

ANALISIS DAN SELEKSI FITUR AUDIO PADA MUSIK TRADISIONAL INDONESIA

Aisha Gemala Jondya ^{1*}, Bambang Heru Iswanto²

¹ School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

² Department of Physics, Universitas Negeri Jakarta, Jakarta, Indonesia

aisha.jondya001@binus.ac.id, bhi@unj.ac.id

Abstract

There are several recommended features from previous research that can be used to find the similarities between one music to another, from manually specified features to automatically extracted features. Researchers have suggested some numerical values of the features that represent audio, but optimal set of features to use in finding music similarities may be different from one study to another. It depends on the method or type of music used in a study. To find the optimal audio features for Indonesian's traditional music, a feature selection process conducted in this study. Four Indonesian's traditional musics from 4 different provinces segmented automatically into 60 audio segments. Eleven sets with total 36 audio features extracted from this 60 audio segments with the combination of mean and standard deviation statistic values. Principal Component Analysis (PCA) used to reduce the set of features. To prove the resulting set of features is an optimal features to use in finding Indonesian's traditional music similarities, a clustering conducted to the 60 audio segments. The clustering process using x-Means method with features obtained in this study shows that every audio segments from 1 song goes to the same cluster.

Keywords: Music Similarities, PCA, clustering

PENDAHULUAN

Indonesia merupakan negara yang kaya akan kebudayaan termasuk berbagai jenis musik tradisional yang tersebar di setiap pulau. Musik tradisional ini memiliki karakteristik khusus meliputi lirik, melodi, dan penggunaan berbagai jenis instrumen. Musik tradisional Indonesia dipengaruhi oleh beberapa hal seperti latar belakang sejarah, agama yang dianut masyarakat dan penyesuaian terhadap budaya asing. Oleh karena itu, musik tradisional dari satu daerah mungkin memiliki kesamaan dengan yang lain.

Salah satu tantangan dalam menganalisis kemiripan musik adalah untuk mengetahui apa yang

memungkinkan manusia untuk membedakan antara gaya musik yang tidak dapat dibandingkan secara langsung [1]. Fitur dari setiap data audio harus dipilih dan diekstrak untuk menemukan set fitur yang efektif. Sebagian besar penelitian dalam analisis audio dilakukan dengan menggunakan nilai-nilai numerik dari fitur-fitur yang mewakili sebuah audio.

Dalam sinyal musik, fitur terkait dengan dimensi utama yang biasanya diukur seperti *pitch*, ritme, harmoni, melodi, *timbre*, dan lokasi spasial. Ada begitu banyak studi yang merekomendasikan cara menghitung fitur semacam itu, tetapi ada kemungkinan bahwa tidak semua fitur relevan untuk studi musik tertentu. Dalam studi ini, peneliti mencoba menemukan set fitur yang optimal untuk menemukan kemiripan di antara musik tradisional Indonesia.

Proses analisis audio telah menjadi penelitian yang menarik baru-baru ini. Proses analisis dilakukan menggunakan fitur yang ditentukan secara manual atau secara otomatis diekstraksi dari sebuah *file* audio. Penelitian pengelompokan musik audio secara otomatis telah dilakukan menggunakan fitur yang ditentukan secara manual oleh beberapa penelitian terdahulu. Pengelompokan musik tradisional dilakukan oleh [2] menggunakan fitur yang ditentukan secara manual untuk membandingkan budaya 16 masyarakat Austronesia. Dalam penelitian ini, setiap lagu didengarkan dan dikodekan secara manual berdasarkan skema yang dikembangkan oleh penulis.

Proses analisis audio menjadi lebih mudah dan akurat menggunakan fitur yang diekstrak secara otomatis dari audio. Fitur audio *low-level* yang populer untuk digunakan dalam beberapa penelitian mengenai klasifikasi audio yang dijelaskan dalam [3]. Selain itu kombinasi penggunaan fitur juga umum digunakan pada sejumlah penelitian. Andreas dan kawan-kawan menggunakan fitur *low-level* dan *mid-level* untuk pengelompokan musik rakyat dari empat negara timur non-Eropa, musik barat dan musik rakyat Siprus [4].

METODOLOGI

Dalam penelitian ini, proses pemilihan fitur diawali dengan memilih 4 buah lagu instrumen dari 4 provinsi yang berbeda, yaitu, pada saat didengarkan ke-empat lagu tersebut dianggap penulis memiliki perbedaan yang signifikan. Ke-empat file lagu ini didapatkan dari berbagai sumber yaitu internet, dan CD lagu daerah. Setiap lagu kemudian diambil bagian 1 menit berukuran 21 MB, lalu bagian lagu tersebut disegmentasi kembali secara otomatis kemudian pada aplikasi MATLAB. Proses ekstraksi fitur dilakukan dengan meng-*cluster* 4 segmen lagu tersebut menggunakan 11 set (total 36) fitur yang direkomendasikan oleh Giannakopoulos [5]. Berikut ini, adalah definisi dari masing-masing fitur.

1. Energy

Let $x_i(n)$, $n = 1, \dots, W_L$ merupakan urutan sampel audio dari *frame* ke- i , di mana W_L adalah panjang *frame*. Energi jangka pendek dihitung berdasarkan:

$$E(i) = \sum_{n=1}^{W_L} |x_i(n)|^2 \quad (1)$$

2. Entropy Energy

Entropi energi jangka pendek dapat diartikan sebagai lonjakan perubahan dalam tingkat energi sinyal audio. entropi, $H(i)$ dari urutan e_j dihitung menurut persamaan:

$$H(i) = -\sum_{j=1}^K e_j \log_2(e_j) \quad (2)$$

3. Zero Crossing Rate

Zero-Crossing Rate (ZCR) dari *frame* audio adalah laju perubahan tanda sinyal dalam satu *frame*. Dengan kata lain, ZCR dapat diartikan sebagai berapa kali sinyal mengubah nilai, dari positif ke negatif dan sebaliknya, dibagi dengan panjang *frame*. ZCR didefinisikan berdasarkan persamaan berikut:

$$Z(i) = \frac{1}{2W_L} \sum_{n=1}^{W_L} |\text{sgn}[x_i(n)] - \text{sgn}[x_i(n-1)]| \quad (3)$$

4. Spectral Centroid

Spectral Centroid merupakan pusat dari 'gravitasi' spektrum. Nilai spektral centroid, C_i , dari *frame* audio ke- i didefinisikan sebagai:

$$C_i = \frac{\sum_{k=1}^{W_{fL}} kX_i(k)}{\sum_{k=1}^{W_{fL}} X_i(k)} \quad (4)$$

5. Spectral Entropy

Spectral Entropy dihitung dengan cara yang mirip dengan cara menghitung *entropy energy*, meskipun, kali ini, perhitungan terjadi dalam

domain frekuensi. *Spectral Entropy* dihitung berdasarkan persamaan:

$$H = -\sum_{f=0}^{L-1} n_f \cdot \log_2(n_f) \quad (5)$$

6. Spectral Flux

Spectral Flux mengukur perubahan spektral antara dua *frame* berurutan dan dihitung sebagai perbedaan kuadrat antara besaran normal spektrum dari dua *frame* jangka pendek berturut-turut:

$$Fl_{(i,i-1)} = \sum_{k=1}^{W_{fL}} (EN_i(k) - EN_{i-1}(k))^2 \quad (6)$$

7. Spectral Roll-off

Fitur ini didefinisikan sebagai frekuensi yang persentase tertentu (biasanya sekitar 90%) dari distribusi besarnya spektrum dikonsentrasikan. Oleh karena itu, jika koefisien DFT ke- m sesuai dengan *spectral roll-off* dari *frame* ke- I , maka memenuhi persamaan berikut:

$$\sum_{k=1}^m X_i(k) = C \sum_{k=1}^{W_{fL}} X_i(k) \quad (7)$$

8. Mel-Frequency Cepstral Coefficient

Mel-Frequency Cepstrum Coefficients (MFCCs) sangat populer di bidang pengolahan audio percakapan (*speech audio*). MFCC sebenarnya merupakan tipe representasi *cepstral* dari sinyal, dimana pita frekuensi didistribusikan menurut skala-mel, bukan pada pendekatan jarak linier. Dalam penelitian ini 13 koefisien fitur MFCC dihitung.

9. Harmonic

Fitur *harmonic* dalam penelitian ini terdiri dari fitur *Harmonic Ratio* dan *Fundamental Frequency*. Rasio harmonik adalah proporsi *harmonic* dalam spektrum. HR menggambarkan tingkat harmonisasi *frame* audio.

$$r(i, k) = \frac{\sum_{j=m}^{m+n-1} s(j)s(j-k)}{(\sum_{j=m}^{m+n-1} s(j)^2 \times \sum_{j=m}^{m+n-1} s(j-k)^2)^{0.5}} \quad (8)$$

$$H(i) = \max_{k=Q} r(i, k) \quad (9)$$

sedangkan *Fundamental Frequency* adalah frekuensi yang setara dengan panjang periode sinyal.

10. Chroma Vector

Chroma Vector merupakan representasi 12 unsur-unsur energi spektral. Fitur ini banyak digunakan sebagai pendeskripsi aplikasi yang berkaitan dengan musik. Koefisien DFT *Chroma Vector* dihitung dengan mengelompokkan *frame* jangka pendek ke dalam 12 kelompok. Setiap kelompok dihitung

berdasarkan persamaan:

$$v_k = \sum_{n \in S_k} \frac{x_l(n)}{N_k}, \quad k \in 0, \dots, \dots, \dots, 11 \quad (10)$$

11. *Spectral Zone*

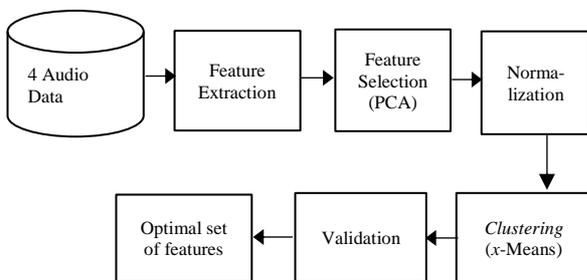
Fitur *Spectral Zone* menunjukkan rentang frekuensi spektrum dalam *frame* audio. *Spectral Zone* dihitung dengan membagi jumlah total *frame* FFT dengan jumlah *frame* FFT di kisaran 100-500 HZ.

Sebelas Set fitur ini diekstraksi dari 60 segmen audio. Segmen audio tersebut merupakan hasil segmentasi otomatis 4 musik instrumen yang memiliki perbedaan signifikan dari 4 provinsi yang berbeda dengan mengkombinasikan nilai statistik rata-rata dan standar deviasi. Empat segmen dari setiap lagu yang panjangnya 1 menit, diekstraksi dengan menggunakan teknik *mid-term windowing* secara otomatis. Berdasarkan teknik ini, setiap *window* disegmentasi menjadi 15 segmen yang lebih pendek (masing-masing segmen memiliki panjang 4 detik) dengan menggunakan *mtWin* dan *mtStep* pada perangkat lunak MATLAB. Segmentasi otomatis ini memperoleh 60 objek data berlabel. Tabel 1 menunjukkan daftar lagu yang digunakan dalam penelitian ini.

Tabel 1. Daftar Lagu

No.	Title	Province
1.	Kromong Dua Belas	Bengkulu
2.	Gilak	Bali
3.	Gamelan	Central Java
4.	Rantak	West Sumatera

Sebelas set fitur tersebut, kemudian direduksi atau dikurangi dengan memilih fitur optimal menggunakan metode PCA. Fitur yang dipilih divalidasi dengan mengelompokkan segmen audio menggunakan metode *x-Means*. Menggunakan metode *clustering*, segmen yang memiliki nilai-nilai fitur serupa akan dikelompokkan dalam *cluster* yang sama. Proses Seleksi Fitur digambarkan pada diagram berikut.



Gambar 1. Proses Seleksi Fitur

Algoritma *x-Means* digunakan karena metode ini memungkinkan kumpulan data dibagi pada jumlah kelompok optimal tanpa pengetahuan sebelumnya [6] dimana metode *x-Means* merupakan pengembangan dari metode *k-Means*. *k-Means* adalah metode *clustering* partisi yang paling populer. Metode ini membagi set data berbasis pusat data atau sentroid. Kelebihan metode ini yang menjadi daya tarik dan menjadikannya populer digunakan adalah kesederhanaan teknik penentuan keanggotaannya. Metode *k-means* mengukur kemiripan data dengan menghitung jarak *euclidean* setiap data terhadap sentroid. Langkah yang dilakukan pada metode *x-means* sesuai dengan langkah awal pada metode *k-means* sebagai metode dasarnya. Data dikelompokkan menjadi *k* kelompok dimana setiap poin data dimiliki oleh satu sentroid. Algoritma yang digunakan terus melacak sentroid setiap kluster pada proses iterasi, dimana sebelum iterasi pertama dimulai nilai sentroid ditentukan secara acak. Algoritma berhenti ketika posisi sentroid cenderung tidak berubah lagi. Pada setiap iterasi terjadi proses berikut:

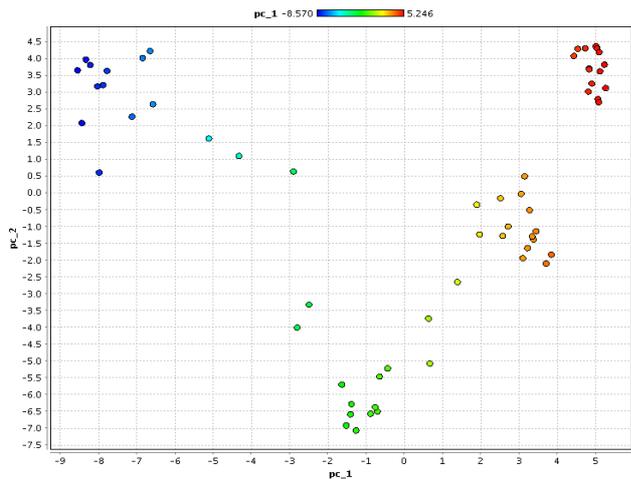
1. Untuk setiap titik *x*, temukan sentroid yang posisinya terdekat dengan *x*. Hubungkan *x* dengan sentroid tersebut.
2. Hitung ulang posisi sentroid, untuk setiap sentroid, dengan memperhitungkan titik pusat titik-titik data yang terhubung pada sentroid tersebut.

Langkah selanjutnya pada metode *x-means* adalah menemukan jumlah kluster *k* yang optimal untuk set data. Algoritma dasar yang digunakan terdiri dari 2 proses yang terus berulang sampai selesai.

1. *Improve-Params*
2. *Improve-Structure*
3. *If k > k_{max}* maka proses berhenti. Laporkan hasil terbaik dari perhitungan struktur. *Else* kembali ke 1.

HASIL DAN PEMBAHASAN

Sebelas set fitur diekstrak dari objek data ini dengan menggabungkan nilai statistik yaitu rata-rata dan standar deviasi. Ekstraksi fitur dilakukan pada aplikasi MATLAB dengan menggunakan fungsi `stFeatureExtraction()`. dan menghasilkan vektor fitur dimensi 60x72. Tujuh puluh dua fitur direduksi dengan menggunakan metode PCA untuk menemukan fitur optimal yang akan digunakan dalam pengelompokan. Proses ini dilakukan pada aplikasi RapidMiner dengan memilih PC1 dan PC2 sebagai 2 komponen utama untuk menentukan fitur. Kedua *principal component* ini dapat memisahkan 60 segmen lagu ke dalam 4 kelompok yang berbeda seperti yang dapat dilihat pada gambar di bawah ini.



Gambar 2. Hasil Plot Proses PCA

Kedua *Principal Component* ini digunakan untuk menafsirkan tabel eigenvector. Zero-Crossing Rate, Energy, Spectral centroid, Spectral Entropy, dan Spectral Roll-off memiliki nilai signifikan untuk dua parameter PC dibandingkan dengan fitur-fitur lainnya. sehingga, diyakini setelah dibuktikan dengan proses clustering nantinya kelima fitur ini akan dapat memisahkan 60 segmen lagu tersebut kedalam 4 kelompok. 5 set fitur dengan kombinasi nilai rata-rata dan standar deviasi kemudian digunakan pada proses *clustering*. Jumlah total fitur yang digunakan adalah 12.

Kemudian, untuk membuktikan 5 set fitur terpilih tersebut merupakan fitur-fitur yang signifikan merepresentasikan ciri dari lagu daerah Indonesia, dilakukan proses clustering terhadap fitur-fitur tersebut. Dimana setiap segmen lagu merupakan objek data dimana fitur yang digunakan adalah kelima set fitur terpilih tersebut. Proses ini diperlukan karena setiap objek data yang memiliki nilai fitur serupa akan berada pada cluster yang sama sehingga membuktikan bahwa fitur-fitur yang terpilih merupakan fitur yang optimal menunjukkan kemiripan dari lagu-lagu tradisional Indonesia.

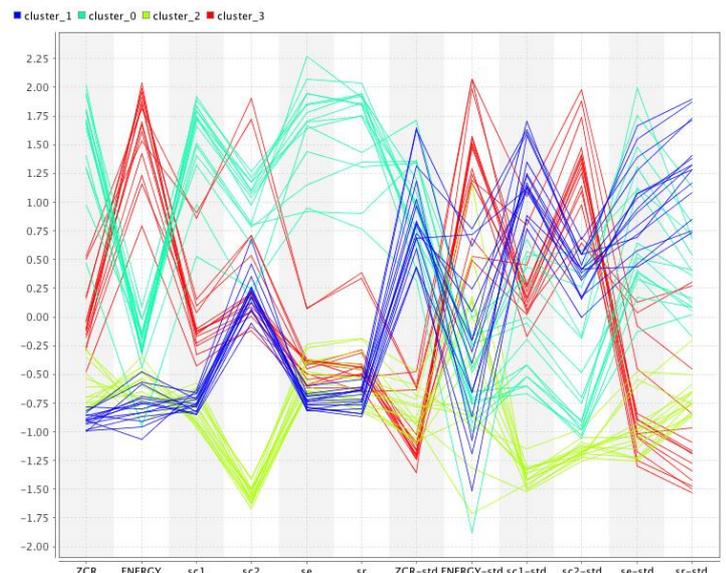
Clustering dengan metode *x-Means* dilakukan pada vektor fitur dimensi 60x12. Proses *clustering* menghasilkan hasil yang memuaskan. Algoritma *x-Means* dapat mendeteksi 4 kelompok secara otomatis tanpa pengetahuan atau training sebelumnya. Semua segmen audio dari satu lagu masuk ke dalam kelompok yang sama. Komposisi setiap *cluster* adalah:

1. Semua segmen audio Bengkulu_Kromong12.wav berada pada *cluster* 0.
2. Semua segmen audio dari Bali_Gilak.wav berada pada *cluster* 1.
3. Semua segmen audio dari Jateng_Gamelan_4.wav berada pada *cluster* 2.

4. Semua segmen audio dari Sumbar_rantak.wav berada pada *cluster* 3.

Hasil clustering divalidasi dengan menghitung nilai *entropy* dan *purity*. Dimana Nilai *entropy* dihitung untuk melihat sebanyak apa variasi anggota kelas dalam sebuah *cluster*. Sebuah *cluster* yang baik akan menghasilkan nilai *entropy* yang kecil hingga mendekati nilai 0. Sedangkan Nilai *purity* dihitung untuk dapat melihat kemurnian anggota sebuah *cluster*. Sebuah *cluster* yang murni hanya memiliki anggota yang berasal dari satu kelas saja, dan begitu pula sebaliknya. Berdasarkan hasil clustering, *clustering* memiliki nilai *purity* 1 dan nilai *entropy* 0. Hal ini membuktikan bahwa hasil clustering terhadap vektor fitur terpilih menghasilkan hasil yang sangat baik sehingga dapat disimpulkan fitur terpilih signifikan merepresentasikan ciri lagu daerah Indonesia.

Hasil *clustering* divisualisasikan dengan grafik koordinat paralel di bawah ini karena dimensi vektor fitur yang tinggi.



Gambar 3. Grafik Koordinat Paralel Hasil Clustering

Dari grafik tersebut juga dapat terlihat nilai-nilai fitur pada setiap lagu daerah tertentu. Seperti lagu Gilak dari Bali cenderung memiliki nilai rata-rata fitur ZCR yang rendah mirip dengan lagu Gamelan dari Jawa Tengah. Begitu pula dengan nilai rata-rata fitur *Spectral Entropy* dan *Spectral Roll-Off* dari lagu Gilak, Gamelan dan Rantak yang cenderung memiliki kedekatan nilai.

KESIMPULAN

Penelitian ini diharapkan menjadi langkah awal dari penelitian selanjutnya tentang musik tradisional Indonesia. Hasil penelitian ini menemukan bahwa penggunaan 5 set fitur; Zero Crossing Rate, Energy, Spectral centroid, Spectral Entropy dan Spectral Roll-off menghasilkan hasil *clustering* yang baik. Berdasarkan hal ini, penelitian selanjutnya diharapkan dapat menemukan kombinasi set fitur yang lain, atau bahkan untuk menemukan fitur yang lebih optimal yang didapatkan dengan menggunakan data audio yang lebih lengkap. Selain itu, kolaborasi penelitian dengan para ahli dari berbagai bidang, seperti antropolog budaya, diyakini dapat menghasilkan lebih banyak informasi bermanfaat.

REFERENSI

- [1] Rini Wongso and Santika Diaz, "Automatic Music Classification Using Dual Tree Complex Wavelet Transform and Support Vector Machine," *Journal of Theoretical and Applied Information Technology*, vol. 63, 2014.
- [2] Tom Rzeszutek, Patrick E. Savage, and Steven Brown, "The Structure of Cross-Cultural Musical Diversity," *Royal Society Publishing*, 2011
- [3] DPW Ellis, X Zeng, and McDermott JH, "Classifying Soundtracks With Audio Texture Features," *ICASSP*, 2011.
- [4] Andreas Neocloud, Maria Pantelli, Ioannou Rafaela, Petkov Nikolai, and Christos Schizas, "A Machine Learning Approach for Clustering Western and non-Western Folk Music Using Low-level And Mid-level Features.," *International Workshop on Machine Learning and Music*, 2013.
- [5] Theodoros Giannakopoulos and Pikrakis, *Introduction to Audio Analysis: A Matlab Approach.*: Elsevier, 2014.
- [6] T Ishioka, "An Expansions of x-Means For Automatically Determining The Optimal Number of Clusters," in *International Conference Computational Inteligence*, pp. 91-96.