

Klasifikasi Genre Musik Menggunakan Metode Deep Learning Convolutional Neural Network dan Mel-Spektrogram

Danny Lionel, Rudy Adipranata, Endang Setyati

Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Kristen Petra

Jl. Siwalankerto 121 – 131 Surabaya 60236

Telp. (031) – 2983455, Fax. (031) – 8417658

E-Mail: dannylie888@gmail.com, rudya@petra.ac.id, endang@stts.edu

ABSTRAK

Pertumbuhan database musik yang berkembang secara sangat pesat menyebabkan sukarnya proses pengelompokan musik dalam kategori tertentu, sehingga dapat berakibat sulitnya pencarian suatu kategori musik dalam jumlah banyak dan skala yang besar. Dengan perkembangan jaman, sekarang telah mulai dikembangkan metode agar suatu audio file dapat dikenali secara otomatis dari fitur-fitur yang telah diekstrak sebelumnya dengan bantuan MFCC (Mel Frequency Cepstral Coefficients) dan ANN (Artificial Neural Network).

Dalam penelitian ini akan digunakan metode Mel-spectrogram. Dimana Mel spektrogram merupakan hasil pemetaan fitur yang telah diambil oleh metode MFCC, yang akan diklasifikasikan dan dimasukkan kedalam *Convolutional Neural Network*. Yang akan dibedakan *activation function* nya yaitu ReLU dan ELU.

Penelitian ini menunjukkan bahwa pengambilan fitur dari audio dengan menggunakan MFCC merupakan metode yang benar, dan dalam hasil pengujian, banyaknya dataset, iterasi *training*, dan spesifikasi komputer sangat mempengaruhi tingkat akurasi dan lama pembuatan neural network model yang optimal. Dalam hasil penelitian ini telah diuji beberapa kali didapatkan hasil akurasi yang paling optimal yaitu 99%.

Kata Kunci: *Digital Image Processing, Artificial Neural Network, Audio Processing, Mel cepstral coefficients*

ABSTRACT

The growth of the music database that has developed very rapidly has caused the difficulty of the process of grouping music in certain categories, so that it can result in the difficulty of finding a large number of music categories and large scale. With the development of the era, methods have now begun to be developed so that an audio file can be automatically recognized from features that have been extracted previously with the help of MFCC (Mel Frequency Cepstral Coefficients) and ANN (Artificial Neural Network).

The activation function that will be distinguished is ReLU and ELU. In this research Mel-spectrogram method will be used. Where the spectrogram Mel is the result of feature mapping that has been taken by the MFCC method, which will be classified and entered into the Convolutional Neural Network.

This study shows that the feature extraction of an audiofile using MFCC is the correct method, and in the test results, the number of datasets, training iterations, and computer specifications strongly influences the accuracy and length of optimal neural

network model manufacturing. In the results of this study tested several times the results of the most optimal accuracy were 99%.

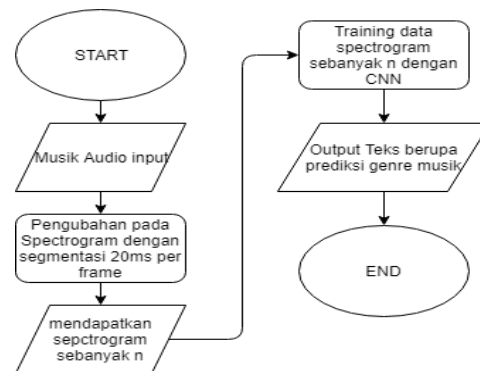
Keywords: *Digital Image Processing, Artificial Neural Network, Audio Processing, Mel cepstral coefficients*

1. PENDAHULUAN

Musik dapat dibagi menjadi banyak kategori yang berbeda berdasarkan dari ritme, gaya dan latar belakang kulturnya. Gaya inilah yang kita sebut dengan *genre* musik[3]. *Genre* musik adalah label yang sifatnya mengkategorikan, dibuat oleh ahli musik dan dapat digunakan untuk mendeskripsikan, mengelompokkan, serta membandingkan lagu, album, dan penulis dalam dunia musik yang luas ini.

Jaman sekarang ini, angka database musik bisa mencapai angka yang sangat tinggi, yaitu puluhan juta lagu. Angka tersebut merupakan perkiraan lagu yang telah dirilis secara resmi, belum menghitung lagu yang tidak resmi. Dengan besarnya angka database musik ini, diperlukan adanya *indexing* atau *labelling* untuk memudahkan pencarian lagu. Dengan *labelling* lagu secara tidak teratur, maka dapat mengakibatkan hasil pencarian yang tidak sesuai dengan keinginan. Menelusuri dan mencari database yang begitu besar akan sangat sulit, sedangkan meng-*assign* setiap data lagu dengan *genre* yang sesuai telah terbukti lebih *user-friendly*, dalam proses pencarian.

Biasanya, *genre* musik diberikan oleh para ahli secara *manual*. Dengan tingginya angka perkembangan musik di jaman sekarang ini, proses pengerjaan *manual* akan banyak menghasilkan bias atau kesalahan, dikarenakan banyaknya aturan-aturan dalam pengklasifikasian musik yang terus-menerus bertambah. Maka dari itu diperlukan suatu cara untuk melakukan otomatisasi pada masalah ini. Gambar 1 merupakan flowchart arsitektur sistem dari program.

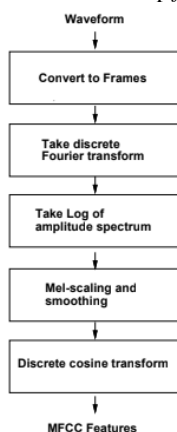


Gambar 1. Flowchart program

2. DASAR TEORI

2.1 MFCC (Mel-Frequency Cepstral Coefficients)

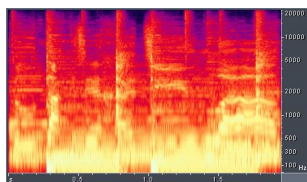
Spektrum suara menampilkan frekuensi berbeda yang ada dalam suara [7]. Mel Frequency Cepstrum (MFC) adalah representasi transformasi linear kosinus dari spektrum daya log waktu singkat dari sinyal ucapan pada skala frekuensi Mel non-linear. MFCC (*Mel-Frequency Cepstral Coefficient*) telah menjadi teknik yang sangat umum dan efisien untuk *signal processing* [4]. MFCC mewakili model telinga, dan dapat menghasilkan pengenalan suara terutama ketika menggunakan jumlah koefisien yang tinggi [6]. Gambar 2 menunjukkan proses dari pembuatan fitur MFCC. Langkah pertama ada membagi sinyal menjadi banyak *frame*, biasanya menggunakan *windowing function* dengan interval tetap. Tujuannya adalah untuk memodelkan bagian kecil (biasanya 20 ms) dari sinyal yang diterima. *Window function*, atau disebut juga *Hamming window*, akan menghilangkan *edge effects*. *Cepstral feature vector* akan dibuat untuk setiap *frame*.



Gambar 2 – Proses untuk membuat fitur MFCC

2.2 Spektrogram

Gambar spektrogram adalah representasi visual dari spektrum frekuensi sinyal [2]. Spektrogram dapat dibentuk dengan menggunakan *Fourier Transform*. Spektrogram didefinisikan sebagai besarnya kuadrat dari STFT, memberikan kekuatan suara untuk frekuensi dan waktu tertentu dalam dimensi ketiga [1]. Membuat spektrogram dengan FFT dilakukan dengan pertama mengambil data yang berada pada *time domain*, dan memecah data-data tersebut menjadi beberapa bagian, dan dilakukan *Fourier Transform* untuk menghitung *magnitude* dari spektrum frekuensi untuk setiap bagian [5]. Setelah itu tiap bagian akan disesuaikan pada garis vertical pada gambar, yang merupakan perbandingan *magnitude* dengan frekuensi dalam waktu tertentu. Setelah itu spektrum-spektrum tersebut akan di *plot* secara bersebelahan untuk membentuk suatu gambar. Contoh gambar spektrogram dapat dilihat pada gambar 3

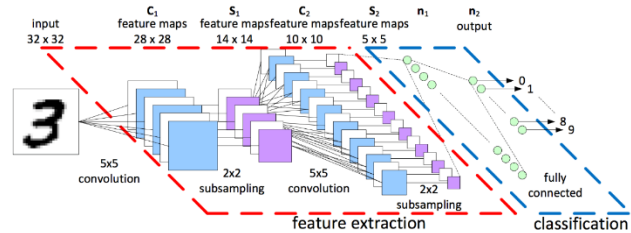


Gambar 3. Gambar Spektrogram

2.3 Convolutional Neural Network

Salah satu metode yang penting dalam *deep learning* adalah *Convolutional Neural Network* (CNN). CNN dirancang khusus untuk pengenalan dan klasifikasi gambar. CNN memiliki beberapa lapisan (*layer*) yang mengekstrak informasi dari gambar dan menentukan klasifikasi dari gambar tersebut

Terdapat tiga layer yang membentuk CNN yaitu *Convolutional Layer*, *Pooling Layer* dan *Fully Connected Layer*. Secara umum, layer tersebut diatur dengan cara menumpuk beberapa *convolutional layer* diikuti dengan *pooling layer*. Pola ini diulangi hingga input gambar semakin mengecil. Layer terakhir adalah *fully connected layer* yang memberikan hasil output akhir, berupa skor klasifikasi. Gambar 3 merupakan visualisasi dari proses *Convolutional Neural Network*.



Gambar 4. Proses Convolutional Neural Network

2.4 Fourier Transform

Fourier Transform (FT) adalah proses yang memecah fungsi waktu menjadi bagian-bagian dari fungsi itu sendiri. Bagian-bagian hasil *Fourier Transform* disebut *Fourier series*.

Discrete Fourier Transform (DFT) adalah FT yang diaplikasikan pada deret nilai kompleks. Penelitian ini menggunakan *Fast Fourier Transform* (FFT) berdasarkan implementasi dari Paul Bourke (1993) dalam FT karena FFT menghasilkan output yang sama dengan *Discrete Fourier Transform* (DFT), namun prosesnya lebih efisien dan cepat. Algoritma DFT adalah sebagai berikut:

Deret nilai kompleks $x(k)$ sepanjang N

$$x_0, x_1, x_2, x_3 \dots x_k \dots x_{N-1} \quad (1)$$

Dimana x adalah nilai kompleks

$$x_i = x_{\text{real}} + j x_{\text{imag}} \quad (2)$$

Transformasi maju untuk mengubah waveform menjadi spectrum adalah seperti persamaan (3)

$$X(n) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) e^{-jk2\pi n/N} \quad \text{for } n=0..N-1 \quad (3)$$

Transformasi mundur untuk mengembalikan spectrum menjadi waveform adalah seperti persamaan (4)

$$x(n) = \sum_{k=0}^{N-1} X(k) e^{jk2\pi n/N} \quad \text{for } n=0..N-1 \quad (4)$$

Meskipun fungsi yang digunakan membutuhkan bilangan imajiner, bilangan real bisa didapatkan dengan mengubah bagian imajiner (X_{imag}) menjadi 0.

3. DESAIN SISTEM

3.1 MFCC

MFCC merupakan metode ekstraksi fitur yang diperlukan untuk mengubah *audio* file menjadi suatu gambar. Pada proses ini, objek yang akan diproses adalah *audio* yang berextension mp3 atau wav, dengan minimal durasi *audio* selama 30 detik. Hasil dari proses ini akan disimpan dalam bentuk gambar (JPEG). Yang nantinya akan digunakan untuk *training* dan *testing*.

3.2 Data Handing

Dalam proses *neural network*, perlu dilakukan proses *data handing* yang digunakan sebagai alat bantu untuk me-load image kedalam satu *neural network*. Proses tersebut diawali dengan menspesifikasi kan *target folder* yang ingin diambil *images file* nya, lalu program akan me-list semua *image file* dan *folder* yang ada didalam *folder* tersebut. Setelah itu masing-masing dari *image file* tersebut akan dicatat dan dimasukkan ke dalam *array*, kemudian kan diberi *label* sesuai dengan index *folder* masing-masing, sehingga setiap *images file* memiliki *label* index masing-masing sesuai dengan *folder* masing-masing. Setelah semua *image* memiliki *labels* index masing-masing, maka data tersebut akan di *append* ke dalam *array* sehingga memudahkan untuk *passing data*.

3.3 Convolutional Neural Network

Sebelum dapat mengklasifikasi gambar karakter, perlu dilakukan training pada model CNN (*Convolutional Neural Network*). *Training* dilakukan dengan jumlah *dataset* sebanyak 2500 gambar dan juga dilakukan validasi hasil training dengan jumlah *dataset* sebanyak 500 karakter. Seluruh *dataset* berukuran 100x100 *pixel* dan merupakan *image*.

4. IMPLEMENTASI SISTEM

Implementasi sistem dilakukan pada computer dengan spesifikasi sebagai berikut, RAM: 8GB, DDR3, SSD: 128 GB, CPU: Core i5-2520M CPU @ 2.50GHz, Chipset: Intel(R) HD Graphics Family, OS: Microsoft Windows 7 Professional 64-bit.

Implementasi pengkodean sistem, menggunakan Bahasa pemrograman Python dengan versi 3.5.0. *GUI* yang digunakan dalam sistem aplikasi ini adalah *python Tkinter*. Adapun beberapa *library* yang mendukung sistem ini adalah: PIL, Matplotlib, Scikit-learn, Numpy, Tensorflow, Librosa, Statistics, Datetime, Os, Threading, Tkinter, Pathlib, Operator.

5. PENGUJIAN SISTEM

5.1 Pengujian terhadap sistem program

Pengujian terhadap program dimulai dari pemilihan lagu yang ingin diproses. Hasil dari spektrogram dapat dilihat pada gambar 5. Dalam spektrogram ini ada banyak nilai dan fitur dari audio yang ditampilkan



Gambar 5. Hasil Spektrogram untuk 1 lagu

5.2 Convolutional Neural Network

Dari proses MFCC, lagu yang dipilih telah diubah menjadi gambar dengan banyak 3 gambar spektrogram. Masing-masing gambar akan dimasukkan kedalam *neural network* untuk diklasifikasikan.

5.3 Perhitungan Akurasi

Perhitungan akurasi dibagi menjadi 2 yaitu perhitungan akurasi untuk *activation function* ELU dan ReLU. Training dilakukan sebanyak 20 *epoch*, 40 *epoch* dan 60 *epoch*.

Tabel 1. Sample dari hasil klasifikasi lagu dengan menggunakan ELU 20 epoch

Genre Lagu	Jumlah Training	Jumlah Testing	Hasil		Akurasi
			Benar	Salah	
Blues	250	50	24	26	48,00%
Classical	250	50	48	2	96,00%
Country	250	50	49	1	98,00%
Disco	250	50	37	13	74,00%
HipHop	250	50	1	49	2,00%
Jazz	250	50	25	25	50,00%
Metal	250	50	47	3	94,00%
Pop	250	50	13	37	26,00%
Reggae	250	50	37	13	74,00%
Rock	250	50	18	32	36,00%

Pada pengujian yang tertera pada tabel 1 dapat dilihat bahwa hasil klasifikasi yang paling benar ada pada genre *country*, yang mempunyai nilai sebesar 98%, sedangkan hasil yang paling kurang adalah *hiphop*, dengan nilai sebesar 2%. Rata-rata akurasi pada pengujian ini adalah 59,8%

Tabel 2. Sample dari hasil klasifikasi lagu dengan menggunakan ReLU 20 epoch

Genre Lagu	Jumlah Training	Jumlah Testing	Hasil		Akurasi
			Benar	Salah	
Blues	250	50	40	10	2,00%
Classical	250	50	47	3	22,00%
Country	250	50	41	9	82,00%
Disco	250	50	23	27	46,00%
HipHop	250	50	29	21	58,00%
Jazz	250	50	34	16	68,00%
Metal	250	50	46	4	8,00%
Pop	250	50	40	10	80,00%
Reggae	250	50	26	24	52,00%
Rock	250	50	29	21	58,00%

Pada pengujian yang tertera pada tabel 2 dapat dilihat bahwa hasil klasifikasi yang paling benar ada pada genre *country*, sama seperti pengujian ELU, namun dengan persentase yang lebih rendah, yaitu 82%. Sedangkan hasil yang paling buruk adalah metal, dengan persentase sebesar 8%. Rata-rata akurasi pada pengujian ini adalah 71 %.

Hasil dari pengujian untuk iterasi sebanyak 20 epoch memiliki akurasi 59.8% dan 71%. Dalam pengujian ini *activation function* ReLU memiliki akurasi yang lebih tinggi daripada *activation function* ELU dengan selisih sebanyak 11.2%.

Tabel 3. Sample dari hasil klasifikasi lagu dengan menggunakan ELU 40 epoch

Genre Lagu	Jumlah Training	Jumlah Testing	Hasil		Akurasi
			Benar	Salah	
Blues	250	50	47	3	94,00%
Classical	250	50	33	17	66,00%
Country	250	50	44	6	88,00%
Disco	250	50	37	13	74,00%
HipHop	250	50	16	34	94,00%
Jazz	250	50	36	14	32,00%
Metal	250	50	50	0	100,00%
Pop	250	50	29	21	58,00%
Reggae	250	50	40	10	80,00%
Rock	250	50	33	17	66,00%

Pada pengujian yang tertera pada tabel 3 dapat dilihat bahwa hasil klasifikasi yang paling benar ada pada genre *metal*, yang mempunyai nilai sebesar 100%, sedangkan hasil yang paling kurang adalah *jazz*, dengan nilai sebesar 32%. Rata-rata akurasi pada pengujian ini adalah 73%.

Tabel 4. Sample dari hasil klasifikasi lagu dengan menggunakan ReLU 40 epoch

Genre Lagu	Jumlah Training	Jumlah Testing	Hasil		Akurasi
			Benar	Salah	
Blues	250	50	41	9	82,00%
Classical	250	50	41	9	82,00%
Country	250	50	36	14	72,00%
Disco	250	50	22	28	44,00%
HipHop	250	50	22	28	44,00%
Jazz	250	50	36	14	72,00%
Metal	250	50	26	24	52,00%
Pop	250	50	27	23	54,00%
Reggae	250	50	32	18	64,00%
Rock	250	50	26	24	52,00%

Pada pengujian yang tertera pada tabel 4 dapat dilihat bahwa hasil klasifikasi yang paling benar ada pada genre *blues* dan *Classical*, dengan keduanya memiliki persentase yang cukup besar yaitu 82%. Sedangkan hasil yang paling buruk adalah *disco* dan *hiphop* dengan besar yang sama juga, yaitu 44%. Rata-rata akurasi pada pengujian ini adalah 61.8 %.

Hasil dari pengujian untuk iterasi sebanyak 40 epoch memiliki akurasi 73% dan 61.8%. Dalam pengujian ini *activation function ELU* memiliki akurasi yang lebih tinggi daripada *activation function ReLU* dengan selisih sebanyak 11.2%.

Tabel 5. Sample dari hasil klasifikasi lagu dengan menggunakan ELU 60 epoch

Genre Lagu	Jumlah Training	Jumlah Testing	Hasil		Akurasi
			Benar	Salah	
Blues	250	50	48	2	96,00%
Classical	250	50	50	0	100,00%
Country	250	50	45	5	90,00%
Disco	250	50	45	5	90,00%
HipHop	250	50	45	5	90,00%
Jazz	250	50	49	1	98,00%
Metal	250	50	50	0	100,00%
Pop	250	50	49	1	98,00%
Reggae	250	50	49	1	98,00%
Rock	250	50	49	1	98,00%

Pada pengujian yang tertera pada tabel 5 dapat dilihat bahwa hasil klasifikasi yang paling benar ada pada genre *metal* dan *classical*, yang mempunyai nilai sebesar 100%, sedangkan hasil yang paling kurang adalah *country*, *disco*, dan *HipHop*, dengan nilai sebesar 90%. Rata-rata akurasi pada pengujian ini adalah 95.8%.

Tabel 6. Sample dari hasil klasifikasi lagu dengan menggunakan ReLU 60 epoch

Genre Lagu	Jumlah Training	Jumlah Testing	Hasil		Akurasi
			Benar	Salah	
Blues	250	50	50	0	100,00%
Classical	250	50	50	0	100,00%
Country	250	50	50	0	100,00%
Disco	250	50	50	0	100,00%
HipHop	250	50	50	0	100,00%
Jazz	250	50	50	0	100,00%
Metal	250	50	49	1	98,00%
Pop	250	50	50	0	100,00%
Reggae	250	50	50	0	100,00%

Pada pengujian yang tertera pada tabel 6 dapat dilihat bahwa hasil klasifikasi telah keluar dengan hasil yang cukup memuaskan. Dengan genre *blues*, *classical*, *country*, *disco*, *hiphop*, *jazz*, *pop*, *reggae* dengan hasil yang sempurna, sebesar 100%. Sedangkan hasil yang paling buruk adalah *metal* dengan besar yang sama juga, yaitu 98%. Rata-rata akurasi pada pengujian ini adalah 99.8 %.

Hasil dari pengujian untuk iterasi sebanyak 60 epoch memiliki akurasi 100% dan 99.8%. Dalam pengujian ini *activation function ELU* memiliki akurasi yang lebih tinggi daripada *activation function ReLU* dengan selisih sebanyak 0.2%.

Dari beberapa pengujian diatas, dapat disimpulkan bahwa semakin banyak iterasi epoch daripada *training* model tersebut, maka semakin baik pula model yang dihasilkan. Namun ada kalanya pada saat iterasi terlalu banyak, maka dapat terjadi yang namanya disebut *Overfitting*. Maka dari itu, iterasi dari training harus tidak berlebihan dan tidak kekurangan.

5.4 Perhitungan Akurasi

Tabel 7 dibawah menunjukkan hasil untuk beberapa lagu yang telah diklasifikasikan secara utuh. Proses klasifikasi dilakukan dengan mengambil *majority vote* dari tiap *classification sample* lagu. Program klasifikasi menggunakan *trained model* dengan menggunakan ELU yang diiterasikan sebanyak 60 epoch.

Pengujian ini dilakukan karena ada pertimbangan bahwa dalam suatu lagu, representasi 10 detik tidak dapat merepresentasikan seluruh durasi lagu. Maka dari itu, perlu dicoba untuk pengujian lagu yang diubah ke dalam spektrogram secara penuh sehingga dapat mengetahui persebaran klasifikasi *genre* dari lagu-lagu tersebut.

Dari tabel 7, dapat ditampilkan beberapa hasil dari klasifikasi lagu yang diuji oleh sistem. Dalam tabel tersebut terlihat bahwa pada lagu nomor 1 diklasifikasikan oleh sistem sebagai lagu *reggae*. Hal ini sesuai dengan genre asli dari lagu tersebut. Selain itu, lagu nomor 9 diklasifikasikan oleh sistem sebagai lagu *rock*, dimana seharusnya lagu nomor 9 merupakan lagu *blues*. maka dapat dihitung akurasi sistem yang tertera pada tabel 8.

Tabel 7. Tabel pengujian untuk beberapa lagu utuh

NO	Panyanyi - Judul Lagu	Frekuensi berdasarkan klasifikasi genre										
		Blues	Classical	Country	Disco	HipHop	Jazz	Metal	Pop	Reggae	Rock	
1	Siti Badriah- Lagi Syantik (Reggae)	2	2	4	0	2	1	0	1	0	0	
2	DNCE - Toothbrush (pop)	0	1	0	3	0	0	0	13	1	1	
3	Charlie Puth - We don't talk anymore (pop)	0	1	0	2	0	0	0	17	1	1	
4	Lobo Loco - Works Wonders (Blues)	0	2	0	0	0	21	0	0	0	0	
5	Derek Clegg - Selfish (Country)	2	3	10	1	0	0	0	0	0	0	
6	Young Kartz - Coupe (HipHop)	0	4	0	0	9	1	0	1	0	0	
7	Scott Holmes - Driven to Success (rock)	0	3	0	0	1	3	0	0	0	11	
8	Dee Yan Key- Weep No More (classical)	0	43	0	0	0	0	0	1	0	2	
9	Council of Elders - James Allen Blues (blues)	2	3	0	0	6	0	0	0	0	11	
10	Justin Bieber - Love Yourself (pop)	0	0	1	2	0	0	0	15	3	5	

Tabel 7 menunjukkan persebaran klasifikasi pada sample waktu tertentu. Seperti pada lagu Siti Badriah – Lagi Syantik, segmen yang terklasifikasi sebagai reggae ada sebanyak 8. Maka lagu tersebut sudah terklasifikasi kan dengan benar yaitu genre reggae. Selain itu, lagu Council of Elders – James Allen Blues yang seharusnya merupakan genre blues diklasifikasi kan oleh program sebagai rock dengan 11 vote.

Tabel 8. Tabel Akurasi sistem

No	Artist-Judul Lagu	Hasil Sistem	Genre Asli	Akurasi Sistem	
				Benar	Salah
1	Siti Badriah- Lagi Syantik	Reggae	Reggae	1	0
2	DNCE - Toothbrush	Pop	Pop	1	0
3	Charlie Puth - We don't talk anymore	Pop	Pop	1	0
4	Lobo Loco - Works Wonders	Jazz	Blues	0	1
5	Derek Clegg - Selfish	Country	Country	1	0
6	Young Kartz - Coupe	Hiphop	Hiphop	1	0
7	Scott Holmes - Driven to Success	Rock	Rock	1	0
8	Dee Yan Key- Weep No More	Classical	Classical	1	0
9	Council of Elders - James Allen Blues	Rock	Blues	0	1
10	Justin Bieber - Love Yourself	Pop	Pop	1	0
Jumlah Akurasi				8	2
Persentase Akurasi				80%	20%

Dari tabel 8 diatas, dapat dilihat akurasi dari klasifikasi sistem yang telah dihasilkan, dengan hasil yang cukup memuaskan, yaitu sebesar 80%, selain itu hasil klasifikasi yang salah sebesar 20%.

6. KESIMPULAN DAN SARAN

6.1 Kesimpulan

Dari hasil pembuatan aplikasi dan *neural network model* baik dengan *activation function ELU* maupun *RELU* untuk klasifikasi *genre* lagu, dapat diambil kesimpulan antara lain :

- Tingkat akurasi saat *testing* pada pembuatan *convolutional neural network model* sangat bergantung pada lama *training* dan banyak *dataset* yang dimasukkan kedalam *neural network*. Dengan semakin lama waktu *training*, dapat dilihat bahwa tingkat akurasi meningkat secara signifikan pada *testing* yang menggunakan *ReLU* . Tingkat akurasi tertinggi dicapai pada epoch ke 60, dengan tingkat akurasi sebesar 99% untuk *RELU* dan 95% untuk *ELU*.
- Pembuatan struktur *neural network model* dan lama *training* sangat bergantung pada *resource memory* yang terdapat pada komputer yang digunakan.
- Aplikasi dibuat berbasis *python GUI* memiliki tampilan sederhana dan *user friendly* untuk memudahkan pengguna untuk mengoperasikan aplikasi.

6.2 Saran

Saran yang diberikan untuk penyempurnaan dan pengembangan lebih lanjut untuk program ini adalah sebagai berikut :

- Dapat dilakukan penambahan *dataset* sehingga akan memperluas jumlah variasi yang dapat dikenali *neural network model*.
- Lama *training* dapat diperpanjang untuk mendapatkan *model* yang optimal
- Menggunakan mesin komputer yang lebih baik sehingga waktu *training* dapat dipersingkat
- Dapat dilakukan modifikasi pada struktur *neural network model* (seperti penambahan *convolution layer*, pengurangan *pooling layer*, mengubah *activation function* yang digunakan) untuk meningkatkan akurasi *neural network model*.

7. DAFTAR PUSTAKA

- [1] Huzafiah, M. 2017. Comparison of Time-Frequency Representations for Environmental Sound Classification using Convolutional Neural Network. *arXiv:1706.07156*.
- [2] Ilyas Ozer, Z. O. 2016. Lanczos Kernel Based Spectrogram Image Features for Sound Classification.
- [3] Politis & Dionysios. 2016. *Digital Tools for Computer Music Production and Distribution*.
- [4] S Lalitha, D. G. 2015. Emotion Detection using MFCC and Cepstrum Features. *Procedia Computer Science 70*, 29-35.
- [5] Shikha Gupta, J. J. 2013. Feature Extraction Using MFCC. *An International Journal (SIPIJ) Vol.4, No.4*.
- [6] Wanli, G. 2013. The research of feature extraction based on MFCC for speaker recognition.
- [7] Wolfe, J. 2014. *What is a Sound Spectrum?* .Diambil kembali dari unsw.edu.au: <http://newt.phys.unsw.edu.au/jw/sound.spectrum.html>