati pada genre blues dan rock dengan nilai 0,44 untuk MFCC, dan genre hiphop serta reggae dengan nilai 0,00 untuk *Filterbank Energies*.

Setelah melewati proses pelatihan, hasil dari pelatihan tersebut digunakan untuk melakukan pengujian terhadap data audio yang dipisahkan dari data untuk pelatihan. Rasio yang digunakan untuk pelatihan dan pengujian yaitu dari 95-5 sampai 50-50, dengan rincian hasil seperti tabel berikut.

TABEL IIIII  
HASIL PENGUJIAN HMM MENGGUNAKAN MFCC DAN *FILTERBANK ENERGIES*

ratio	mfcc		filterbank	
	accuracy	time	accuracy	time
95-5	70%	4.11.43	56%	1.43.16
90-10	62%	5.30.16	51%	1.40.21
85-15	55%	4.54.39	38%	1.28.23
80-20	50%	3.3.27	41%	1.15.16
75-25	52%	3.17.39	44%	1.8.45
70-30	49%	2.53.58	44%	54.25
65-35	48%	2.44.54	44%	55.17
60-40	46%	1.55.22	46%	1.7.55
55-45	44%	2.19.22	41%	1.2.15
50-50	39%	1.52.48	38%	42.23



Gbr 8. Grafik hasil pengujian HMM

Dari gambar di atas, pengujian HMM menggunakan MFCC dan *Filterbank Energies* terjadi penurunan tren akurasi seiring rasio mendekati 50-50. Pada tabel di atas menunjukkan hasil pengujian pada rasio 95-5 dari MFCC mendapatkan 70% akurasi dalam waktu 4 jam 11 menit 43 detik, sementara *Filterbank Energies* mendapatkan 56% akurasi dalam waktu 1 jam 43 menit 16 detik.

## V. PENUTUP

Dari penelitian yang telah dilakukan untuk melakukan identifikasi genre musik menggunakan metode *Mel-Frequency Cepstral Coefficients* dan *Filterbank Energies* dalam proses *Hidden Markov Model*, kesimpulan yang didapatkan adalah:

1. Hasil implementasi dari metode *Mel-Frequency Cepstral Coefficients* dalam proses *Hidden Markov Model* untuk melakukan identifikasi genre musik, yaitu berupa data dalam *normalized confusion matrix* yang dilakukan menggunakan 10 parameter training-test. Pada data yang diambil, diperoleh bahwa genre pop mendapatkan nilai tertinggi di semua parameter training-test yang dilakukan. Sedangkan untuk nilai terendahnya didapati berbeda-beda di semua parameter training-test yang dilakukan, dengan genre blues yang

sering muncul sebagai nilai terendah dari semua parameter training-test yang dilakukan.

2. Hasil implementasi dari metode *Filterbank Energies* dalam proses *Hidden Markov Model* untuk melakukan identifikasi genre musik, yaitu berupa data dalam *normalized confusion matrix* yang dilakukan menggunakan 10 parameter training-test. Pada data yang diambil, diperoleh bahwa genre pop mendapatkan nilai tertinggi di hampir semua parameter training-test yang dilakukan. Sedangkan untuk nilai terendahnya didapati berbeda-beda di semua parameter training-test yang dilakukan, dengan genre hiphop yang sering muncul sebagai nilai terendah dari semua parameter training-test yang dilakukan.
3. Perbandingan hasil akurasi yang didapatkan dari metode *Mel-Frequency Cepstral Coefficients* dan *Filterbank Energies* dalam proses *Hidden Markov Model* untuk melakukan identifikasi genre musik, diperoleh setelah menghitung metrik evaluasi akurasi yang didapatkan dari hasil *normalized confusion matrix* metode *Mel-Frequency Cepstral Coefficients* dan *Filterbank Energies* dalam proses *Hidden Markov Model*. Pada data yang diambil, diperoleh bahwa metode *Mel-Frequency Cepstral Coefficients* mendapatkan nilai akurasi yang lebih tinggi daripada metode *Filterbank Energies* di hampir semua parameter training-test yang dilakukan. Untuk hasil akurasi terbaik yang didapat dalam melakukan identifikasi genre musik ini, didapatkan dari parameter training-test (95-5) dengan hasil akurasi 70% dari penggunaan metode *Mel-Frequency Cepstral Coefficients* dan 56% dari penggunaan metode *Filterbank Energies*.

#### REFERENSI

- [1] G. Roy, William, dkk. (2010). "What Is Sociological about Music?". *Annu. Rev. Sociol.* 2010. 36:183–203
- [2] Lerch, Alexander. 2022. "An Introduction to Audio Content Analysis: Music Information Retrieval Tasks and Applications". John Wiley & Sons.
- [3] Sebastian D.V. (2017). "Classification Of Musical Genres Using Hidden Markov Models". Faculty of Science, Centre for Mathematical Sciences, Mathematical Statistics, Master's thesis 2017:E20.
- [4] Liu, Xinru. (2019). "Hidden Markov Models for Music Classification". Faculty of Wheaton College. Norton, Massachusetts.
- [5] Nadeu, Climent dkk. (2001). "Time and frequency filtering of filterbank energies for robust HMM speech recognition". Elsevier Science B.V. *Speech Communication* 34 (2001): 93–114.
- [6] Swarga, L.T. Santoso, Joan. Setyati, Endang. (2022). "Pengenalan Lirik Lagu Otomatis Pada Video Lagu Indonesia Menggunakan Hidden Markov Model Yang Dilengkapi Music Removal". *Journal of Information System, Graphics, Hospitality and Technology*, Vol. 04, No. 02.
- [7] Sanneh, Kelefa. (2021). "Major Labels: A History of Popular Music in Seven Genres". Penguin Press.
- [8] Välimäki, Vesa. Reiss, Joshua D. (2016). "All About Audio Equalization: Solutions and Frontiers". Basel, Switzerland. MDPI *Applied Sciences: Volume 6, Issue 5*.
- [9] Shao, Xi. Xu, Changsheng. Kankanhalli, M.S. (2004). "Unsupervised Classification of Music Genre Using Hidden Markov Model". *IEEE International Conference on Multimedia and Expo (ICME)*.
- [10] Nasr, M.A. dkk. (2018). "Speaker identification based on normalized pitch frequency and Mel Frequency Cepstral Coefficients". *International Journal of Speech Technology*
- [11] Manju D.P., Rajendra D.K. (2020). "Convolution neural network based automatic speech emotion recognition using Mel-frequency Cepstrum coefficients". *Multimedia Tools and Applications* (2021) 80:15563–15587
- [12] Hardik B.S. dkk. (2017). "Unsupervised Filterbank Learning Using Convolutional Restricted Boltzmann Machine for Environmental Sound Classification". *INTERSPEECH 2017*. Stockholm, Sweden
- [13] Rishabh N.T. dkk. (2017). "Novel Phase Encoded Mel Filterbank Energies for Environmental Sound Classification". *Speech Research Lab, Dhirubhai Ambani Institute of Information and Communication Technology, (DA-IICT), Gandhinagar, India*
- [14] M. G. J. Harry Khesa S, dkk. (2016). "Perbandingan Metode Hidden Markov Model Dan Vector Quantization Untuk Aplikasi Identifikasi Suara". *E-Journal SPEKTRUM*, Vol. 3, No. 2.
- [15] Awad, M., Khanna, R. (2015). "Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers". Apress, Berkeley, CA. *Efficient Learning Machines* (2015): 81-104
- [16] Westhead, D.R. Vijayabaskar, M.S. (2017). "Hidden Markov Models: Methods and Protocols". Springer Nature, New York. U.S.A
- [17] Nahar, K.M.O. dkk. 2020. "Handicapped Wheelchair Movements Using Discrete Arabic Command Recognition". *Scientific Journal of King Faisal University (Basic and Applied Sciences)* Vol.21.