

BAB I

PENDAHULUAN

1.1. Latar Belakang

Musik merupakan salah satu hiburan yang paling mudah didapatkan pada zaman sekarang. Dengan adanya perkembangan teknologi internet sekarang yang semakin maju, sehingga memudahkan untuk men-*download* digital musik dari internet. Semua golongan masyarakat dapat menikmati musik dengan menggunakan berbagai teknologi pemutar musik yang dapat dengan mudah dijumpai. Sehingga tidak heran jika satu orang saja mempunyai banyak koleksi musik dalam bentuk lagu-lagu .

Namun demikian dalam mengoleksi lagu-lagu, sering kali para penikmat musik kesulitan dalam memilih dan memilah lagu sesuai dengan genre lagu yang mereka sukai. Pada umumnya pengelompokan lagu hanya dilakukan dengan mendengar musik satu per satu atau melihat dari meta data yang terdapat di dalam lagu. Namun hal ini tidak efisien karena harus mendengarkan lagu satu persatu. Selain itu, jika ingin melihat dari meta data, banyak lagu yang tidak mempunyai meta data yang menjelaskan tipe genre lagu tersebut sehingga menyusahkan para kolektor lagu memilih lagu-lagu tersebut.

Telah banyak peneliti yang telah meneliti terkait perihal klasifikasi musik berdasarkan genre dari lagu tersebut. Salah satunya penelitian yang telah dilakukan oleh Timotius et al, (2009) yang meneliti tentang klasifikasi genre musik dengan menggunakan fitur musik dan *neural network*. Klasifikasi genre musik dengan menggunakan 40 data uji dari 4 jenis genre dengan mengekstraksi *timbre features* dengan menggunakan MFCC (*Mel Frequency Cepstral Coefficient*) dan *rhythm features* serta *neural network*. Hasil penelitian menunjukkan tingkat keakuratan 80%. Penelitian yang lain salah satunya dilakukan oleh Chang-Hsing Lee, et all (2007) tentang *automatic music genre classification using modulation spectral contrast feature*. Hasil penelitian yang telah dilakukan manunjukkan tingkat keakuratan 84,03%. Darma Putra dan Adi Resmawan juga melakukan pengenalan suara dengan

menggunakan MFCC. Berdasarkan penelitiannya yang berjudul verifikasi biometrika suara menggunakan metode MFCC dan DTW, disimpulkan bahwa MFCC merupakan metode yang baik dalam pengenalan ekstraksi fitur dalam pengenalan suara.

Self Organizing Maps merupakan salah satu metode yang digunakan untuk proses *clustering*. Metode ini sudah banyak digunakan dalam penelitian, salah satunya seperti yang dilakukan oleh Wahyumianto, et al (2009) tentang Identifikasi Tumbuhan Berdasarkan *Minutiae* Tulang Daun Menggunakan SOM Kohonen yang menyimpulkan bahwa berdasarkan penelitian tersebut metode Kohonen dapat mengenali jenis daun tumbuhan dengan tingkat akurasi di atas 80%.

Dari penelitian sebelumnya didapatkan kesimpulan bahwa klasifikasi genre musik dapat dilakukan dengan menggunakan MFCC dan *neural network*. MFCC (*Mel Frequency Cepstral Coefficient*) dapat mengenali fitur musik mendekati pendengaran manusia. Sedangkan metode *neural network* yang akan digunakan adalah menggunakan model SOM (*Self Organizing Maps*).

1.2. Rumusan Masalah

Berdasarkan dari latar belakang yang telah dijelaskan, maka rumusan masalah penelitian ini adalah bagaimana mengklaster genre sebuah musik dari *database* yang ada dengan menggunakan ekstraksi fitur MFCC dan SOM.

1.3. Batasan Masalah

Agar pembahasan sesuai dengan tujuan penulisan, maka diberikan batasan masalah sebagai berikut.

1. Penelitian ini hanya untuk mempelajari pengenalan musik dengan MFCC dan SOM.
2. Genre yang digunakan untuk proses klaster hanya 3 yaitu pop, rock, dan dangdut.
3. Objek penelitian ini adalah lagu *.mp3 yang dikonversi dalam format *.wav.
4. Panjang lagu yang diuji dibatasi hanya 5,10, dan 20 detik yang diambil pada intro (setelah 5 detik), setelah menit ke-1, dan setelah menit ke-2 lagu.

5. Fitur seleksi yang digunakan adalah *mean*, *max*, *min*, dan *median* dari hasil MFCC.

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah berhasil mengimplementasikan MFCC (*Mel Frequency Cepstral Coefficient*) dan SOM (*Self Organizing Maps*) untuk kluster lagu serta dapat menganalisis fitur seleksi dan durasi lagu yang lebih tepat.

1.5. Sistematika Penulisan

Berikut merupakan rencana susunan sistematika penulisan laporan Tugas Akhir yang akan dibuat :

BAB I Pendahuluan

Bab ini berisi penjelasan mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan dari Tugas Akhir yang dibuat.

BAB II Landasan Teori

Bab ini membahas tentang teori-teori yang berhubungan dengan tugas akhir ini yaitu tentang genre dan pengenalan genre lagu dengan ekstraksi fitur, fitur seleksi, *clustering*, hingga penghitungan akurasi.

BAB III Metodologi Penelitian

Bab ini membahas langkah-langkah yang dilaksanakan dalam proses penelitian, yaitu penelitian pendahuluan, pengumpulan data, analisa, implementasi, pengujian, kesimpulan dan saran.

BAB IV Analisa Metode MFCC Dan SOM

Bab ini berisi pembahasan mengenai analisa data masukan, proses pengenalan fitur ekstraksi dengan MFCC, fitur seleksi, dan pengenalan pola dengan SOM.

BAB V Pengujian

Bab ini berisi tentang data masukan, skenario pengujian, hasil pengujian, dan analisa hasil pengujian.

BAB VI Kesimpulan dan Saran

Bagian ini berisi kesimpulan yang dihasilkan dari pembahasan tentang analisa dan implementasi MFCC (*Mel Frequency Cepstral Coefficient*) dan SOM (*Self Organizing Maps*) untuk klaster musik.

BAB II

LANDASAN TEORI

2.1. Genre

Menurut Kamus Besar Bahasa Indonesia, genre adalah jenis, tipe, atau kelompok sastra atas dasar bentuknya. Jadi genre musik merupakan pengelompokan musik berdasarkan kemiripan jenis, tipe, ataupun tema musik tersebut.

Jenis – jenis pembagian genre musik yang sering didengar, antara lain:

1. Musik populer atau Musik pop

Jenis aliran musik yang banyak didengar oleh kalangan luas yang biasanya diperjualbelikan. Genre pop mempunyai irama yang ringan dan berperan sebagai media hiburan dan pertunjukan (Errie 2009). Genre ini pertama kali berkembang di Amerika Serikat pada tahun 1920 yang pertama kali dibuat berdasarkan penemuan Thomas Edison, dibedakan dengan Musik Klasik, Musik Jazz, Musik Tradisional, Musik Blues, kemudian juga berkembang ke negara-negara lain sedunia.

2. Rock

Genre musik mulai dikenal pada pertengahan tahun 50-an. Musik rock diadaptasi dari berbagai musik lainnya, termasuk musik rakyat (*folk music*), jazz dan musik klasik, blues dan country. Bunyi khas dari musik rock adalah penggunaan gitar listrik atau gitar akustik, dan penggunaan back beat yang pada bagian ritme dengan gitar bass dan drum, serta organ, piano atau sejak 70-an, synthesizer. Pada akhir tahun 60-an dan awal 70-an, musik rock berkembang menjadi beberapa jenis. Yang bercampur dengan musik folk (musik daerah di amerika) menjadi *folk rock*, dengan blues menjadi blues-rock dan dengan jazz, menjadi jazz-rock *fusion*. Pada tahun 70an, rock menggabungkan pengaruh dari *soul*, *funk*, dan musik latin. Juga di tahun 70an, rock berkembang menjadi berbagai subgenre (sub-kategori) seperti soft rock, glam rock, heavy metal, hard rock, progressive rock, dan punk rock. Sub kategori rock yang mencuat ditahun 80an termasuk New Wave, hardcore punk dan alternative rock. Pada tahun 90an terdapat grunge, Britpop, indie rock dan nu metal. Ciri khas dari genre ini adalah ritme yang tidak konvensional (bukan 4/4 atau sinkopasi), penguasaan alat music yang mahir dengan permainan solo yang rumit dan lagu-lagu yang berdurasi panjang. (Harry 2007)

3. Dangdut

Musik dangdut merupakan salah satu seni music yang berkembang di Indonesia. Bentuk musik ini berakar dari musik melayu pada tahun 1940-an. Dalam proses evolusinya dipengaruhi unsur musik india (tabla) dan arab (pada cengkok dan harmonisasi). Penyebutan nama dangdut berdasarkan ciri khas suara permainan tabla (dalam dunia dangdut disebut gendang) yang khas dan didominasi oleh suara dang dan ndut. (Harry 2007)

2.2 Pengenalan Genre lagu

Untuk mengenal masing-masing karakteristik genre musik diperlukan beberapa langkah, yakni:

2.2.1 Ekstraksi Fitur

Ekstraksi fitur berfungsi untuk mengenal karakteristik dari suatu musik atau lagu. Setiap genre lagu menggunakan alat musik yang dominan yang berbeda-beda (Timotius 2009). Contohnya, musik pop lebih dominan menggunakan gitar, drum dan keyboard. Sedangkan musik rock lebih dominan menggunakan gitar listrik, serta dangdut yang lebih dominan menggunakan gendangnya.

Untuk mengenal karakteristik ini maka diperlukan sebuah metode, yakni MFCC (*Mel Frequency Cepstral Coefficient*) (Tzanetakis 2001)

MFCC mempunyai kemampuan seperti pendengaran manusia, jadi MFCC dapat mengadaptasi pendengaran manusia dalam bentuk digital. Selain itu, penggunaannya adalah untuk mengekstrak data sinyal suara.

Beberapa keunggulan dari MFCC, antara lain (Putra 2011):

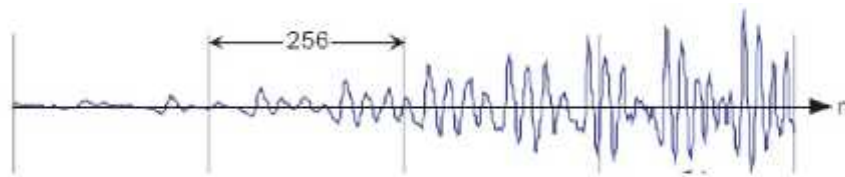
- a. Mampu untuk menangkap karakteristik suara yang sangat penting bagi pengenalan suara atau dapat menangkap informasi yang penting dalam sinyal suara.
- b. Menghasilkan data seminimal mungkin tanpa menghilangkan informasi penting yang ada di dalam sinyal.
- c. Mempunyai kemampuan seperti organ pendengaran suara dalam mengenali sinyal suara.

Tahapan proses *Mel Frequency Cepstral Coefficient* adalah sebagai berikut:

1. *Frame Blocking*, proses ini membagi sinyal audio yang masuk kedalam bentuk *frame*, dan setiap *frame* terdiri atas N-sample. Sinyal audio yang masuk merupakan sebuah *non-*

stationer signal. Sedangkan untuk menganalisa sinyal audio, maka sinyal yang masuk harus menjadi *stationer signal* sehingga perlu dibagi-bagi menjadi *frame-frame*.

Frame akan dibagi menjadi N *sample* dan jarak antar frame overlap adalah M dimana $M < N$. nilai N biasanya adalah 256 yang mewakili 30 ms dengan spasi 100. Panjang daerah *overlap* umumnya yang digunakan adalah 30% sampai 50% dari panjang frame yang berfungsi untuk menghindari hilangnya ciri atau karakteristik suara pada perpotongan frame. (Putra 2011)



Gambar 2.1 Frame Blocking

(sumber : Satrya 2010)

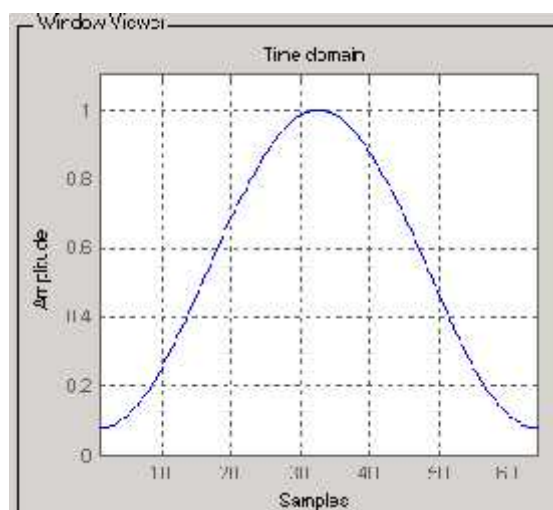
2. *Windowing*, proses ini dibutuhkan untuk meminimalisir diskontinuitas antar frame yang menyebabkan kehilangan informasi yang terdapat dalam suatu sinyal. (Satrya 2010)
Proses *windowing* yang dipakai adalah proses *Hamming Window* dengan persamaan :

$$w(n) = 0,54 - 0,46 \cos \left(\frac{2\pi n}{(N-1)} \right) \quad (2.1)$$

$w(n)$ = *windowing*

N = Jumlah *sample* pada masing-masing frame

$n = 0, 1, 2, 3, \dots, N-1$



Gambar 2.2 Hamming Window

(sumber : matlab help)

3. *Fast Fourier Transform* (FFT), yang berfungsi untuk mengubah domain waktu menjadi domain frekuensi (Sung 2008). FFT merupakan cara yang lebih efisien daripada DFT (*Discrete Fourier Transform*) (Satrya 2010). Berikut adalah persamaan dari *Fast Fourier Transform* :

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N} \quad 2.2$$

X_n = frekuensi

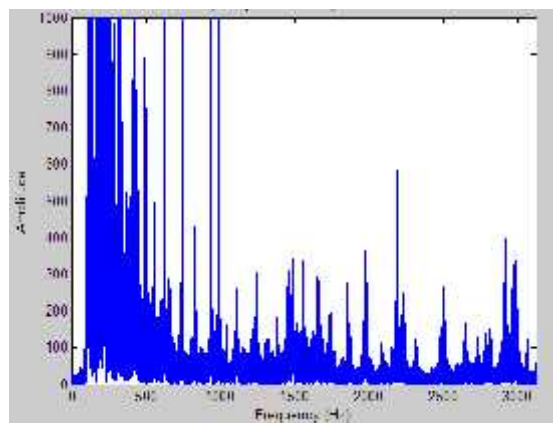
$k = 0, 1, 2, \dots, (N-1)$

N = jumlah *sample* pada masing-masing *frame*

j = bilangan imajiner ($\sqrt{-1}$)

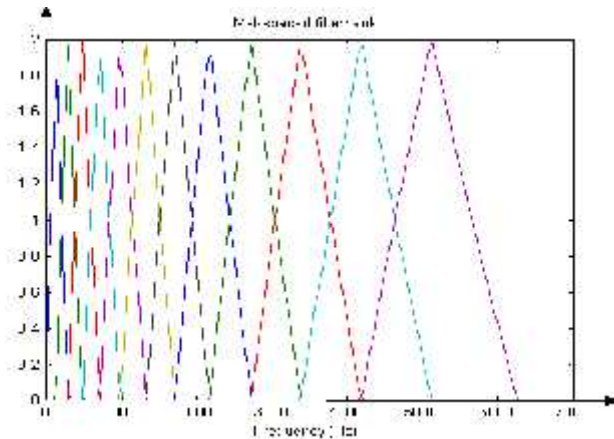
$n = 1, 2, 3, \dots (N-1)$

Hasil dari data yang telah melalui proses FFT, seperti gambar di bawah ini.



Gambar 2.3 Hasil FFT Dari Salah Satu Contoh Lagu

4. *Mel-frequency Wrapping*, merupakan selang frekuensi di bawah 1000 hz dan selang algoritmik di atas 1000 hz. Skala *mel* diperlukan karena pendengaran manusia tidak berupa skala linear, oleh karena itu, tinggi subjektif setiap nada bisa diukur dengan skala *mel*. (Zilvan 2011)



Gambar 2.4 Mel-frequency Wrapping

(sumber : <https://engineering.purdue.edu/~malcolm/interval/1998-010/>)

Terlihat pada gambar 2.4, mel filter bank adalah merupakan segitiga yang saling tumpang tindih. Sumbu y mewakili power (db) sinyal suara, dan sumbu x merupakan frekuensi (hz). (Inggih 2011)

Persamaan yang digunakan untuk menghitung skala *mel* untuk frekuensi dalam Hz adalah sebagai berikut.

$$Mel\ f = 2595 * \log_{10} \left(1 + \frac{f}{700} \right) \quad (2.3)$$

5. *Cepstrum*, merupakan langkah terakhir dalam proses MFCC. Pada tahap ini, akan menghasilkan *mel frequency cepstrum coefficient (MFCC)* yang merupakan hasil konversi log mel *cepstrum* ke dalam domain waktu (Tiwari 2010). Untuk mengubah ke dalam domain waktu, maka diperlukan DCT (*Discrete Cosine Transform*) dengan persamaan berikut.

$$C_n = \sum_{k=1}^k \log S_k \cos \left(n \left(k - \frac{1}{2} \right) * \frac{\pi}{k} \right) \quad (2.4)$$

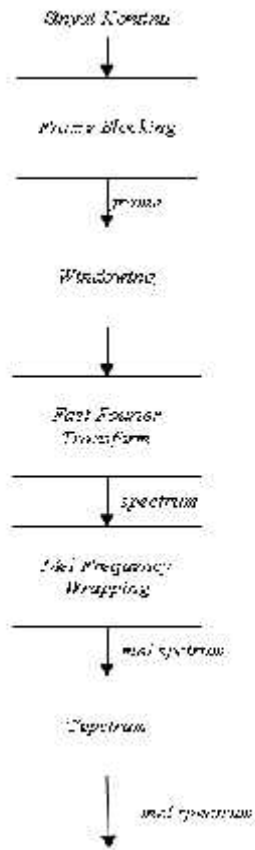
$n = 1, 2, \dots, k$

$C_n = MFCC$

$S_k =$ mel frekuensi

$k =$ jumlah coefficient

Jadi , tahapan dalam MFCC adalah seperti gambar di bawah ini.



Gambar 2.5 Tahapan-tahapan dalam MFCC
(sumber : Fansuri 2011)

2.2.2 Fitur Seleksi

Setelah melalui proses MFCC selanjutnya data tersebut akan dicari nilai max, mean, median, dan min-nya untuk selanjutnya data tersebut akan dikenali pola atau jenis kelompoknya dengan menggunakan metode SOM.

Nilai max merupakan nilai tertinggi, nilai mean adalah nilai rata-rata data, nilai median adalah nilai tengah, dan min merupakan nilai minimum dari data yang ada. Berikut adalah rumus mencari nilai max, mean, median, dan min.

$$\text{Max} = \max (\text{data}) \quad (2.5)$$

$$\text{Mean} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.6)$$

$$\text{Median} = x_{\left(\frac{n+1}{2}\right)} \quad (2.7)$$

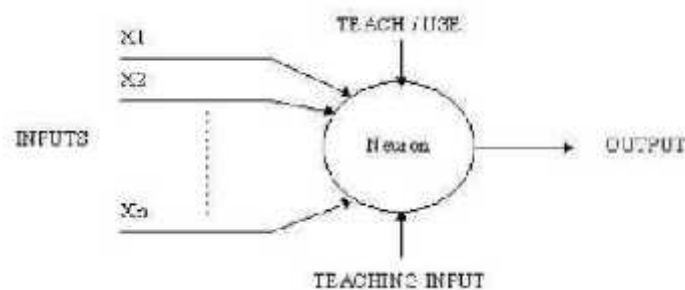
$$\text{Min} = \min (\text{data}) \quad (2.8)$$

Dimana n adalah banyaknya data, x adalah data ke-i dan i mempunyai nilai 1 sampai n.

2.2.3 Clustering

Clustering merupakan proses untuk menemukan sekumpulan model yang berbeda dan terbagi menjadi beberapa kelas sesuai dengan data yang ada dari masing-masing objek atau data yang kelasnya belum diketahui. (Fansuri 2011)

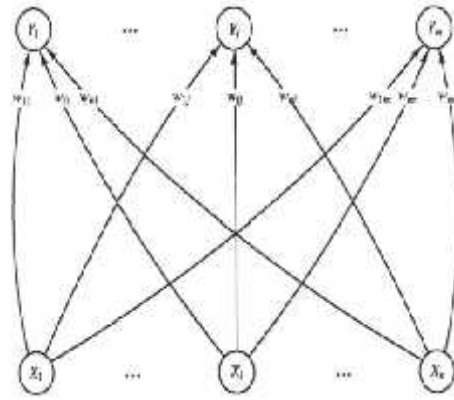
Jaringan syaraf tiruan merupakan representasi buatan otak manusia untuk proses simulasi pembelajaran dan menggunakan computer untuk proses penghitungannya. (Kusumadeni 2003) Jaringan ini mirip dengan jaringan syaraf biologi dimana ada masukan informasi yang akan diproses untuk dikirimkan ke neuron-neuron dan menghasilkan sebuah keluaran informasi. Gambar dibawah ini merupakan arsitektur sederhana dari jaringan syaraf tiruan.



Gambar 2.6 Arsitektur jaringan syaraf tiruan
(sumber : Mirawanti 2010)

SOM (*Self-Organizing Maps*) merupakan model dari jaringan syaraf tiruan yang paling populer dan termasuk jaringan syaraf *unsupervised* yang menyerupai model jaringan syaraf manusia. SOM dikenal juga dengan nama jaringan Kohonen karena ditemukan oleh Prof. Teuvo Kohonen di tahun 1982. Pada jaringan ini terdapat sebuah cluster, dimana neuron-neuro masukan akan membentuk kelasnya masing-masing secara otomatis berdasarkan nilai-nilai tertentu dari informasi masukan tersebut.

Metode SOM digunakan untuk pemetaan clustering dimana prosesnya adalah suatu objek yang masuk pada cluster ke-j tergantung pada jarak yang ditetapkan. Berikut adalah arsitektur dari SOM.



Gambar 2.7 Arsitektur SOM
(sumber : Mirawanti 2010)

- X : vektor inputan pembelajaran (X_1, X_2, \dots, X_n)
- α : *learning rate*
- R : radius *neighborhood*
- X_i : neuron / node input
- Y_j : neuron / node output
- W_{ij} : bias pada neuron output ke-j

Jaringan SOM mempunyai dua lapisan (*layer*), yaitu lapisan masukan (*input*) dan lapisan keluaran (*output*) dimana setiap neuron pada lapisan *output* merepresentasikan kelas dari *input*-an yang diberikan.

Dan berikut ini adalah tahapan dalam algoritma pembelajaran jaringan SOM. (Siang 2009)

- Langkah 0: inialisasi bobot (W_{ij})
 - Tentukan parameter jarak (R)
 - Tentukan *learning rate* (α)
- Langkah 1: jika kondisi *stop* salah, lakukan langkah 2-8
- Langkah 2: untuk semua *input* vector x lakukan langkah 3-5
- Langkah 3: untuk setiap j,

$$D_j = \sum_i^n (W_{ij} - X_i)^2 \quad (2.9)$$
- Langkah 4: dapatkan j sehingga D(j) minimal
- Langkah 5: untuk semua j yang berada dalam jarak dan untuk semua i,

$$W_{ij} \text{ baru} = W_{ij} \text{ lama} + \alpha(X_i - W_{ij}) \text{ lama}$$

(2.10)

Langkah 6: *update learning rate*

Langkah 7: *Update jarak*

Langkah 8: Uji kondisi stop

(Stop jika $|W_{ij}(\text{baru}) - W_{ij}(\text{lama})| \leq \epsilon$)

Model jaringan SOM telah banyak digunakan di dunia industri, diantaranya untuk mesin monitoring, identifikasi kesalahan, dan robot kontrol.

2.2.4 Penghitungan Akurasi

Hasil yang diamati pada penelitian ini adalah tingkat akurasi dari penggunaan MFCC dan SOM dalam mengkluster lagu. Tingkat akurasi akan diperoleh dengan persamaan: (Lee 2007)

$$akurasi = \frac{\sum data\ benar\ dicluster}{\sum data} \times 100\% \quad (2.11)$$