

PERBANDINGAN METODE EXTREME GRADIENT BOOSTING DAN METODE DECISION TREE UNTUK KLASIFIKASI GENRE MUSIK

Sally Lutfiani¹, Triando Hamonangan Saragih², Friska Abadi³, Mohammad Reza Faisal⁴, Dwi Kartini⁵

^{1,2,3,4,5}Ilmu Komputer, Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Jalan Jenderal Ahmad Yani KM 36, Banjarbaru, Kalimantan Selatan 70714

¹sallylutfiani27@gmail.com, ²triando.saragih@ulm.ac.id, ³friska.abadi@ulm.ac.id, ⁴reza.faisal@ulm.ac.id, ⁵dwikartini@ulm.ac.id

Abstrak

Musik merupakan sebuah “bahasa” yang mampu dimengerti dan dipahami oleh semua orang. Dalam musik sendiri, terdapat banyak *genre* musik yang berkembang yang dipengaruhi oleh budaya dari daerah-daerah yang berbeda-beda, seperti musik *jazz*, *reggae*, *pop*, *rock*, *punk*, dan masih banyak lagi *genre* musik yang ada seperti musik tradisional. Bertambahnya jumlah musik dalam bentuk digital secara pesat menyebabkan pemberian label *genre* secara manual menjadi tidak efektif. Pemberian label *genre* secara otomatis dapat dilakukan dengan menerapkan algoritma kecerdasan buatan yaitu salah satunya klasifikasi yang dapat mengelompokkan jenis musik berdasarkan *genre* dengan menggunakan fitur-fitur musik. Salah satu metode klasifikasi yang cukup sering digunakan adalah metode *Extreme Gradient Boosting*. Selain itu ada juga metode yang sering digunakan dalam melakukan klasifikasi yaitu metode *Decision Tree* yang merupakan metode pohon keputusan. Pada penelitian ini melakukan perbandingan tingkat akurasi metode klasifikasi *Extreme Gradient Boosting* dan metode klasifikasi *Decision Tree* dengan melakukan pengujian parameter menggunakan nilai parameter terbaik yang didapatkan. Berdasarkan hasil penelitian metode klasifikasi *Extreme Gradient Boosting* dengan pengujian parameter menggunakan nilai parameter terbaik yang didapatkan menghasilkan kinerja akurasi yang lebih baik dibandingkan dengan metode klasifikasi *Decision Tree* yaitu sebesar 72% karena pada metode *Extreme Gradient Boosting* ini mampu meminimalisir galat dengan menggunakan data residu atau kesalahan prediksi pada model sebelumnya sehingga sehingga model pembelajaran mesin mendapatkan hasil akurasi yang lebih baik.

Kata kunci : Musik, *Genre*, *Extreme Gradient Boosting*, *Decision Tree*, Klasifikasi.

1. Pendahuluan

Musik merupakan sebuah “bahasa” yang mampu dimengerti dan dipahami oleh semua orang. Semua perasaan manusia mampu dicurahkan dalam alunan musik bernada pentatonis maupun diatonis yang diaransemen sesuai suasana hati sang pencipta lagu, baik lagu bernuansa senang, sedih, bersemangat, horor, dan lainnya. Dalam musik sendiri, terdapat banyak *genre* musik yang berkembang yang dipengaruhi oleh budaya dari daerah-daerah yang berbeda-beda, seperti musik *jazz*, *reggae*, *pop*, *rock*, *punk*, dan masih banyak lagi *genre* musik yang ada seperti musik tradisional. Bertambahnya jumlah musik dalam bentuk digital secara pesat menyebabkan pemberian label *genre* secara manual menjadi tidak efektif. Pemberian label *genre* secara otomatis dapat dilakukan dengan menerapkan algoritma kecerdasan buatan yaitu salah satunya klasifikasi yang dapat mengelompokkan jenis musik berdasarkan *genre* dengan menggunakan fitur-fitur musik (Giri, 2017).

Menurut Rahmawati (2016) klasifikasi *genre* musik penting dilakukan agar memudahkan masyarakat yang menyukai musik untuk memilih dan mencari musik

berdasarkan *genre* nya. Pengklasifikasian *genre* musik secara otomatis dapat menjadi hal yang sangat membantu dalam pengembangan sistem temu-kembali untuk data audio, maupun untuk penggunaan pengklasifikasian musik bagi individu.

Wahyudi (2020) melakukan pengklasifikasian *genre* musik dengan mengimplementasikan *Random Forest* sebagai metode klasifikasinya. Hasil dari penelitian ini mendapatkan akurasi tertinggi sebesar 66,3%. Peneliti mengemukakan bahwa akurasi dapat ditingkatkan dengan menggunakan metode pengklasifikasian yang lain agar mendapatkan nilai akurasi yang lebih baik.

Salah satu metode klasifikasi yang cukup sering digunakan adalah metode *Extreme Gradient Boosting*. Metode ini seringkali digunakan karena efisiensi dan skalabilitasnya untuk memecahkan beragam masalah klasifikasi ataupun regresi (Chen, 2016). Pada penelitian yang dilakukan oleh Pamina et al., (2019) yang melakukan perbandingan tiga metode klasifikasi yaitu *K-Nearest Neighbour* (K-NN), *Random Forest*, dan *Extreme Gradient Boosting* untuk memprediksi *customer churn* pada perusahaan telekomunikasi. Hasil dari penelitian ini menunjukkan bahwa kinerja dari metode *Extreme Gradient Boosting* mendapatkan tingkat akurasi

paling tinggi sebesar 79,8% yang lebih baik dibandingkan dengan metode lainnya.

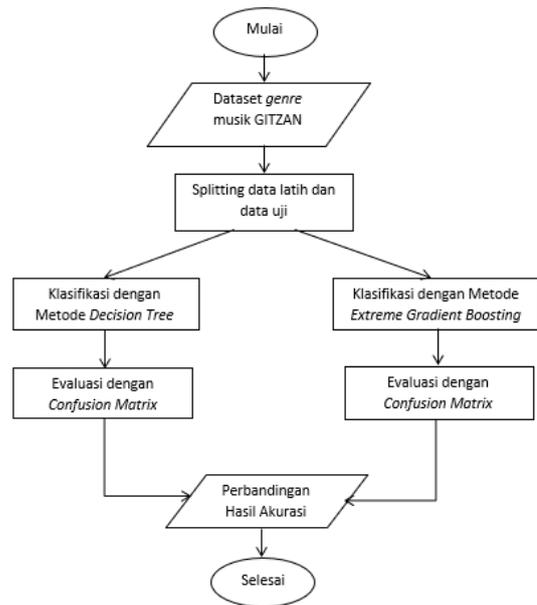
Selain itu, metode *Decision Tree* juga sering digunakan dalam melakukan klasifikasi. Menurut Achmad (2012) metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Pada penelitian lain oleh Wibowo (2015) yang melakukan perbandingan kinerja metode klasifikasi *Decision Tree* algoritma CART dengan metode *Random Forest*. Berdasarkan hasil dari penelitian ini menunjukkan bahwa kinerja metode *Decision Tree* lebih baik dibandingkan dengan metode *Random Forest* dengan hasil tingkat akurasi sebesar 90%.

Berdasarkan paparan beberapa penelitian sebelumnya yang telah disebutkan, maka dapat diketahui bahwa kinerja metode *Extreme Gradient Boosting* dan metode *Decision Tree* termasuk metode terbaik dalam klasifikasi karena menghasilkan tingkat akurasi yang cukup baik. Oleh karena itu, pada penelitian ini ingin melakukan perbandingan tingkat akurasi metode *Extreme Gradient Boosting* dan metode *Decision Tree*. Karena kedua metode ini termasuk dalam rumpun keluarga pohon atau *ensemble learning*, maka dengan dilakukannya perbandingan antara kedua metode tersebut diharapkan dapat mengetahui kinerja metode mana yang lebih baik dalam mengklasifikasikan genre musik dan dapat menghasilkan tingkat akurasi yang lebih baik dari penelitian sebelumnya. Dalam melakukan perbandingan kedua metode tersebut juga melakukan uji parameter untuk meningkatkan hasil akurasi yang didapatkan.

Data dari penelitian ini memakai *dataset* GITZAN seperti *dataset* penelitian yang dilakukan oleh Wahyudi (2020). Pada penelitian ini, akan menggunakan seluruh fitur yang ada di dalam *dataset* tersebut dengan 10 *genre* musik yang akan digunakan diantaranya *blues, classical, country, disco, hiphop, jazz, metal, pop, reggae* dan *rock* dan data pada *dataset* berjumlah 1000 data.

2. Metode

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

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data *genre* musik yang diperoleh dari GITZAN yang berjumlah 1000 data dengan 26 fitur dan 10 label. Data ini juga digunakan dalam penelitian yang dilakukan oleh Amini dkk pada tahun 2022.

2.2 Klasifikasi

Pada tahap ini, terdapat empat pengujian yang akan dilakukan. Pada pengujian pertama dilakukan klasifikasi menggunakan metode *Extreme Gradient Boosting* dengan parameter *default* atau bawaan, kemudian pengujian kedua dilakukan menggunakan metode *Extreme Gradient Boosting* dengan menggunakan parameter terbaik yang didapatkan pada pengujian. Kemudian selanjutnya dilakukan pengujian menggunakan metode *Decision Tree* dengan parameter *default* atau bawaan, dan yang terakhir dilakukan pengujian menggunakan metode *Decision Tree* menggunakan parameter terbaik yang didapatkan saat pengujian.

2.3 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*

	Positive	Negative
Positive	TP	FP
Negative	FN	N

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data *genre* musik yang diperoleh dari GITZAN yang berjumlah 1000 data dengan 26 fitur dan 10 label. Data ini juga digunakan dalam penelitian yang dilakukan oleh Amini dkk pada tahun 2022

3.2 Klasifikasi

Penelitian ini menggunakan dua metode klasifikasi yaitu metode *Xtreme Gradient Boosting* dan metode *Decision Tree*. Pembagian data *training* (data latih) dan data *testing* (data uji) dilakukan menggunakan metode *splitting* data dengan perbandingan rasio 90:10.

Pada proses klasifikasi juga dilakukan uji parameter dengan menggunakan beberapa parameter. Parameter yang diuji pada proses klasifikasi menggunakan metode *Extreme Gradient Boosting* yaitu (i) parameter *max depth* merupakan tingkat dalamnya pohon, (ii) parameter *gamma* berfungsi untuk meminimumkan pengurangan kerugian, (iii) parameter *learning rate* berfungsi untuk membantu mempersingkat langkah dalam pembaruan model, dan (iv) parameter *n estimator* merupakan banyaknya pohon yang digunakan untuk proses klasifikasi. Parameter yang diuji pada proses klasifikasi menggunakan metode *Decision Tree* yaitu (i) parameter *max depth* merupakan tingkat dalamnya pohon, (ii) parameter *min_samples_split* merupakan jumlah minimum sampel yang diperlukan untuk membagi node internal, (iii) parameter *min_samples_leaf* merupakan jumlah minimum sampel yang diperlukan untuk berada di simpul daun, (iv) parameter *max_features* merupakan jumlah fitur yang perlu dipertimbangkan saat mencari pembagian terbaik.

3.3 Extreme Gradient Boosting

Klasifikasi pertama yang dilakukan setelah pembagian data dengan *splitting* data berdasarkan perbandingan rasio 90:10 yaitu menggunakan metode *Extreme Gradient Boosting* dan melakukan uji parameter untuk mengetahui perbedaan hasil akurasi dari parameter yang diuji yaitu parameter *max depth*, *learning rate*, *n estimator*, dan *gamma*.

a. Extreme Gradient Boosting Default

Pada klasifikasi pertama hasil akurasi yang dihasilkan oleh peneliti menggunakan metode klasifikasi *Extreme Gradient Boosting* dapat dilihat pada tabel 4.

Tabel 2 Hasil Akurasi parameter bawaan (*default*)

Nilai Rasio	Data Training	Data Testing	Hasil Akurasi	Total Data
90:10	900	100	66%	1000

b. Extreme Gradient Boosting Uji Parameter

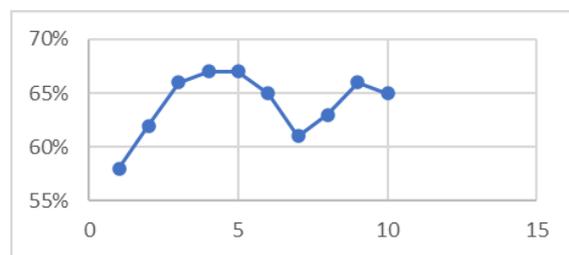
1. Parameter Max Depth

Pada uji parameter *max depth* peneliti melakukan uji parameter *max depth*= 1- 10, dengan jarak loncat 1 (1,2,3,...,10) dengan parameter *learning rate*, *n estimator* dan *gamma* menggunakan nilai default. Hasil dari uji parameter *max depth* dapat dilihat pada tabel 5.

Tabel 3 Hasil Akurasi Parameter *Max Depth*

Nilai Parameter	Akurasi
1	58 %
2	62 %
3	66 %
4	67 %
5	67 %
6	65 %
7	61 %
8	63 %
9	66 %
10	65 %

Pada hasil uji parameter diatas nilai parameter terbaik dari *max depth* adalah 5 dengan hasil akurasi 67%, dimana hasil akurasi dari uji tersebut dapat dilihat pada gambar 2 grafik hasil akurasi parameter *max depth*.



Gambar 2 Grafik hasil akurasi parameter *max depth*

2. Parameter Learning Rate

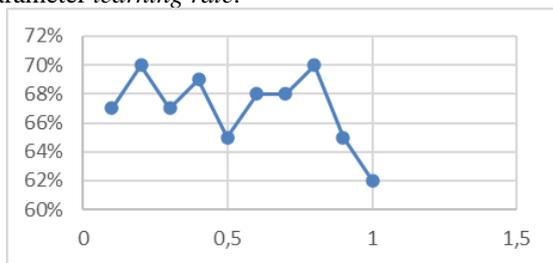
Pada uji parameter *learning rate* peneliti melakukan uji parameter *learning rate*= 0.1-1, dengan jarak loncat 0.1 (0.1,0.2,0.3,...,1) dengan parameter *max depth* menggunakan nilai parameter terbaik yang sudah didapatkan dengan parameter *n estimator* dan *gamma*

menggunakan nilai default. Hasil dari uji parameter *learning rate* dapat dilihat pada tabel 6.

Tabel 4 Hasil akurasi parameter *learning rate*

Nilai Parameter	Akurasi
0,1	67 %
0,2	70 %
0,3	67 %
0,4	69 %
0,5	65 %
0,6	68 %
0,7	68 %
0,8	70 %
0,9	65 %
1	62 %

Pada hasil uji parameter diatas nilai parameter terbaik dari *learning rate* yang digunakan adalah 0,2 dengan hasil akurasi 70%, dimana hasil akurasi dari uji tersebut dapat dilihat pada gambar 3 grafik hasil akurasi parameter *learning rate*.



Gambar 3 Grafik hasil akurasi parameter *learning rate*

3. Parameter *N-Estimator*

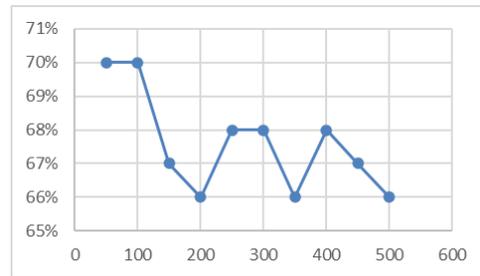
Pada uji parameter *n-estimator* peneliti melakukan uji parameter *n estimator*= 50-500, dengan jarak loncat 50 (50,100,150,...,500) dengan parameter *max depth* dan *learning rate* menggunakan nilai parameter terbaik yang sudah didapatkan dan parameter *gamma* menggunakan nilai parameter *default*. Hasil dari uji parameter *n estimator* dapat dilihat pada tabel 5.

Tabel 5 Hasil akurasi parameter *n-estimator*

Nilai Parameter	Akurasi
50	70 %
100	70 %
150	67 %
200	66 %
250	68 %
300	68 %
350	66 %
400	68 %
450	67 %
500	66 %

Pada hasil uji parameter diatas terdapat nilai parameter terbaik dari *n estimator* adalah 50 yang

menghasilkan akurasi 70%, dimana hasil akurasi dari uji tersebut dapat dilihat pada gambar 4 grafik hasil akurasi percobaan parameter *n estimator*.



Gambar 4 grafik hasil akurasi percobaan parameter *n estimator*

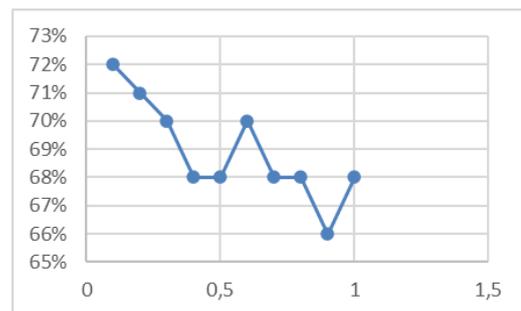
4. Parameter *Gamma*

Pada uji parameter *gamma* peneliti melakukan uji parameter *gamma*= 0.1-1, dengan jarak loncat 0.1 (0.1,0.2,0.3,...,1) dengan parameter *max depth*, *learning rate* dan *n estimator* menggunakan nilai parameter terbaik yang sudah didapatkan dari pengujian sebelumnya. Hasil dari uji parameter *gamma* dapat dilihat pada tabel 6.

Tabel 6 Hasil akurasi percobaan parameter *gamma*

Nilai Parameter	Akurasi
0,1	72 %
0,2	71 %
0,3	70 %
0,4	68 %
0,5	68 %
0,6	70 %
0,7	68 %
0,8	68 %
0,9	66 %
1	68 %

Pada hasil uji parameter diatas nilai parameter terbaik dari *gamma* yang digunakan yaitu 0,1 dengan hasil akurasi 72%. Hasil akurasi dari uji tersebut dapat dilihat pada gambar 5 grafik hasil akurasi parameter *gamma*.



Gambar 5 grafik hasil akurasi parameter *gamma*

3.4 Decision Tree

Klasifikasi kedua yang dilakukan setelah pembagian data dengan *splitting* data berdasarkan perbandingan rasio 90:10 yaitu menggunakan metode *Decision Tree* dan melakukan uji parameter untuk mengetahui perbedaan hasil akurasi dari parameter yang diuji yaitu parameter *max depth*, *min_samples_split*, *min_samples_leaf* dan *max features*.

a. Decision Tree Default

Pada klasifikasi pertama hasil akurasi yang dihasilkan oleh peneliti menggunakan metode klasifikasi *Decision Tree* dapat dilihat pada tabel 7.

Tabel 7 Hasil akurasi parameter bawaan (*default*)

Nilai Rasio	Data Training	Data Testing	Hasil Akurasi	Total Data
90:10	900	100	40%	1000

b. Decision Tree Uji Parameter

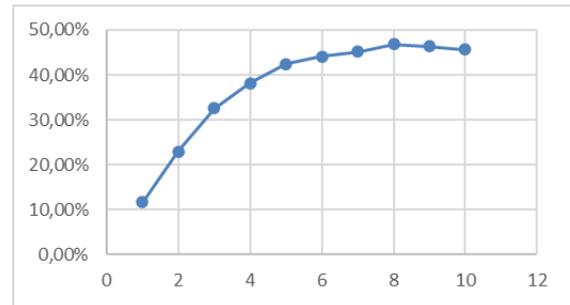
1. Parameter Max Depth

Pada uji parameter *max depth* peneliti melakukan uji parameter *max depth*= 1- 10, dengan jarak loncat 1 (1,2,3,...,10) dengan parameter *min samples split*, *min samples leaf* dan *max features* menggunakan nilai *default*. Hasil dari uji parameter *max depth* dapat dilihat pada tabel dibawah ini.

Tabel 8 Hasil akurasi parameter *max depth*

Nilai Parameter	Akurasi
1	11,7 %
2	22,9 %
3	32,6 %
4	38,2 %
5	42,4 %
6	44,1 %
7	45,2 %
8	46,8 %
9	46,4 %
10	45,6 %

Pada hasil uji parameter diatas nilai parameter terbaik dari *max depth* adalah 8 dengan hasil akurasi 46,8%, dimana hasil akurasi dari uji tersebut dapat dilihat pada gambar 6 grafik hasil akurasi parameter *max depth*.



Gambar 6 Grafik hasil akurasi parameter *max depth*

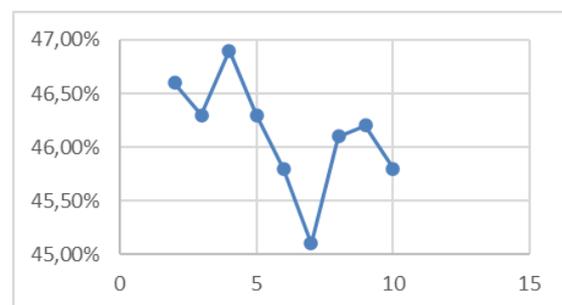
2. Parameter Min Samples Split

Pada uji parameter *min samples split* peneliti melakukan uji parameter *min samples split* = 1- 10, dengan jarak loncat 1 (1,2,3,...,10) dengan parameter *max depth* menggunakan nilai parameter terbaik yang sudah didapatkan pada pengujian sebelumnya dan parameter *min samples leaf* dan *max features* menggunakan nilai *default*. Hasil dari uji parameter *min samples split* dapat dilihat pada tabel dibawah ini.

Tabel 9 Hasil akurasi parameter min samples split

Nilai Parameter	Akurasi
1	Error
2	46,6 %
3	46,3 %
4	46,9 %
5	46,3 %
6	45,8 %
7	45,1 %
8	46,1 %
9	46,2 %
10	45,8 %

Pada hasil uji parameter diatas nilai parameter terbaik dari *min samples split* adalah 4 dengan hasil akurasi 46,9%, dimana hasil akurasi dari uji tersebut dapat dilihat pada gambar 7 grafik hasil akurasi parameter *min samples split*.



Gambar 7 Grafik hasil akurasi parameter min samples split

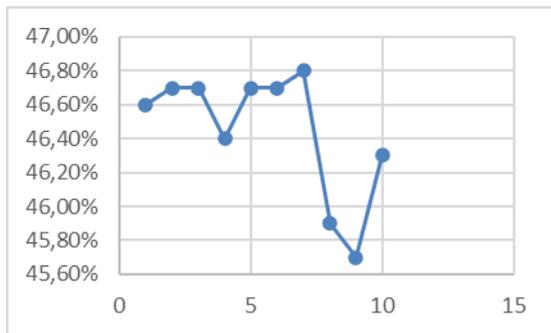
3. Parameter Min Samples Leaf

Pada uji parameter *min samples leaf* peneliti melakukan uji parameter *min samples leaf* = 1- 10, dengan jarak loncat 1 (1,2,3,...,10) dengan parameter *max depth* dan *min samples split* menggunakan nilai parameter terbaik yang sudah didapatkan pada pengujian sebelumnya dan parameter *max features* menggunakan nilai *default*. Hasil dari uji parameter *min samples leaf* dapat dilihat pada tabel 10.

Tabel 10 Hasil akurasi parameter *min samples leaf*

Nilai Parameter	Akurasi
1	46,6 %
2	46,7 %
3	46,7 %
4	46,4 %
5	46,7 %
6	46,7 %
7	46,8 %
8	45,9 %
9	45,7 %
10	46,3 %

Pada hasil uji parameter diatas nilai parameter terbaik dari *min samples leaf* adalah 7 dengan hasil akurasi 46,7%, dimana hasil akurasi dari uji tersebut dapat dilihat pada gambar 8 grafik hasil akurasi parameter *min samples leaf*.



Gambar 8 Grafik hasil akurasi parameter *min samples leaf*

4. Parameter Max Features

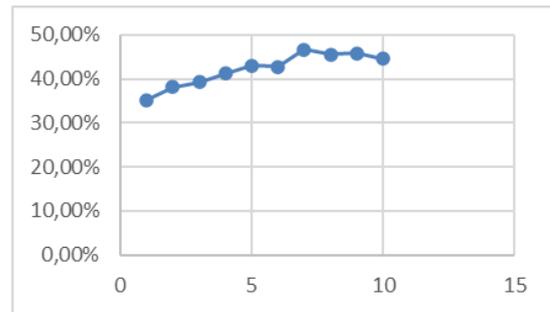
Pada uji parameter *max features* peneliti melakukan uji parameter *max features* = 1- 10, dengan jarak loncat 1 (1,2,3,...,10) dengan parameter *max depth*, *min samples split*, dan *min samples leaf* menggunakan nilai parameter terbaik yang sudah didapatkan pada pengujian sebelumnya. Hasil dari uji parameter *max features* dapat dilihat pada tabel 11.

Tabel 11 Hasil akurasi parameter *max features*

Nilai Parameter	Akurasi
1	35,2 %
2	38,1 %
3	39,3 %
4	41,2 %
5	43 %
6	42,8 %

7	46,7 %
8	45,6 %
9	45,9 %
10	44,6 %

Pada hasil uji parameter diatas nilai parameter terbaik dari *max features* adalah 7 dengan hasil akurasi 46,7% dimana hasil akurasi dari uji tersebut dapat dilihat pada gambar 9 grafik hasil akurasi parameter *max features*.



Gambar 9 grafik hasil akurasi parameter *max features*

3.5 Evaluasi

Klasifikasi pertama dilakukan menggunakan metode *Extreme Gradient Boosting* dengan parameter bawaan, setelah proses klasifikasi selesai, data hasil prediksi dilakukan pengujian menggunakan *confusion matrix* sehingga diperoleh akurasi sebesar 66%. Tabel *confusion matrix* klasifikasi *Extreme Gradient Boosting* dapat dilihat pada Tabel 12.

Pada tabel 12 merupakan hasil *confusion matrix* untuk metode *Extreme Gradient Boosting* yang menggunakan parameter bawaan. Untuk blues ada 9 jumlah data yang benar, classical 5, country 6, disco 6, hiphop 4, jazz 11, metal 11, pop 5, reggae 7, sementara rock hanya 2 data yang benar terklasifikasi oleh sistem. Berdasarkan Tabel 13 tersebut dilakukan kinerja klasifikasi berupa akurasi.

Tabel 13 Perhitungan akurasi klasifikasi *extreme gradient boosting*

Perhitungan Akurasi	
Rumus	$\frac{\text{jumlah data yang benar}}{\text{jumlah seluruh data}} \times 100\%$
Hasil	$\frac{9 + 5 + 6 + 6 + 4 + 11 + 11 + 5 + 7 + 2}{100} \times 100\%$ = 66%

Dilakukan evaluasi *Extreme Gradient Boosting* dengan menggunakan data *Genre Musik* yang menghasilkan akurasi sebesar 66% yang didapat dengan menghitung jumlah data prediksi yang terklasifikasi secara benar dan dibagi dengan keseluruhan data prediksi.

Klasifikasi selanjutnya yaitu metode *Extreme Gradient Boosting* dengan melakukan uji parameter menggunakan nilai parameter terbaik yang didapatkan

dan data *Genre Musik*, setelah proses klasifikasi selesai, data hasil prediksi akan dilakukan pengujian dengan menggunakan *confusion matrix* sehingga diperoleh hasil akurasi sebesar 72%. Dibawah ini terdapat Tabel 14 *Confusion Matrix* Klasifikasi *Extreme Gradient Boosting* Uji Parameter.

Pada tabel 14 atas merupakan hasil *confusion matrix* untuk metode *Extreme Gradient Boosting* yang menggunakan nilai parameter terbaik setelah dilakukan pengujian. Untuk blues ada 10 jumlah data yang benar, classical 6, country 8, disco 6, hiphop 4, jazz 10, metal 11, pop 5, reggae 7, sementara rock 5 data yang benar terklasifikasi oleh sistem. Berdasarkan Tabel 15 tersebut dilakukan kinerja klasifikasi berupa akurasi.

Tabel 15 Perhitungan akurasi klasifikasi *extreme gradient boosting* uji parameter

Perhitungan Akurasi	
Rumus	$\frac{\text{Jumlah data yang benar}}{\text{Jumlah seluruh data}} \times 100\%$
Hasil	$\frac{10 + 6 + 8 + 6 + 4 + 10 + 11 + 5 + 7 + 5}{100} = 72\%$

Dilakukan evaluasi *Extreme Gradient Boosting* uji parameter dengan menggunakan data *Genre Musik* yang menghasilkan akurasi sebesar 72% yang didapat dengan menghitung jumlah data prediksi yang terklasifikasi secara benar dan dibagi dengan keseluruhan data prediksi.

Penelitian selanjutnya yaitu menggunakan metode *Decision Tree* sebagai metode klasifikasi untuk data *Genre Musik*. Setelah proses klasifikasi selesai, data hasil prediksi dilakukan pengujian menggunakan *confusion matrix* sehingga diperoleh akurasi sebesar 40%. Tabel *confusion matrix* klasifikasi *Decision Tree* dapat dilihat pada Tabel 16.

Pada tabel 16 merupakan hasil *confusion matrix* untuk metode *Decision Tree* yang menggunakan nilai parameter bawaan. Untuk blues ada 6 jumlah data yang benar, classical 4, country 6, disco 2, hiphop 2, jazz 4, metal 7, pop 3, reggae 4, sementara rock hanya 2 data yang benar terklasifikasi oleh sistem. Berdasarkan Tabel 17 tersebut dilakukan kinerja klasifikasi berupa akurasi.

Tabel 17 Perhitungan akurasi klasifikasi *decision tree*

Perhitungan Akurasi	
Rumus	$\frac{\text{Jumlah data yang benar}}{\text{Jumlah seluruh data}} \times 100\%$
Hasil	$\frac{6 + 4 + 6 + 2 + 2 + 4 + 7 + 3 + 4 + 2}{100} = 40\%$

Dilakukan evaluasi *Decion Tree* uji parameter dengan menggunakan data *Genre Musik* yang menghasilkan akurasi sebesar 40% yang didapat dengan menghitung jumlah data prediksi yang terklasifikasi secara benar dan dibagi dengan keseluruhan data prediksi.

Klasifikasi selanjutnya yaitu metode *Decision Tree* dengan melakukan uji parameter menggunakan nilai parameter terbaik yang didapatkan dan data *Genre Musik*. Setelah proses klasifikasi selesai, data hasil prediksi akan dilakukan pengujian dengan menggunakan *confusion matrix* sehingga diperoleh hasil akurasi sebesar 51%. Dibawah ini terdapat Tabel 18 *Confusion Matrix* Klasifikasi *Decision Tree* Uji Parameter.

Pada tabel 18 merupakan hasil *confusion matrix* untuk metode *Decision Tree* yang menggunakan nilai parameter terbaik setelah dilakukan pengujian. Untuk blues ada 9 jumlah data yang benar, classical 4, country 7, disco 2, hiphop 3, jazz 6, metal 8, pop 3, reggae 7, sementara rock 2 data yang benar terklasifikasi oleh sistem. Berdasarkan Tabel 19 tersebut dilakukan kinerja klasifikasi berupa akurasi.

Tabel 19 Perhitungan akurasi klasifikasi *decision tree* uji parameter

Perhitungan Akurasi	
Rumus	$\frac{\text{Jumlah data yang benar}}{\text{Jumlah seluruh data}} \times 100\%$
Hasil	$\frac{9 + 4 + 7 + 2 + 3 + 6 + 8 + 3 + 7 + 2}{100} = 51\%$

Dilakukan evaluasi *Decision Tree* uji parameter dengan menggunakan data *Genre Musik* yang menghasilkan akurasi sebesar 51% yang didapat dengan menghitung jumlah data prediksi yang terklasifikasi secara benar dan dibagi dengan keseluruhan data prediksi.

4. Kesimpulan

Makalah ini membandingkan dua metode populer pembelajaran mesin yaitu *Decision Tree* dan *Extreme Gradient Boosting*. Hasil uji parameter terbaik dari *Decision Tree* dan *Extreme Gradient Boosting* masing-masing sekitar 51% dan 72%. Hasil ini lebih baik tanpa adanya penyesuaian uji parameter yang diperoleh peningkatan sekitar 11% dan 6% untuk masing-masing metode. Pada kasus dataset *genre musik*, *Extreme Gradient Boosting* menghasilkan performa akurasi 12% lebih tinggi dibandingkan *Decision Tree*. Oleh karenanya, *Extreme Gradient Boosting* direkomendasikan sebagai *baseline* dalam klasifikasi genre musik.

Adapun saran yang dapat disampaikan untuk penelitian selanjutnya adalah untuk melakukan penelitian menggunakan metode *ensemble learning* yang lain sehingga dapat diketahui perbandingan hasil klasifikasi dengan metode yang berbeda, juga bisa melakukan penelitian dengan *dataset* yang berbeda dengan jumlah yang lebih besar atau *dataset* yang bersifat kategorik.

Daftar Pustaka:

- Achmad, Budanis Dwi Meilani., & Slammat, Fauzi. 2012, *Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree*, *Jurnal IPTEK*, Vol 16 No.1.
- Amini, dkk. 2022, *Implementasi Algoritma Genetika Untuk Seleksi Fitur Pada Klasifikasi Genre Musik Menggunakan Metode Random Forest*, *Jurnal Informatika Polinema*. Volume X Edisi X.
- Ashuman, G., Sheezan, M., Masood, S. & Saleem, A. 2016, *Genre Classification of Songs Using Neural Network*, *Institute of Electrical and Electronics Engineers Inc*.
- Attenberg, J. & Ertekin, S., 2013, *Class Imbalance and Active Learning*, *Imbalanced Learning: Foundations, Algorithms, and Applications*, pp. 101-149.
- Budi, S. 2017, *Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means*, *Techno.COM*. Vol. 16. No. 1.
- Budiman, Haldi. 2016, *Analisis Dan Perbandingan Akurasi Model Prediksi Rentet Waktu Support Vector Machines Dengan Support Vector Machines Particle Swarm Optimization Untuk Arus Lalu Lintas Jangka Pendek*, *SYSTEMIC*. Vol. 02, No. 01, 19-24.
- Chen, T. & Guestrin, C., 2016, *XGBoost: A Scalable Tree Boosting System*, *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794
- Clark, A. 2003, *Pre-processing Very Noisy Text*. *Proceedings of Workshop on Shallow Processing of Large Corpora Lancaster*, Lancaster University. pp. 12-22.
- Cui, S. Yunqiang Yin, Dujuan Wang, Zhiwu Li, & Yanzhang Wang. 2020, *A stacking-based ensemble learning method for earthquake casualty prediction*. *Applied Soft Computing Journal*.
- Das, Joy Krishan., Arkha Ghosh, Abhijit Kumar Pal, & Amitabha Chakrabarty. 2020, *Urban Sound Classification Using Convolutional Neural Network and Long Short Term Memory Based on Multiple Features*, *In Proceedings 4th International Conference on Intelligent Computing in Data Sciences*, Fez, Maroko.
- Dubey, R., Jiayu Zhou B.S., Yalin Wang, Paul M. Thompson, & Jieping Ye. 2014, *Analysis Of Sampling Techniques For Imbalanced Data: An N = 648 ADNI study*, *Neuroimage*, pp. 220-241.
- Galar, M., Fernandez, A., Barrenechea, E. & Bustince, H., 2012, *A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches*, *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, pp. 463-484.
- Giri, Gst. Ayu Vida Mastrika., 2017, *Klasifikasi Musik Berdasarkan Genre dengan Metode K-Nearest Neighbor*, *Jurnal Ilmu Komputer VOL. XI No. 2*.
- Han, J., Kamber, M. & Pei, Jian. 2012, *Data Mining Concepts and Techniques*, *Elsevier Inc*, Vol.1, 978-381479.
- Haryoto, Parawystia Prabasini., Harly Okprana, & Ilham Syahputra Saragih. 2021, *Algoritma C4.5 Dalam Data Mining Untuk Menentukan Klasifikasi Penerimaan Calon Mahasiswa Baru*, *TIN: Terapan Informatika Nusantara*. Vol 2. No 5. 358-364.
- Hasan, M. K. Md. Ashraful Alam, Dola Das, & Eklas Hossain. 2020, *Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers*, *IEEE Access*, Volume 8, pp. 76516-76531.
- Ismanto, Edi., & Novaliya, Melly. 2021, *Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas*, *Techno.COM*, Vol. 20, No.3 : 400-410.
- Jijo, B. T. & Abdulazeez, A. M., 2021, *Classification Based on Decision Tree Algorithm for Machine Learning*, *JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS*, Volume 02, pp. 20-28.
- Khammassi, C & Saoussen, K. 2017, *A GA-LR Wrapper Approach For Feature Selection in Network Intrusion Detection*, *Journal of Computers & Security*. 70: 255-277.
- Khasanah FN, Setiyadi D. 2019, *Uji Sensitivitas Metode Simple Additive Weighting Dan Weighted Product Dalam Menentukan Laptop*, *Bina Insani ICT Journal*. 6(2): 165 - 174.
- Mujilahwati, S, 2016. *Pre-Processing Text Mining Pada Data Twitter*, *Seminar Nasional Teknologi Informasi dan Komunikasi*, 2089-9815.
- Pamina, J., J. Beschi Raja., S. Sathya Bama., S. Soundarya., M.S. Sruthi., S. Kiruthika., V.J. Aiswaryadevi., G. Priyanka. 2019, *An Effective Classifier for Predicting Churn in Telecommunication*, *Journal of Advanced Research in Dynamical and Control Systems*, 11(1 Special Issue), 221–229.
- Perdana, Ramadhan Rosihadi. 2017, *Implementasi Ekstraksi Fitur Untuk Pengelompokan Berkas Musik Berdasarkan Kemiripan Karakteristik Suara*, *Skripsi S-1, FTI Institut Teknologi Sepuluh Nopember, Surabaya*.
- Prasetyo, Pandu Deski., I Gede Pasek Suta Wijaya., & Ario Yudo Husodo. 2019, *Klasifikasi Genre Musik Menggunakan Metode Mel Frequency Cepstrum Coefficients (Mfcc) Dan K-Nearest Neighbors Classifier*, *JTIKA*, Vol. 1, No. 2.
- Priyanka & Kumar, D., 2020, *Decision Tree classifier: a detailed survey*, *J. Information and Decision Sciences*, Volume 12.
- Rahman, Iqbal Fathur. 2020, *Implementasi Metode SVM, MLP Dan XGBOOST Pada Data Ekspresi Gen*, *Skripsi S-1, FMIPA UII, Yogyakarta*.
- Rahmawati, R., Magdalena, R. & Ramatryana, I, N, A. 2016, *Perbandingan dan Analisis K-Narest Neighbor dan Linier Discriminant Analysis Untuk*

- Klasifikasi Genre Musik, e-Proceeding of Engineering*, Vol.3. No. 2.
- Religia, Yoga. 2019, *Feature Extraction Untuk Klasifikasi Pengenalan Wajah Menggunakan Support Vector Machine Dan K-Nearest Neighbor*, *Pelita Teknologi: Jurnal Ilmiah Informatika, Arsitektur dan Lingkungan*, 14 (2) 85-92.
- Riadi, M. 2017, *Pengertian, Fungsi, Proses, dan Tahapan Data Mining*.
<https://www.kajianpustaka.com/2017/09/data-mining.html>
(diakses tanggal 22 Januari 2022)
- Saini, R. & Ghosh, S., 2017, *Ensemble classifiers in remote sensing: A review*. In *International Conference on Computing, Communication and Automation*.
- Socrates, I. G. 2016, *Optimasi Naive Bayes Dengan Pemilihan Fitur Dan Pembobotan Gain Ratio*, *LONTAR KOMPUTER*, VOL. 7, No.1.
- Tama, B. A., Nkenyereye, L., Islam, S. M. R. & Kwak, K.-s., 2020, *An Enhanced Anomaly Detection in Web Traffic Using a Stack of Classifier Ensemble*, *IEEE Access*, Volume 8, pp. 24120-24134.
- Tamatjita, Nurmiyati. & Harjoko, Agus. 2014, *Klasifikasi Lagu Berdasarkan Genre pada Format WAV*, *IJCCS*. Vol.8, No.2, pp. 145-154.
- Wahyudi., 2020, *Implementasi Metode Random Forest Untuk Klasifikasi Genre Music*, Skripsi S-1, FMIPA ULM, Banjarbaru.
- Wang, J., Zhengwei Sun, Bin Bao, & Dongyuan Shi., 2019, *Malicious Synchronphasor Detection Based on Highly Imbalanced Historical Operational Data*, *CSEE Journal of Power and Energy System*, Volume 15, pp. 11-20.
- 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.

Lampiran

Tabel Hasil Penelitian

Tabel 12 *Confusion matrix extreme gradient boosting*

Nilai Prediksi	Nilai sebenarnya									
	Blues	Classical	Country	Disco	Hiphop	Jazz	Metal	Pop	Raggae	Rock
Blues	9	0	2	0	1	1	0	0	0	0
Classical	0	5	0	0	0	1	0	0	0	0
Country	0	0	6	0	0	4	0	1	1	1
Disco	0	0	0	6	1	0	0	1	1	1
Hiphop	0	0	0	0	4	0	1	0	1	0
Jazz	0	0	0	1	0	11	1	0	0	0
Metal	0	0	0	1	0	0	11	0	1	0
Pop	0	0	1	0	1	0	0	5	1	0
Raggae	0	0	1	0	1	0	0	0	7	0
Rock	0	0	1	2	1	1	0	0	2	2

Tabel 14 *Confusion matrix extreme gradient boosting uji parameter*

Nilai Prediksi	Nilai sebenarnya									
	Blues	Classical	Country	Disco	Hiphop	Jazz	Metal	Pop	Raggae	Rock
Blues	10	1	1	0	0	1	0	0	0	0
Classical	0	6	0	0	0	0	0	0	0	0
Country	0	0	8	0	0	2	0	1	1	1
Disco	0	0	1	6	1	0	0	0	1	1
Hiphop	0	0	0	0	4	0	1	0	1	0
Jazz	0	0	0	1	0	10	1	1	0	0
Metal	0	0	0	2	0	0	11	0	1	0
Pop	0	0	1	0	1	0	0	5	1	0
Raggae	0	0	1	0	1	0	0	0	7	0
Rock	0	0	0	1	1	1	0	0	1	5

Tabel 16 *Confusion matrix decision tree*

Nilai Prediksi	Nilai sebenarnya									
	Blues	Classical	Country	Disco	Hiphop	Jazz	Metal	Pop	Raggae	Rock
Blues	6	0	3	1	0	0	0	0	1	2
Classical	0	4	1	0	0	1	0	0	0	0
Country	2	0	6	0	0	1	0	0	1	
Disco	0	1	0	2	3	1	1	1	0	1
Hiphop	0	0	0	1	2	0	0	2	1	0
Jazz	1	1	4	0	2	4	0	0	0	1
Metal	1	0	0	2	2	0	7	0	0	1
Pop	0	0	1	0	3	0	0	3	0	1
Raggae	0	0	0	0	2	0	1	2	4	0
Rock	1	0	0	1	2	0	0	0	3	2

Tabel 12 *Confusion matrix decision tree uji parameter*

Nilai Prediksi	Nilai sebenarnya									
	Blues	Classical	Country	Disco	Hiphop	Jazz	Metal	Pop	Raggae	Rock
Blues	9	1	1	0	0	1	0	0	0	1
Classical	0	4	1	0	0	0	0	0	0	1
Country	1	0	7	0	1	2	0	1	0	1
Disco	0	0	1	2	2	0	0	2	1	2
Hiphop	0	0	0	0	3	0	1	0	2	0
Jazz	1	2	0	1	0	6	0	0	1	2
Metal	2	0	0	0	2	0	8	0	1	0
Pop	0	0	1	1	2	0	0	3	0	1
Raggae	0	0	0	0	1	0	0	0	7	1
Rock	1	0	1	1	3	1	0	0	0	2