

IMPLEMENTASI ALGORITMA GENETIKA UNTUK SELEKSI FITUR PADA KLASIFIKASI GENRE MUSIK MENGGUNAKAN METODE RANDOM FOREST

Nurlatifah Amini¹, Triando Hamonangan Saragih², Mohammad Reza Faisal³, Andi Farmadi⁴, Friska Abadi⁵

Ilmu Komputer, Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Jalan Jenderal Ahmad Yani KM 36, Banjarbaru, Kalimantan Selatan 70714

Corresponding author. Email: ¹nurlatifah.amini@gmail.com, ²triando.saragih@ulm.ac.id, ³reza.faisal@ulm.ac.id, ⁴andifarmadi@gmail.com, ⁵friska.abadi@ulm.ac.id

Abstrak

Musik memiliki jenis yang beragam di Dunia. Adapun jenis musik yang paling populer diantaranya yaitu pop, disco, country, dangdut, jazz, blues, reggae, hiphop, rock, metal. Penelitian ini mengenai Klasifikasi *genre* musik menggunakan metode *Random Forest* menggunakan dataset dari GitHub atau GTZAN tentang *genre* musik dengan jumlah *label* ada 10, memiliki 26 fitur dan jumlah keseluruhan data ada 1000. Penelitian ini dibagi menjadi beberapa tahap, yaitu dengan klasifikasi seluruh data, mengklasifikasi dengan data yang dinormalisasi, melakukan klasifikasi dengan data asli menggunakan tahap seleksi fitur Algoritma Genetika, dan mengklasifikasi pada data yang dinormalisasi dengan seleksi fitur menggunakan Algoritma Genetika. Parameter yang digunakan pada Algoritma Genetika yaitu menggunakan Probabilitas Crossover, Probabilitas Mutasi. Pada penelitian ini *Min-Max* digunakan untuk metode normalisasi data, dan untuk perhitungan akurasi menggunakan metode *Confusion Matrix*. Performasi terbaik dari parameter GA untuk Pc dan Pm menggunakan kombinasi 0.5 dan 0.2. Performasi populasi *size* terbaik adalah 26 dan iterasi atau *max* generasi terbaik ada pada 100 iterasi. Akurasi yang dihasilkan ketika menggunakan seluruh data menghasilkan akurasi sebesar 62%, 59% dengan data yang dinormalisasi, 64% dengan semua data menggunakan seleksi fitur Algoritma Genetika dan didapatkan akurasi sebesar 67% dengan menggunakan seleksi fitur Algoritma Genetika yang datanya dinormalisasi. Hasil ini memberi pengetahuan nilai rata-rata akurasi menunjukkan peningkatan dengan diterapkannya seleksi fitur Algoritma Genetika.

Kata kunci: klasifikasi, *genre* musik, *random forest*, algoritma genetika, seleksi fitur.

1. Pendahuluan

Mendengarkan musik sudah menjadi aktivitas yang biasa dilakukan orang-orang saat ini, apalagi saat ini mendengarkan musik sangat digemari anak-anak milenial yang bisa didengarkan melalui *smartphone*, TV, radio, tempat makan dan mencari lagu kesukaan di *youtube* atau aplikasi *streaming* lainnya (Shaleha, 2019). Ada berbagai jenis *genre* musik, yang paling populer diantaranya yaitu pop, dangdut, jazz, blues, disco, country, rock, metal, reggae, classical dan hiphop. Mengenali *genre* musik umumnya dapat diketahui dengan kemiripan melodi, irama, nada, tangga nada dan lainnya dan dengan cara langsung (Prasetyo, dkk 2019). Cara memudahkan mengenali perbedaannya, diperlukan sebuah *machine learning* yang dapat membantu mengatasi data yang banyak dengan menerapkan sebuah metode komputer. Tujuan dari *machine learning* ini dapat membuat sebuah keputusan ataupun melakukan prediksi dilakukan berdasarkan data yang sudah tersedia. Semakin bagus metode *machine learning*, akan semakin bagus juga hasil yang didapatkan.

Klasifikasi merupakan sebuah proses pembagian atau pengelompokkan sesuatu menjadi kategori atau jenis didasari kesamaan pola ataupun

sifat yang terdapat pada setiap data. Pada penelitian sebelumnya melakukan klasifikasi *genre* musik menggunakan *Naïve Bayes* oleh (Navisa, dkk. 2021) dengan menggunakan pola penelitian kerangka kerja dari CRISP-DM yang menghasilkan akurasi 58.91%. Penelitian sebelumnya oleh Sugianto (2019) melakukan klasifikasi *genre* music menggunakan CNN dari *Subsample Melspectogram* menghasilkan akurasi uji 63% dan akurasi *vouting* 71.87%. Penelitian terdahulu oleh Giri (2018) dengan metode K-NN menggunakan 11 fitur menghasilkan akurasi 44.8%.

Metode *Random Forest* termasuk metode klasifikasi. Menurut Supriyadi et al (2020) Metode *Random Forest* ini memiliki kelebihan yang mana bisa meningkat hasil dari jika terjadi kehilangan data atau *resisting outliers*. Selain itu kelebihan dari metode ini juga efisien untuk menyimpan sebuah data serta membantu perbaikan metode-metode dari *Decision Tree* yang mudah *overfitting*. Pada penelitian yang dilakukan oleh Wahyudi (2020) tentang pengklasifikasian *genre* musik menggunakan metode *Random Forest* dengan hasil akurasi yang didapatkan yaitu sebesar 63%, selanjutnya berfokus meningkatkan akurasi dengan menggunakan tahap fitur seleksi.

Seleksi fitur dilakukan agar dapat mengurangi dimensi data dan fitur yang tidak relevan. Menurut Gunawan, dkk (2020) Algoritma Genetika merupakan metode yang digunakan mampu menyeleksi fitur yang yang mampu meningkatkan hasil akurasi dari klasifikasi. Algoritma genetika adalah metode penyelesaian optimasi serta masalah pencarian (Pane, dkk. 2019).

Dari uraian diatas, penelitian ini akan menggunakan metode klasifikasi Random Forest dan dipadu dengan Algoritma Genetika dengan tujuan meningkatkan akurasi dari dataset GTZAN dengan menerapkan algoritma genetika sebagai seleksi fitur serta meningkatkan kinerja algoritma *Random Forest* dengan mengoptimalkan parameter menggunakan algoritma genetika.

2. Tinjauan Pustaka

2.1 Algoritma Genetika

Algoritma genetik merupakan algoritma yang terinspirasi oleh teorinya evolusi Darwin yang merupakan paradigma perangkat lunak yang relative baru dan populer (Adnyana. 2018). Metode Algoritma genetika ini dapat diterapkan dalam sistem data mining untuk mengklasifikasikan data agar mendapat informasi yang berguna pada data mining (Bakhtiar. 2020).

Kelebihan dari algoritmaa genetika ini mampu untuk menangani permasalahan kompleks dan parallel. Metode ini menangani berbagai macam optimasi tergantung dengan fungsi objektifnya linier atau tidak, seimbang atau tidak, berkesinambungan atau tidak atau dengan *random noise* (Putri & Ermatita, 2017).

Algoritma genetika ini mencakup pada pengkodean fungsi optimasi sebagai array yang berisi bit-bit atau karakter berbentuk *string* untuk menggambarkan kromosom, manipulasi operasi string dengan operator genetic, dan seleksi sesuai dengan kebugaran (fitness) yang bertujuan untuk menemukan solusi terbaik dan optimasi yang dihadapi. Operator utama algoritma genetika yaitu:

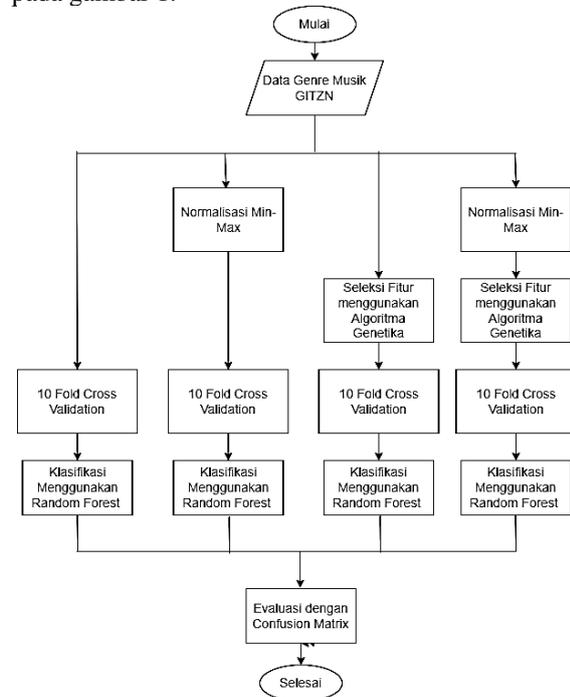
1. *Crossover* yaitu proses menukar bagian berdasarkan solusi (kromosom) menggunakan bagian "parent" lain buat membentuk jenis kromosom yg tidak sinkron yg mungkin sebagai solusi baru buat merampungkan permasalahan. Peran utamanya merupakan buat menaruh pencampuran solusi & konvergensi pada sub ruang (membentuk solusi yg baru).
2. Mutasi yaitu pergantian keliru satu bagian solusi yg dipilih secara acak, yg menaikkan keragaman berdasarkan populasi & membentuk prosedur
3. Seleksi fitness atau elitisme yaitu penggunaan solusi dengan nilai fitness yang tinggi untuk lulus ke generasi berikutnya, yang sering dilakukan dalam hal beberapa bentuk seleksi dari solusi terbaik (Prasetyo et al. 2018).

2.2 Random Forest

Random Forest pertama kali diterbitkan secara resmi oleh Leo Breiman pada tahun 2001. *Random Forest* dikembangkan untuk memperbaiki metode pohon keputusan yang cenderung *overfitted* (Widiastuti, 2018). *Random Forest* adalah metode yang dapat digunakan untuk klasifikasi, regresi atau tujuan lain. *Random Forest* sebuah metode yang bisa meningkatkan akurasi, karena dalam mengembangkan simpul pada setiap *node* yang dilakukan secara *random*. *Random Forest* digunakan untuk membangun metode pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan mengambil atribut dan data secara *random* sesuai dengan ketentuan (Ramli dan Sibanon. 2022)

3. Metode Penelitian

Alur penelitian dilakukan dengan langkah-langkah yang dirancang, agar penelitian dapat berjalan dengan baik. Alur penelitian dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

3.1 Pengumpulan Data

Dataset yang digunakan adalah *dataset* musik dari GITZAN yang berjumlah 1000 data memiliki 26 fitur dan terdiri dari 10 *genre* yang mencakup blueses, classical, country, disco, hiphcp, jass, metal, pcp, reggei dan rcck.

3.2 Pengolahan Data Awal

Sebelum melakukan klasifikasi pada *dataset* yang digunakan, harus dilakukan *preprocessing* terlebih dahulu terhadap data agar format data tersebut sesuai dengan yang diperlukan oleh algoritma klasifikasi yang digunakan

3.3 Permodelan

Pada tahap ini, dataset genre musik dilakukan klasifikasi menggunakan Random Forest dan Algoritma Genetika sebagai metode seleksi fitur dengan populasi *size* 26 dan generasi maksimum 100 dan dilakukan validasi dengan teknik *10-Fold Cross Validation*. Menurut Wei et al. (2019) *Cross Validation* membagi data menjadi data training dan data testing. Ten fold adalah definisi untuk nilai K, dimana nilai K = 10. Satu set data akan digunakan menjadi data training dan sisanya akan digunakan sebagai data testing dan seterusnya berurutan dan bergantian untuk setiap set nya.

Pada tahap ini terdapat empat tahap yang akan dilakukan. Pada tahap pertama dilakukan klasifikasi menggunakan Random Forest tanpa seleksi fitur, kemudian tahap kedua dilakukan klasifikasi menggunakan *Random Forest* dengan data yang sudah dinormalisasi tanpa seleksi fitur, kemudian selanjutnya adalah dilakukan klasifikasi menggunakan *Random Forest* dengan seleksi fitur, terakhir dilakukan klasifikasi menggunakan *Random Forest* dengan data yang sudah dinormalisasi dengan seleksi fitur.

3.4 Evaluasi

Pada tahap ini akan dievaluasi dengan menggunakan Confusion Matrix untuk mendapatkan hasil kinerja dari metode pengklasifikasian.

Tahap evaluasi bertujuan untuk mengetahui tingkat akurasi dari hasil penggunaan metode. Pada proses pengujian dikenal sebagai *Confusion Matrix* yang merepresentasikan kebenaran dari sebuah klasifikasi. *Confusion Matrix* berisi informasi actual dan prediksi pada sistem klasifikasi. Ada empat nilai yang dihasilkan di dalam tabel *confusion matrix*, di antaranya True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN). Ilustrasi tabel *confusion matrix* dapat dilihat pada tabel 1 berikut

Tabel 1. Ilustrasi tabel confusion matrix

		Nilai Aktual	
		Positive	Negative
Nilai Prediksi	Positive	TP	FP
	Negative	FN	N

4. Hasil dan Pembahasan

4.1. Hasil Pengumpulan Data

Data yang digunakan pada penelitian ini berisi 10 label ekstraksi genre music dengan jumlah 1000 data audio yang masing-masing berdurasi 30 detik. Dataset tersebut dari GITZAN atau GitHub dan dataset tersebut memiliki 26 fitur. Datasat yang sudah dimasukkan ke Google Collab yang digunakan dalam penelitian seperti pada tabel 1.

Tabel 2. Data Asli

No	c_stft	rmse	S_c	s_b	Sr_off	Zcr	mfcc1	mfcc20	label
1	0.430 8941	0.19622 1590	19.465 .657	19.799 .099	39.558 .677	0.0974 5368	- 67.770. 980	0.660037 07	blues
2	0.338 8961	0.25135 0194	21.414 .617	21.680 .156	46.279 .970	0.1051 5129	- 29.362. 093	130.481. 237	blues
3	0.320 2247	0.03895 6139	14.748 .034	18.133 .165	28.099 .413	0.0740 4604	- 305.396 .503	- 692.883. 467	classical
4	0.216 3284	0.04383 1587	12.112 .963	13.377 .625	21.047 .039	0.0709 3910	- 306.752 .771	133.380. 327	classical
5	0.390 5153	0.13019 9984	21.669 .047	22.776 .589	47.945 .627	0.1023 7580	- 94.067. 936	223.435. 319	country
6	0.382 1416	0.11459 7984	33.288 .039	31.118 .215	7.196. 340	0.1690 5151	- 83.556. 910	2.784.53 0.980	disco
....
1000	0.435 2317	0.17827 9251	31.677 .958	29.350 .139	6.730. 526	0.1510 0324	- 32.143. 401	5.848.71 7.131	rock

4.2. Pengolahan data awal.

Pengolahan data awal sebelum melakukan klasifikasi pada dataset yang digunakan, harus dilakukan *preprocessing* terlebih dahulu terhadap data agar format data tersebut sesuai dengan yang diperlukan oleh algoritma klasifikasi yang digunakan. Adapun langkah *preprocessing* yang digunakan yaitu sebagai berikut:

a. Pengkodean Label (*Label Encoding*)

Pengkodean Label atau *Label Encoding* merupakan proses pengubahan label ke dalam bentuk numerik. Proses pengkodean label ini membantu mengubah bentuk label pada dataset menjadi bentuk yang lebih mudah dibaca oleh mesin, karena sebagian besar algoritma pembelajaran mesin bekerja lebih baik dengan data numerik. Tabel 2 dapat dilihat perubahan label awal menjadi label dalam bentuk numerik pada dataset *Genre Musik*.

Tabel 3. Hasil dari label encoding

No	Label awal	Hasil label encoding
1	Blues	0
2	Classic	1
3	Country	2
4	Disco	3
5	Hip-Hop	4
6	Jazz	5
7	Raggae	6
8	Rock	7
9	Metal	8
10	Pop	9

b. Normalisasi

Normalisasi adalah teknik penskalaan di mana nilai-nilai diskalakan ulang sehingga akhirnya berkisar biasanya antara 0 dan 1. Teknik ini juga dikenal sebagai penskalaan *Min-Max*.

4.3 Seleksi fitur Algoritma Genetika

Adapun beberapa langkah yang terdapat dalam metode algoritma genetika untuk mengoptimasikan fitur seleksi pada *Random Forest*. Pada langkah awal adalah pengkodean yakni semua proses kodefikasi pada solusi dari permasalahan. Hasil dari tersebut terbentuk sebuah string yang merupakan *re-presentasi* pada suatu kromosom. Langkah selanjutnya adalah *selection* yakni ditentukannya kromosom yang akan tetap tinggal pada generasi berikutnya. Selanjutnya *crossover* dimana pada tahap ini menghasilkan kromosombaru yang menggantikan kromosom lama. Dan kemudian

mutase, tahap ini memungkinkan adanya kromosom baru secara *unpredictable*.

a. Inisialisasi Pupulasi

Menginisialisasi populasi dengan cara membangkitkan bilangan biner. Jadi pada masalah ini menggunakan pengkodean yang berbentuk bilangan biner. Ukuran populasi atau populasi size yang telah ditentukan dalam proses manualisasi ini adalah 26. Dan maksimum generasinya ada 100.

b. Evaluasi Populasi

Proses berikutnya yakni evaluasi populasi dengan ditentukannya parameter dari GA sebagai berikut:

- Pc atau Probabilitasn penyilangan/Cross over
- Pm/Probabilitas

Evaluasi populasi yang terbentuk dengan cara menghitung nilai fitness dengan menggunakan fungsi dari objektif yang dibentuk pada studi kasus.

c. Penyilangan pada kromosom

Model penyilangan yang digunakan terhadap percobaan ini yaitu metode penyilangan satu titik (*one point crossover*). Pertama pilih bangun bilangan acak sebanyak jumlah kromosom dalam populasi. Memilih bilangan acak lagi mulai dari 0 sampai dengan panjang kromosom / gen-1 disebut posisi *cut-point crossover* dimana posisi itu akan menentukan posisi gen yang akan disilangkanantara kromosom induk yang sebelumnya sudah dipilih.

d. Mutasi Algoritma Genetika

Kromosom sebelum mutase	Kromosom setelah mutase
1001011101001101101111010	1001011101001101101101010

Proses mutasi melakukan pergantian 1 gen acak dengan nilai baru sehingga tidak selalu menjamin jika setelah proses mutasi akan diperoleh kromosom dengan fitness yang terbaik. Proses yang dilakukan adalah menginversi nilai bit pada posisi tertentu yang terpilih secara acak (atau menggunakan skema tertentu pada kromosom, yang disebut inversi bit.

e. Evaluasi Populasi Baru

Setelah melewati proses mutasi, dan dihasilkan populasi baru yang mana adalah generasi 1. Pada penelitian ini akan ditentukan max generation atau iterasi yakni 100, maka proses algoritma genetika akan berhenti ketika sudah mencapai 100 generasi. Jika sudah didapatkan hasil akhir dari proses evaluasi populasi baru hanya saja sudah dipilih yang memiliki nilai fitness tertinggi dan yang memiliki nilai fitness dibawah rata-rata akan dihilangkam. Pada pengujian ini melakukan pengujian kombinasi parameter Pc dan Pm pada Algoritma genetika yang dapat dilihat pada tabel 3.

Tabel 4. Kombinasi Pengujian Parameter Algoritma Genetika

No.	Probabilitas Crossover	Probabilitas Mutation
1.	0.5	0.5
2.	0.3	0.5
3.	0.8	0.5
4.	1	1
5.	0.6	0.1
6.	1	0.5
7.	0.5	0.2

Pada parameter Algoritma Genetika terdapat nilai probabilitas crossover (Pc) dan probabilitas mutase (Pc) yang akan digunakan pada sistem. Performasi kombinasi parameter Pc dan Pm dapat dilihat pada tabel 4.

Tabel 5. Performansi dari pengaruh Pc dan Pm

No.	Pc	Pm	RF	Normalisasi-RF-GA
1.	0.1	0.5	62%	62%
2.	0.3	0.5	62%	63%
3.	0.8	0.5	63%	60%
4.	1	1	60%	61%
5.	0.6	0.1	63%	57%
6.	1	0.5	62%	61%
7.	0.5	0.2	64%	67%

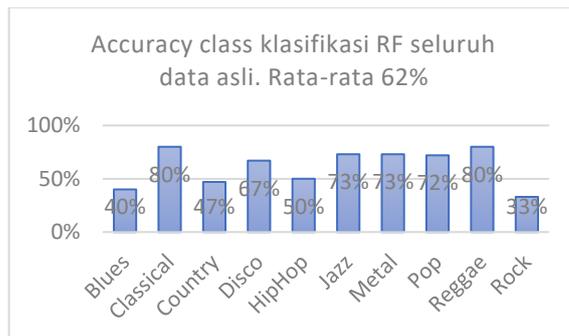
Pada Tabel 4, didapat hasil perhitungan rata-rata performansi dari kombinasi pada pengujian Algoritma Genetika yang menunjukkan kombinasi antara Pc dan Pm untuk menghasilkan populasi yang terbaik ditunjukkan dengan nilai rata-rata terbesar terdapat pada kombinasi nomor tujuh. Kombinasi ini yang nantinya akan digunakan pada proses Algoritma Genetika untuk seleksi fitur *Random Forest*.

4.3. Model Data Mining

Metode yang diusulkan Metode yang diusulkan mengintegrasikan algoritma genetika untuk seleksi fitur dan algoritma random forest yang telah dioptimasi untuk klasifikasi genre musik. Diawali dengan tahapan pembagian dataset menjadi data training dan data testing menggunakan 10 *k-fold validation*. Algoritma genetika diimplementasikan untuk seleksi fitur. Metode klasifikasi yang digunakan dalam penelitian ini adalah menggunakan metode Random Forest sebagai klasifikasi dataset genre musik untuk klasifikasi RF tanpa seleksi fitur Algoritma Genetika dan dengan menggunakan seleksi fitur Algoritma Genetika dengan menggunakan 100, 200, 300, 400, 500, 600, 700, 700, 800, 900, 1000 max generation.

a. Klasifikasi Random Forest dengan Seluruh Data

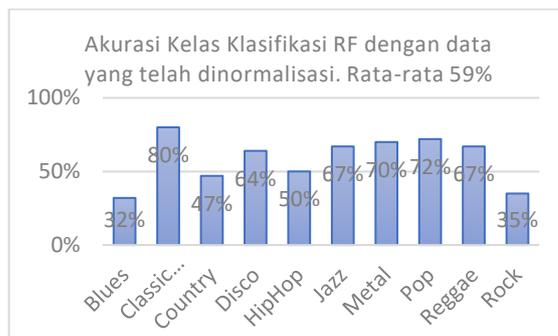
Dalam penelitian ini pengujian pertama adalah melakukan klasifikasi *Random Forest* dengan menggunakan seluruh data atau seluruh fitur yang ada pada dataset yang menghasilkan akurasi sebesar 62%. Dalam pengujiannya dengan menggunakan 10 genre music yaitu blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, dan rock. Pengujian ini juga menggunakan seluruh fitur yaitu C-STFT, RMSE, SR- OFF, SB, SC, ZCR, MFCC1 sampai MFCC20. Grafik hasil perbandingan akurasi setiap label dapat dilihat pada gambar 2.



Gambar 2. Perbandingan Accuracy Class disetiap labelnya dengan seluruh fitur

b. Klasifikasi Random Forest Dengan Data Normalisasi

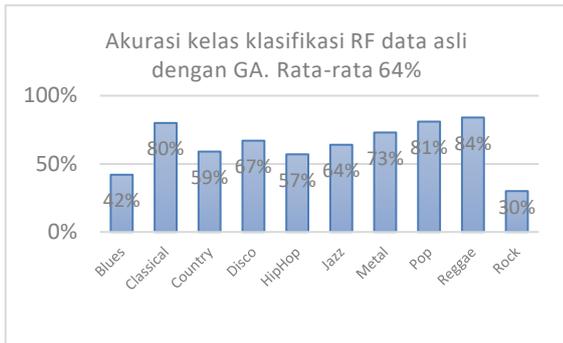
Pengujian kedua ini dengan menggunakan seluruh data dengan datanya sudah dilakukan tahap normalisasi. Diperoleh hasil nilai akurasi sebesar 59%. Pada akurasi setiap label yang didapatkan dari penelitian ini dapat dilihat pada gambar 3.



Gambar 3. Perbandingan akurasi class data yang telah dinormalisasi

c. Klasifikasi Random Forest dengan Seleksi Fitur Menggunakan Algoritma Genetika

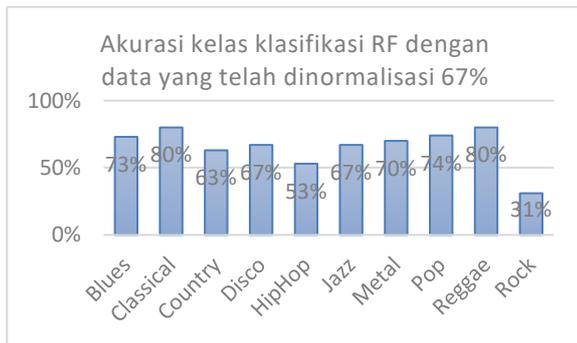
Pengujian ketiga ini dengan menggunakan seluruh data dengan seleksi fitur menggunakan Algoritma Genetika. Pada pengujian ini menggunakan hasil kombinasi parameter Pc dan Pm yang terbaik, yaitu kombinasi Pc = 0.5, dan Pm = 0.2 dengan percobaan iterasi/max generation 100. Hasil yang menggunakan seluruh data dengan seleksi fitur dapat dilihat pada gambar 4 sampai dengan gambar 4.



Gambar 4. Perbandingan Accuracy Class data asli dan seleksi fitur menggunakan Algoritma Genetika dengan 100 generasi

d. **Klasifikasi Random Forest Pada Data Normalisasi dengan Seleksi Fitur Menggunakan Algoritma Genetika**

Pengujian ketiga ini dengan data yang dinormalisasi dengan seleksi fitur menggunakan Algoritma Genetika. Pada pengujian ini menggunakan hasil kombinasi parameter Pc dan Pm yang terbaik, yaitu kombinasi Pc = 0.5, dan Pm = 0.2 dengan percobaan iterasi/max generation 100. Hasil yang menggunakan seluruh data dengan seleksi fitur dapat dilihat pada gambar 5.



Gambar 5. Perbandingan Accuracy Class data yang dinormalisasi dan seleksi fitur menggunakan Algoritma Genetika dengan 100 generasi

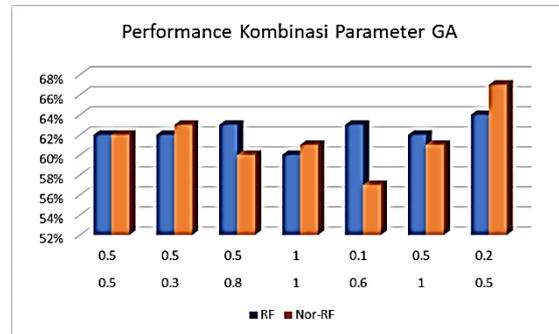
Dilihat pada grafik akurasi kelas diatas, dapat disimpulkan bahwa dari beberapa kali pengujian pada klasifikasi genre musik didapatkan akurasi tertinggi pada genre Classical dengan nilai akurasi 80%, 80%, 80%, 80%.

4.4 Evaluasi

Pada tahap klasifikasi digunakan metode Random Forest. Pada penelitian ini melakukan beberapa kali pengujian, yaitu mengklasifikasi dengan seluruh fitur, mengklasifikasi dengan data yang dinormalisasi, melakukan klasifikasi dengan data asli menggunakan tahap seleksi fitur Algoritma Genetika, dan mengklasifikasi pada data yang dinormalisasi dengan seleksi fitur menggunakan Algoritma Genetika.

Pada proses seleksi fitur melakukan beberapa percobaan kombinasi parameter crossover

dan mutasi, kombinasi Pc dan Pm menghasilkan performasi. Hasil performasi yang terbaik digunakan pada proses Algoritma Genetika untuk fitur seleksi Random Forest. Perbandingan Dari performansi kombinasi parameter GA dapat dilihat pada gambar 6:

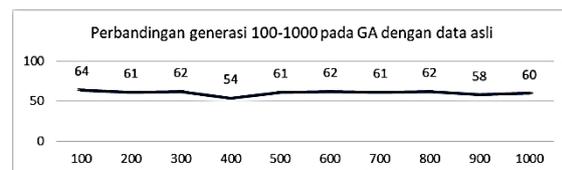


Gambar 6. Grafik Perbandingan Kombinasi Parameter Pc dan Pm pada GA

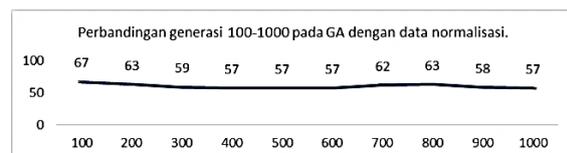
Selanjutnya pada ukuran populasi atau size population melakukan percobaan dengan 26, 30 dan 50 dan melakukan percobaan iterasi/max generasi 100, 200, 300, 400, 500, 600, 700, 800, 900, dan 1000. Performasi dari percobaan beberapa size populasi dapat dilihat pada Tabel 5 dan hasil yang diperoleh pada percobaan semua iterasi dengan data asli dan data yang sudah dinormalisasi dapat dilihat pada gambar 7 dan 8.

Tabel 6. Performance size populasi

populasi	RF-GA	Nor RFGA
27	64%	67%
30	62%	61%
50	59%	67%



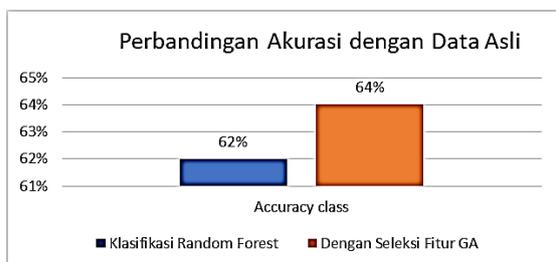
Gambar 7. Perbandingan generasi 100-1000 pada GA dengan data asli.



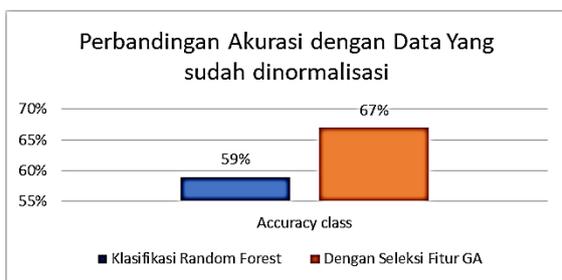
Gambar 8. Perbandingan generasi 100-1000 pada GA dengan data normalisasi.

Setelah melakukan beberapa tahapan, maka tahap evaluasi ini menggunakan *confusion matrix*, didapatkan akurasi sebesar 62% dengan klasifikasi menggunakan seluruh fitur, 59% dengan data yang

dinormalisasi, 64% dengan semua data menggunakan seleksi fitur Algoritma Genetika dan didapatkan akurasi sebesar 67% dengan menggunakan seleksi fitur Algoritma Genetika yang datanya dinormalisasi. Grafik hasil perbandingan akurasi klasifikasi *Random Forest* dengan Algoritma Genetika dapat dilihat pada gambar 9 dan 10.



Gambar 9. Grafik Perbandingan RF dan GA-RF



Gambar 10. Grafik Perbandingan data Normalisasi RF dan GA-RF

5. Kesimpulan dan Saran

Dari pengujian tersebut, akurasi klasifikasi menggunakan data asli dengan metode *Random Forest* menghasilkan rata-rata nilai sebesar 62% namun setelah data dinormalisasi hasilnya menurun menjadi 59%. Kemudian menerapkan seleksi fitur Algoritma Genetika, nilai rata-rata akurasi data asli meningkat menjadi 64% dan data yang dinormalisasi menjadi 67%. Hal ini menunjukkan bahwa penerapan seleksi fitur Algoritma Genetika mempengaruhi kinerja dari Klasifikasi *genre* musik yang mana dapat meningkatkan nilai akurasi.

Pada dataset GTZAN memiliki data mentah berupa file audio 16-bit Mono 22050Hz dalam format .wav. Saran yang diberikan berdasarkan penelitian ini adalah selanjutnya dapat mencoba algoritma deep learning lain ataupun melakukan penambahan fitur ekstraksi yang lain pada data audio seperti *spectral centroid and speed*, *spectral entropy*, *spectral flux*, dan *chroma vector* agar mendapatkan hasil yang maksimal. Selain itu juga mencoba menggunakan algoritma lain agar didapatkan hasil yang lebih baik lagi.

Daftar Pustaka:

Adnyana, IMB. (2018). *Implementasi Algoritma Genetika untuk Penjadwalan Asisten Dosen di STIKOM Bali*. Jurnal Sistem Dan Informatika Vol. 12, No. 2.

Bakhtiar, M. Y. (2020). *Klasifikasi Penelitian Dosen Menggunakan Naïve Bayes Classifier dan Algoritma Genetika*. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 5(2), 134.

Fajriyanto, A. S. & Supriyanto, C. 2018. *Optimasi prediksi tingkat produksi bawang merah nasional menggunakan metode Backpropagation Neural Network berbasis Algoritma Genetika*. Jurnal Teknologi Informasi, Vol. 13 No. 2.

G. A. V. M. Giri. 2018. *Klasifikasi Musik Berdasarkan Genre dengan Metode KNearest Neighbor*, Jurnal Ilmu Komputer, vol. 11, no. 2, Art. no. 2, Oct.

Navisa, S., Hakin, L., & Nabilah, A. 2021. *Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM*. Jurnal Sistem Cerdas (2021) Vol 04 - No 02 eISSN: 2622-8254 Hal: 114 – 125.

Pane, S. F., Awangga, R.M., Rahmadani, E.V & Permana, S. 2019. *Implementasi Algoritma Genetika Untuk Optimasi Pelayanan Kependudukan*. Jurnal Tekno Insentif. Vol. 13 No. 2.

Prasetyo, R. T., Rismayadi, A. A., & Anshori, I. F. (2018). *Implementasi Algoritma Genetika pada k-nearest neighbours untuk Klasifikasi Kerusakan Tulang Belakang*. Jurnal Informatika, 5(2), 186–194.

Prasetyo, P. D., Suta Wijaya, I. G. P., & Yudo Husodo, A. (2019). *Klasifikasi Genre Musik Menggunakan Metode Mel-Frequency Cepstrum Coefficients dan K-Nearest Neighbors Classifier*. Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTIIKA), 1(2), 189–197.

Putri, N. M. E., & Ermatita. (2017). *Review : Data Mining menggunakan Algoritma Genetika*. *Prosiding Annual Research Seminar 2017*, 3(1), 235–238.

Ramli, R.G., & Sibaroni, Y. 2022. *Klasifikasi Topik Twitter menggunakan Metode Random Forest dan Fitur Ekspansi Word2Vec*. *e-Proceeding of Engineering: Vol.9, No.1*

Shaleha, RRA. 2019. *Do Re Mi: Psikologi, Musik, dan Budaya*. Buletin Psikolog Vol. 27, No. 1, 43-51.

Sugianto. 2019. *Klasifikasi Genre Musik dari Subsample Melspectrogram menggunakan Convolutional Neural Network*. Universitas Telkom, S1 Informatika, Bandung.

Supriyadi, R., Gata, W., Maulidah, N., & Fauzi, A. (2020). *Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah*. *E-Bisnis : Jurnal Ilmiah Ekonomi Dan Bisnis*, 13(2), 67–75.

Wahyudi. 2020. *Implementasi Metode Random Forest Untuk Klasifikasi Genre Music*. Universitas Lambung Mangkurat, Banjarbaru.

- Wei, H., Hu, C., Chen, S., Xue, Y., & Zhang, Q. 2019. *Establishing a software defect prediction model via effective dimension reduction. Information Sciences*. Volume 28, 27: 1088-1099.
- Widiastuti, J. 2018. *Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode Random Forest dengan Teknik Sampling Kelas Imbalanced. Skripsi Program Sarjana Statistika, Universitas Islam Indonesia, Yogyakarta.*