

Perbandingan Algoritma C4.5 dan Random Forest untuk Klasifikasi Kualitas Tembakau Berdasarkan Pola Aroma Menggunakan *Electronic Nose*

Rahmat Tegar Patriot Hari Lambang¹, Dimas Saputra², Muh. Mashdarul Hilmi Aufa³,
Ifnu Wisma Dwi Prastya⁴

¹²³⁴Universitas Nahdlatul Ulama Sunan Giri, Indonesia

rahmatkarorio@gmail.com, dimsanggara35@gmail.com, mashdarulhilmi25@gmail.com, ifnuprastya@unugiri.ac.id

Info Artikel

Riwayat Artikel:

Received 2025-12-30

Revised 2026-06-16

Accepted 2026-06-20

Abstract – Tobacco quality is influenced by changes in volatile compounds that occur during the drying process and can be represented through the sensor response patterns on the electronic nose (E-Nose) device. Various machine learning algorithms have been used to classify tobacco based on E-Nose data, but the performance of each algorithm can vary depending on the characteristics of the data used. Therefore, this study aims to analyze and compare the performance of the C4.5 and Random Forest algorithms in classifying tobacco quality based on volatile compound data obtained using the E-Nose. A total of 375 tobacco sample data representing four drying conditions were used in this study. The preprocessing stage was carried out using the Interquartile Range (IQR) method to remove outliers and Moving Average to reduce noise in the MQ-4, MQ-7, and MQ-135 sensor signals. Furthermore, both algorithms were evaluated using stratified 10-fold cross-validation to obtain stable performance estimates. The results indicated that Random Forest performed better than C4.5 in all evaluation metrics. Random Forest achieved an accuracy of 94.44%, a Cohen's Kappa value of 0.9259, an MCC of 0.9262, a balanced accuracy of 0.945, and a cross-entropy log loss of 0.4291. These results indicate that Random Forest is more effective in classifying tobacco quality based on volatile compound data obtained using E-Nose.

Keywords: C4.5; Classification; Electronic Nose; Random Forest; Tobacco Quality.

Corresponding Author:

Rahmat Tegar Patriot Hari Lambang

Email: rahmatkarorio@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak – Kualitas tembakau dipengaruhi oleh perubahan senyawa volatil yang terjadi selama proses pengeringan dan dapat direpresentasikan melalui pola respons sensor pada perangkat elektronik nose (E-Nose). Berbagai algoritma pembelajaran mesin telah digunakan untuk melakukan klasifikasi berdasarkan data E-Nose, namun performa masing-masing algoritma dapat berbeda bergantung pada karakteristik data yang digunakan. Oleh karena itu, penelitian ini ditujukan untuk menganalisis dan membandingkan kinerja dari algoritma C4.5 dan Random Forest dalam mengklasifikasikan kualitas tembakau berdasarkan data senyawa volatil yang diperoleh menggunakan E-Nose. Sebanyak 375 data sampel tembakau yang merepresentasikan empat kondisi pengeringan digunakan dalam penelitian ini. Tahap pra-proses dilakukan menggunakan metode Interquartile Range (IQR) untuk menghilangkan outlier dan Moving Average untuk mereduksi noise pada sinyal sensor MQ-4, MQ-7, dan MQ-135. Selanjutnya, kedua algoritma dievaluasi menggunakan stratified 10-fold cross-validation untuk memperoleh estimasi performa yang stabil. Hasil penelitian mengindikasikan bahwa Random Forest menunjukkan kinerja yang lebih baik dibandingkan C4.5 pada setiap metrik evaluasi. Random Forest memperoleh akurasi sebesar 94,44%, nilai Cohen's Kappa sebesar 0,9259, MCC sebesar 0,9262, balanced accuracy sebesar 0,945, dan cross-entropy log loss sebesar 0,4291. Hasil tersebut menunjukkan bahwa Random Forest lebih efektif dalam mengklasifikasikan kualitas tembakau berdasarkan data senyawa volatil yang diperoleh menggunakan E-Nose.

Kata Kunci: C4.5; Electronic Nose; Klasifikasi; Kualitas Tembakau; Random Forest.

I. PENDAHULUAN

Tembakau atau *Nicotiana tabacum* merupakan salah satu komoditas perkebunan penting di Indonesia dengan nilai ekonomi tinggi. Tembakau tidak hanya memiliki nilai agronomis sebagai tanaman perkebunan musiman, tetapi juga memainkan peran penting dalam perekonomian negara Indonesia. Indonesia adalah salah satu produsen tembakau terbesar di dunia dengan kontribusi signifikan terhadap ekspor dan penerimaan negara melalui cukai hasil tembakau [1].

Peningkatan produksi tembakau Indonesia berdampak langsung terhadap penerimaan bea cukai negara dan memberikan efek positif terhadap kesejahteraan masyarakat serta pertumbuhan ekonomi nasional. Industri hasil tembakau juga memiliki peranan krusial dalam menyerap tenaga kerja serta menjaga keseimbangan ekonomi saat menghadapi perubahan di pasar internasional [2]. Oleh karena itu, tembakau menjadi komoditas strategis yang tidak hanya bernilai ekonomi tinggi tetapi juga memiliki keterkaitan erat dengan kebijakan fiskal dan perdagangan internasional [3].

Keberlanjutan peran ekonomi tersebut sangat bergantung pada kualitas tembakau yang dihasilkan. Sebagai salah satu negara produsen tembakau, Indonesia perlu menjaga konsistensi mutu agar dapat bersaing di pasar global. Kualitas tembakau sangat dipengaruhi oleh proses pascapanen, terutama tahap pengeringan (*curing*) yang menentukan warna, aroma, dan kadar nikotin pada daun [4].

Proses pengeringan yang tidak terkendali dapat menyebabkan penurunan mutu, aroma yang tidak konsisten, serta kesulitan dalam standarisasi produk akhir. Oleh karena itu, pengawasan mutu pada tahap pengeringan menjadi aspek krusial dalam rantai produksi tembakau [5]. Proses pengecekan mutu tembakau yang dilakukan secara tradisional sering kali memerlukan waktu lebih lama, bersifat destruktif, dan sangat bergantung pada keahlian operator [4].

Untuk mengatasi hal tersebut, berbagai penelitian mulai memperkenalkan metode pengukuran cepat dan non-destruktif berbasis sensor, salah satunya menggunakan teknologi *Electronic Nose* (E-Nose). Teknologi E-Nose bekerja dengan mendeteksi pola respons sensor terhadap campuran senyawa volatil (*volatile organic compounds* / VOC) yang berasal dari daun tembakau. Pola respons tersebut kemudian digunakan sebagai representasi aroma untuk membedakan kualitas tembakau [6]. Penelitian sebelumnya oleh Esteves, dkk. (2018) menunjukkan bahwa E-Nose mampu mengidentifikasi tingkat kualitas aroma dengan akurasi tinggi [7].

Penelitian ini dilakukan untuk membandingkan kinerja algoritma C4.5 dan Random Forest dalam mengklasifikasikan kualitas tembakau berdasarkan pola aroma yang diperoleh dengan perangkat E-Nose.

Dalam *machine learning*, klasifikasi merupakan teknik yang digunakan untuk mempelajari pola dari data berlabel sehingga model dapat memprediksi kategori atau kelas pada data baru. Proses ini dilakukan dengan mempelajari pola dari data yang sudah ada sebelumnya (data berlabel), sehingga sistem mampu mengenali karakteristik tertentu dan menentukan kelas dari data baru yang belum diketahui [8].

Algoritma klasifikasi bekerja berdasarkan pendekatan *supervised learning*, di mana data pelatihan (*training data*) digunakan untuk membangun model yang dapat memprediksi label atau kelas berdasarkan nilai dari variabel *input*. Dengan adanya proses pembelajaran ini, model dapat mempelajari hubungan antara fitur dan label secara optimal sehingga hasil prediksi yang dihasilkan memiliki tingkat akurasi yang tinggi [9].

Beberapa algoritma klasifikasi yang sering digunakan antara lain Decision Tree, Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), dan Random Forest. Pemilihan algoritma bergantung pada karakteristik data seperti ukuran, jumlah fitur, dan jenis kelas. Tren penelitian terkini menunjukkan bahwa berbagai algoritma klasifikasi dibandingkan untuk mencapai tingkat akurasi dan efisiensi terbaik [10].

Beberapa penelitian sebelumnya telah banyak menerapkan E-Nose untuk klasifikasi aroma produk pertanian seperti teh, kopi, dan tembakau. Algoritma C4.5 adalah salah satu algoritma yang tergolong dalam kelompok *decision tree*, yang merupakan teknik klasifikasi yang tampak seperti pohon yang dibalik. Algoritma C4.5 dikembangkan oleh Quinlan, dkk (1993) sebagai pengembangan dari ID3, dengan kemampuan menangani atribut numerik, nilai yang hilang (*missing values*), dan melakukan *pruning* untuk mengurangi *over-fitting* [11], [12].

Lebih lanjut, Lu dan Chen (2015) menjelaskan bahwa C4.5 memilih atribut dengan *gain ratio* sebagai kriteria pemisahan, kemudian melakukan pemangkasan pohon untuk mempertahankan generalisasi [13]. Namun, temuan oleh Nugroho dan Sugiharti (2021), ketika algoritma C4.5 digunakan sendirian masih menunjukkan penurunan akurasi jika data memiliki *noise* atau distribusi yang kompleks [14].

Penelitian yang dilakukan oleh Mula Agung Barata, dkk. (2023) berhasil meningkatkan akurasi klasifikasi aroma teh dengan menerapkan metode *feature selection* sebelum algoritma C4.5 dan menghasilkan akurasi tinggi dari 93,65% menjadi 94,27 %, dimana presentase menjadi lebih tinggi dibandingkan dengan akurasi algoritma C4.5 secara individu [15]. Sehingga dalam penelitian ini C4.5 akan digunakan sebagai *baseline* dan dibandingkan dengan Random Forest, yaitu sebuah metode *ensemble* berbasis *decision tree* yang lebih canggih, untuk meningkatkan performa klasifikasi [16].

Random Forest terdiri dari beberapa pohon keputusan yang dibangun dari berbagai sampel data, lalu mengintegrasikan semua hasil prediksi dengan cara *voting* untuk mencapai keputusan akhir yang lebih stabil [17], [18]. Setiap pohon dilatih pada sampel dari data pelatihan dan pada setiap *node* pembelahan hanya dipertimbangkan subset fitur secara acak, sehingga hubungan antar-pohon berkurang dan *overfitting* pada pohon tunggal dapat diminimalkan. Prediksi akhir diperoleh melalui mekanisme *majority voting* untuk klasifikasi, sedangkan nilai *out-of-bag* (OOB) dapat dimanfaatkan sebagai estimasi kesalahan prediksi tanpa perlu *cross-validation* tambahan [19].

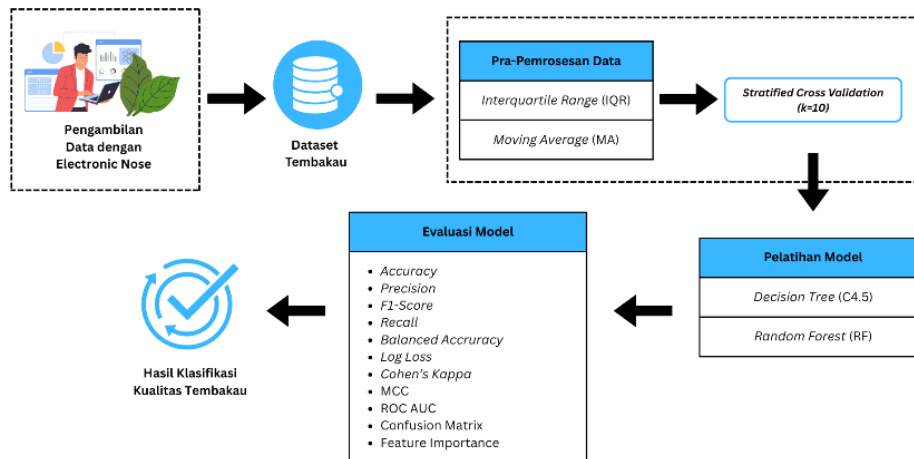
Penelitian yang dilakukan oleh Ismanto dan Novalia (2021), memberikan dasar penting dalam pemilihan metode klasifikasi berbasis *decision tree*. Dalam studi tersebut, dilakukan perbandingan antara algoritma C4.5, Random Forest, dan Gradient Boosting dalam mengklasifikasikan data terkait komoditas perkebunan di Provinsi Riau. Temuan dari pengujian menunjukkan bahwa algoritma Random Forest memberikan performa paling unggul dengan tingkat akurasi mencapai 98,96% saat menerapkan metode *shuffle sampling* dan kriteria *gain ratio*. [20].

Namun, penelitian yang secara khusus membandingkan performa algoritma C4.5 dan Random Forest untuk klasifikasi kualitas tembakau berdasarkan pola aroma yang diperoleh menggunakan *electronic nose* masih terbatas. Sebagian besar penelitian sebelumnya lebih berfokus pada penerapan satu algoritma klasifikasi atau peningkatan performa melalui teknik *feature selection* dan optimasi model. Dengan demikian, penelitian ini dilakukan untuk

mengisi kesenjangan tersebut melalui perbandingan analisis antara dua algoritma menggunakan berbagai parameter evaluasi untuk menentukan metode yang paling efektif dalam mengklasifikasikan kualitas tembakau berdasarkan pola aroma.

II. METODE

Penelitian ini memakai pendekatan eksperimental kuantitatif guna menciptakan dan membandingkan suatu model klasifikasi mutu tembakau berdasarkan ciri aroma yang ditunjukkan oleh respon sensor terhadap campuran senyawa volatil dalam beragam kondisi pengeringan. Sistem pengumpulan data dilakukan menggunakan perangkat *E-Nose* berbasis sensor gas seri MQ yang dihubungkan dengan mikrokontroler ESP32, sedangkan proses analisis data dilakukan dengan metode *data mining* dan *machine learning* menggunakan algoritma C4.5 dan Random Forest.



Gambar 1. Alur metodologi penelitian

A. Electronic Nose

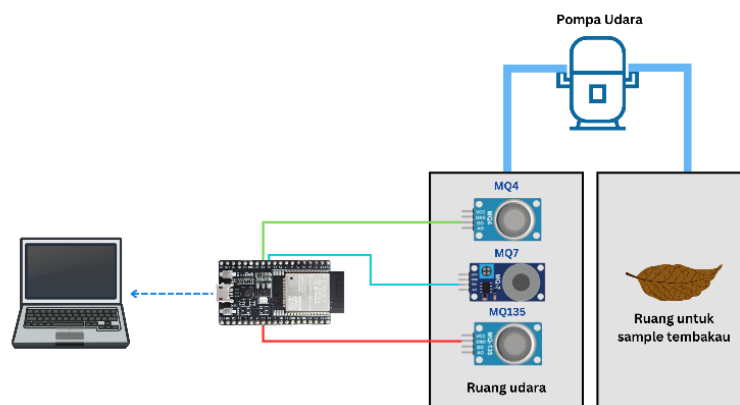
Tembakau mengandung berbagai komponen kimia yang berperan dalam pembentukan aroma khasnya, termasuk nikotin, protein, pati, serta berbagai senyawa volatil (*volatile compounds*). Senyawa volatil tersebut dapat dideteksi menggunakan perangkat *E-Nose* dan digunakan sebagai representasi aroma pada proses klasifikasi [21]. Perangkat e-nose yang diterapkan dalam mengumpulkan data tembakau dalam penelitian ini terdiri dari berbagai komponen, termasuk mikrokontroler ESP32 dengan beberapa sensor Seri MQ [22]. Maka dipilihlah sensor gas pada tabel 1.

TABEL 1
SENSOR YANG DIGUNAKAN PADA ELECTRONIC NOSE

Sensor	Gas yang dideteksi
MQ-4	Natural Gas, Metana (CH ₄)
MQ-7	Karbon Monoksida (CO)
MQ-135	Amonia (NH ₃), Benzena, VOC

Sensor gas yang terpasang pada perangkat *E-Nose* mendeteksi gas yang masuk ke ruang sirkulasi udara. Ruang sirkulasi udara berfungsi untuk mengalirkan dan meratakan campuran senyawa volatil yang dihasilkan sampel tembakau sehingga respons sensor dapat diukur secara lebih stabil. Skema konfigurasi dalam desain *electronic nose* dapat dilihat pada Gambar 2.

Perancangan perangkat *E-Nose* yang digunakan dalam penelitian ini dirancang dengan mengacu pada berbagai penelitian sebelumnya yang telah mengembangkan sistem serupa untuk beragam kebutuhan. Teknologi *E-Nose* telah banyak diterapkan, antara lain untuk mengidentifikasi kualitas kopi, mengevaluasi kualitas teh, serta membedakan jenis daging seperti daging sapi dan daging babi. Berdasarkan referensi tersebut, dilakukan perancangan perangkat *E-Nose* yang disesuaikan dengan kebutuhan penelitian ini. Hasil perancangan perangkat dapat dilihat pada Gambar 3.



Gambar 2. Skema konfigurasi E-Nose



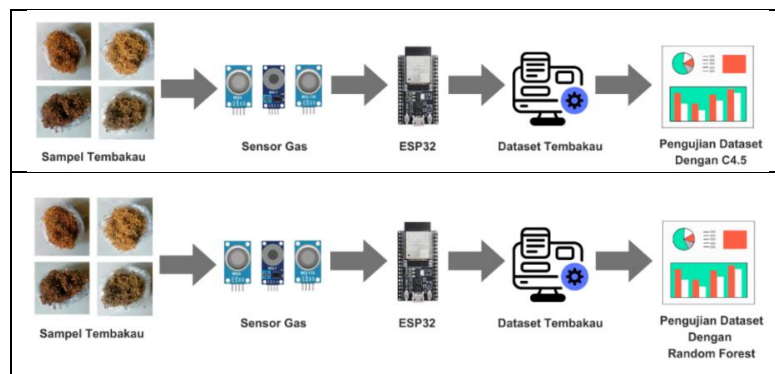
Gambar 3. Perangkat electronic nose

B. Alur Kerja Sistem Electronic Nose

Alur sistem yang dirancang pada perangkat *E-Nose*, yang mencakup interaksi antara perangkat keras dan perangkat lunak untuk mengumpulkan dataset kualitas tembakau, ditunjukkan pada Gambar 4. Sampel diuji dalam bentuk tembakau rajang, untuk memaksimalkan jumlah gas yang dikeluarkan oleh daun tembakau [23]. Keempat sampel tembakau disedot ke dalam ruang udara dibantu oleh pompa udara. Aliran gas tembakau ke dalam ruang udara yang telah dipasang sensor gas, direkam respons sensor terhadap campuran senyawa volatil dalam satuan analog, yaitu dari MQ4, MQ7, dan MQ135.

Sensor MQ4, MQ7, dan MQ135 dipilih karena telah banyak digunakan pada penelitian *E-Nose* untuk proses klasifikasi berdasarkan pola aroma atau senyawa volatil. Meskipun MQ4 secara umum dirancang untuk mendeteksi natural gas dan gas metana (CH_4) dan MQ7 untuk mendeteksi karbon monoksida (CO), pada sistem *E-Nose* sensor-sensor tersebut tidak digunakan untuk mengidentifikasi gas tertentu secara spesifik. Sebaliknya, sensor digunakan sebagai *sensor array* terhadap berbagai senyawa volatil (*volatile organic compounds/VOC*) sehingga menghasilkan pola respons gabungan yang dapat digunakan sebagai representasi aroma sampel. Penelitian sebelumnya oleh Wakhid dkk. (2021) menunjukkan bahwa kombinasi sensor MQ-135, MQ-2, MQ-3, MQ-4, dan MQ-7 mampu menghasilkan pola respons aroma yang efektif untuk membedakan kopi luwak dan non-kopi luwak menggunakan *E-Nose* [22]. Hasil penelitian tersebut menunjukkan bahwa pola respons gabungan beberapa sensor gas dapat digunakan sebagai representasi karakteristik aroma suatu sampel tanpa harus mengidentifikasi senyawa volatil secara individual. Oleh karena itu, penelitian ini memanfaatkan MQ4, MQ7, dan MQ135 sebagai *sensor array* untuk memperoleh pola respons terhadap senyawa volatil yang dihasilkan tembakau dan menggunakannya sebagai fitur pada proses klasifikasi kualitas tembakau.

Pengumpulan data hasil pengukuran tembakau berlangsung selama 10 menit untuk setiap kondisi pengeringan tembakau. Data yang terkumpul, termasuk nilai konsentrasi keempat kondisi tembakau, ditangkap oleh tiga sensor gas terpasang yang dikontrol oleh ESP32, yang terhubung ke komputer sebagai perangkat pemantau selama proses pengumpulan data. Data tersebut disimpan di komputer, dan kemudian dilakukan visualisasi data untuk mengetahui kondisi data yang diperoleh. Pra-pemrosesan data merupakan salah satu proses yang dilakukan sebelum dataset diuji sebagai model algoritmik. Alur sistem komunikasi perangkat *E-Nose* dapat dilihat pada Gambar 4.



Gambar 4. Sistem komunikasi perangkat E-Nose

C. Dataset

Objek penelitian ini adalah data tembakau dari tanaman *Nicotiana tabacum* yang diolah menjadi tembakau rajang. Data diperoleh dari hasil pengukuran menggunakan perangkat E-Nose yang terdiri atas tiga sensor gas, yaitu MQ4, MQ7, dan MQ135. Dataset yang dipakai terdiri dari 375 data dengan tiga atribut, yaitu nilai respons dari masing-masing sensor. Proses pelabelan (*labelling*) dilakukan berdasarkan kondisi pengeringan tembakau pada saat pengambilan data yang dibagi menjadi empat kelas, yaitu kelas 1 (1 hari panas optimal), kelas 2 (1 hari 1 malam), kelas 3 (2 hari panas tidak optimal), dan kelas 4 (6 jam atau tembakau basah). Perbedaan kondisi pengeringan tersebut menghasilkan karakteristik senyawa volatil yang berbeda sehingga dapat digunakan sebagai dasar dalam proses klasifikasi kualitas tembakau.

TABEL 2
DATASET TEMBAKAU

Waktu	MQ4	MQ7	MQ135	Kelas
29/11/2025 11:09	2190	2642	1111	1
29/11/2025 11:09	2727	3411	1232	1
29/11/2025 11:09	2886	3664	1303	1
29/11/2025 11:09	2919	3749	1331	1
...	...	...	...	...
29/11/2025 11:25	2089	2747	1220	2
29/11/2025 11:25	2523	3338	1264	2
29/11/2025 11:25	2752	3615	1328	2
29/11/2025 11:25	2842	3706	1354	2
...	...	...	...	...
29/11/2025 11:43	2416	2864	1200	3
29/11/2025 11:43	2593	3082	1255	3
29/11/2025 11:43	2591	3056	1215	3
29/11/2025 11:43	2822	3469	1297	3
...	...	...	...	...
29/11/2025 12:09	2427	3406	1263	4
29/11/2025 12:09	2427	3417	1264	4
29/11/2025 12:09	2431	3415	1269	4
29/11/2025 12:09	2352	3280	1173	4

D. Pra-Pemrosesan

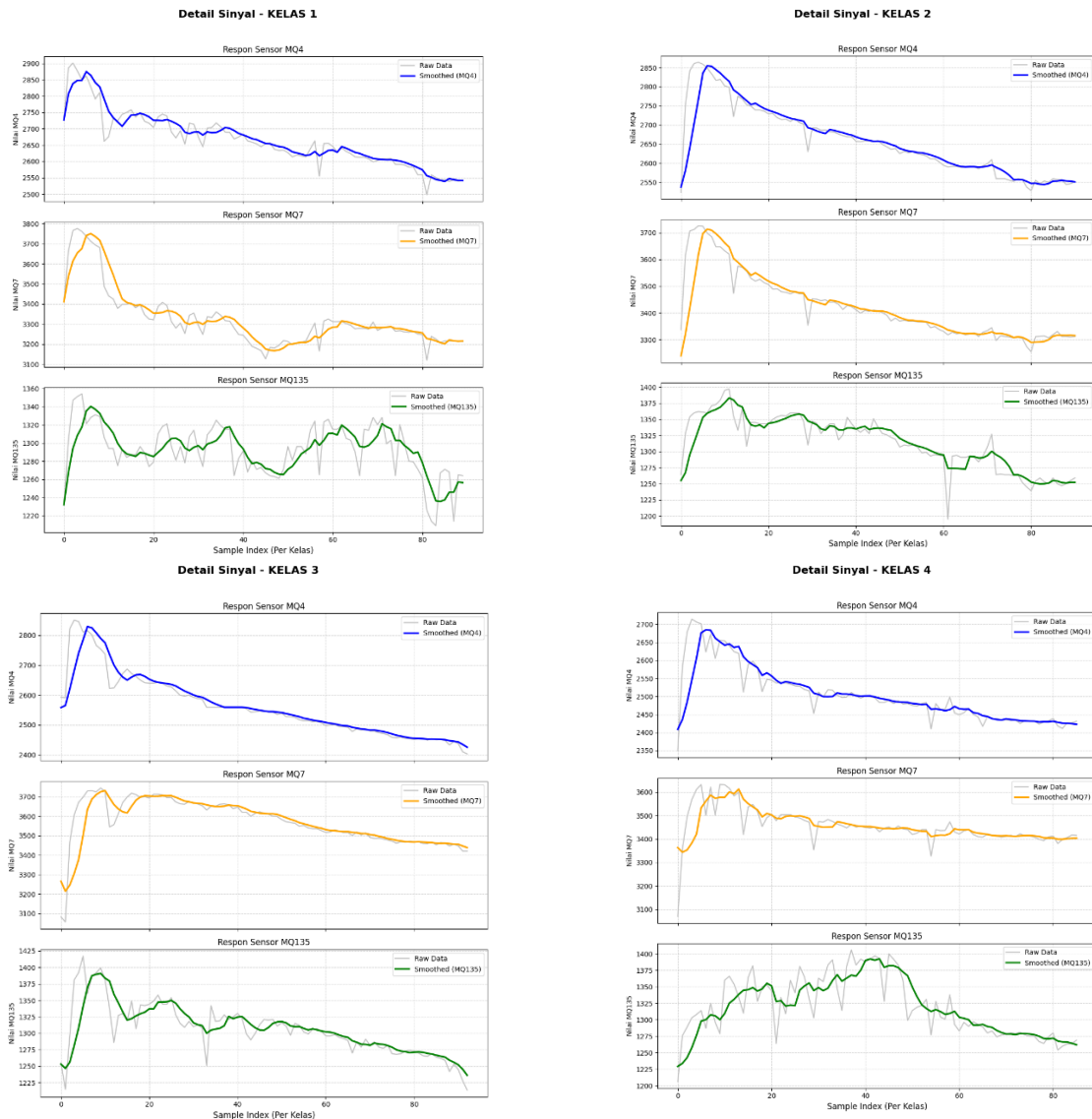
Data mentah yang diperoleh dari sensor gas E-Nose seringkali mengandung gangguan (*noise*) dan anomali sinyal. Oleh karena itu, dilakukan dua tahapan pra-pemrosesan untuk menjamin kualitas data sebelum masuk ke tahap klasifikasi, yaitu penanganan *outlier* dan penghalusan sinyal.

Langkah pertama bertujuan untuk mengeliminasi data pencilon (*outlier*) yang menyimpang jauh dari distribusi normal, yang biasanya disebabkan oleh gangguan teknis pada sensor sesaat. Metode IQR menghitung selisih antara kuartil ketiga $Q3$ dan kuartil pertama $Q1$. Data dinyatakan sebagai *outlier* dan dihapus jika nilainya berada di luar rentang batas yang ditentukan oleh persamaan dengan rumus 1 dan 2 dibawah ini:

$$\text{BatasBawah} = Q1 - 1.5 \times IQR \quad (1)$$

$$\text{BatasAtas} = Q3 + 1.5 \times IQR \quad (2)$$

Setelah data bersih dari *outlier*, langkah kedua adalah mereduksi fluktuasi kecil (*noise*) pada sinyal agar pola tren gas lebih mudah dikenali oleh algoritma. Metode yang digunakan adalah *Moving Average (MA)* dengan ukuran *window* = 5. Metode ini bekerja dengan mengambil rata-rata dari 5 data berturut-turut untuk menghasilkan grafik respon sensor yang lebih halus.



Gambar 5. Hasil Smoothing Data Sensor E-Nose Perbandingan 4 Kelas

E. Algoritma

Salah satu algoritma yang bisa diterapkan dalam membentuk pohon keputusan (*decision tree*) adalah algoritma C4.5. Algoritma C4.5 ini pertama kali diperkenalkan oleh Quinlan, yang merupakan penyempurnaan dari algoritma ID3 [11]. Pohon model algoritma C4.5 dibangun dengan membagi data secara rekursif, teknik di mana suatu fungsi memanggil dirinya sendiri untuk memecahkan suatu masalah dengan cara yang berulang, hingga setiap bagian terdiri dari data dari kelas yang sama. Proses pertama algoritma C4.5 adalah melakukan perhitungan untuk mendapatkan entropi global dari dataset kualitas tembakau dengan rumus 3.

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (3)$$

S adalah himpunan kasus i , n adalah total sampel, p_i adalah proporsi kelas.

Setelah entropi global diperoleh, hitung perolehan informasi pada semua atribut dataset kualitas tembakau dengan rumus 4.

$$Gain(A) = Entropy(S) - \sum_{i=1}^n \left(\frac{|S_i|}{|S|} \right) * Entropy(S_i) \quad (4)$$

S merupakan kumpulan kasus, A adalah atribut, n merujuk pada total keseluruhan sampel, $|S_i|$ menunjukkan rasio kasus dalam partisi, dan $|S|$ adalah jumlah kasus yang terdapat dalam S.

Atribut yang memiliki perolehan informasi paling signifikan kemudian digunakan sebagai simpul akar. Kemudian, pisahkan dataset berdasarkan cabang pada simpul 1 dengan rumus 5.

$$Split\ Info(S, A) = \sum_{i=1}^0 \frac{|S_t|}{|S|} * \log_2 \left(\frac{|S_t|}{|S|} \right) \quad (5)$$

S adalah himpunan kasus, A adalah atribut, dan $|S_i|$ adalah jumlah kasus pada partisi,

Setelah memperoleh nilai gain dan informasi split, Langkah berikutnya adalah menghitung nilai gain ratio menggunakan rumus 6.

$$Gain\ Ratio(S, A) = \frac{Gain\ Ratio(S, A)}{Split\ Information(S, A)} \quad (6)$$

S merupakan himpunan kasus, dan A merupakan atribut.

Random Forest adalah teknik *ensemble learning* yang menciptakan beberapa pohon keputusan dengan cara acak melalui teknik pengambilan *bootstrap sampling* dan pemilihan subset fitur di setiap *node*. Setiap pohon menghasilkan prediksi, dan keputusan akhir ditentukan menggunakan *majority voting*, sehingga model menjadi lebih konsisten dan tidak rentan terhadap *overfitting*. Metode ini terbukti efektif dalam memproses data sensor gas yang bersifat non-linear, sebagaimana ditunjukkan oleh Pilat-Rożek dkk (2023) yang menerapkan Random Forest untuk klasifikasi sinyal *E-Nose* dengan akurasi tinggi [24]. Oleh karena itu, Random Forest digunakan dalam penelitian ini sebagai salah satu algoritma pembandingan untuk mengklasifikasikan berdasarkan pola aroma yang diperoleh dari data sensor gas.

Random Forest terdiri dari K pohon keputusan. Setiap pohon memberikan hasil prediksi kelas terhadap data x . Rumus keputusan akhir menggunakan *majority voting*, dalam kasus klasifikasi Random Forest menggunakan rumus 7 dibawah ini:

$$\hat{y} = mode\{h_1(x), h_2(x), \dots, h_k(x)\} \quad (7)$$

$h_k(x)$ merupakan hasil prediksi dari pohon keputusan yang ke- k , K adalah jumlah seluruh pohon, *mode* adalah kelas yang paling sering muncul.

Pada setiap node pohon di Random Forest, pemilihan split dilakukan dengan Gini Index:

$$Gini(t) = 1 - \sum_{i=1}^c p_i^2 \quad (8)$$

C adalah jumlah kelas, p_i adalah proporsi data pada kelas ke- i di node t . Node dengan impurity terendah dipilih sebagai split terbaik.

F. Validasi

Validasi model pengujian pada dataset kualitas tembakau dilakukan menggunakan *stratified 10-fold cross-validation* untuk memperoleh estimasi performa yang lebih stabil. Dataset dibagi menjadi sepuluh bagian, proses ini dilakukan sebanyak sepuluh kali hingga seluruh *fold*. Akurasi akhir diperoleh dari rata-rata akurasi pada seluruh iterasi [25].

G. Evaluasi Kinerja Algoritma

Evaluasi kinerja model dilaksanakan dengan beberapa metrik standar dalam pengklasifikasian, yaitu akurasi, presisi, *recall*, dan *F1-score*. Disamping itu, *confusion matrix* dipakai untuk mengkaji sebaran prediksi di antara kelas dan mengenali kelas mana yang paling sering mengalami kesalahan. Seluruh metrik dihitung berdasarkan hasil validasi yang melibatkan *stratified 10-fold cross-validation* untuk menjamin hasil yang konsisten dan tidak bias terhadap pembagian data tertentu.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan temuan dari eksperimen klasifikasi kualitas tembakau berdasarkan tingkat kondisi pengeringan *stratified 10-fold cross-validation* untuk menjamin keandalan model dan mengurangi bias dalam pembagian data latih dan data uji. Metrik yang diterapkan meliputi Akurasi, Presisi, *Recall*, *F1-Score*, serta analisis stabilitas model melalui Standar Deviasi.

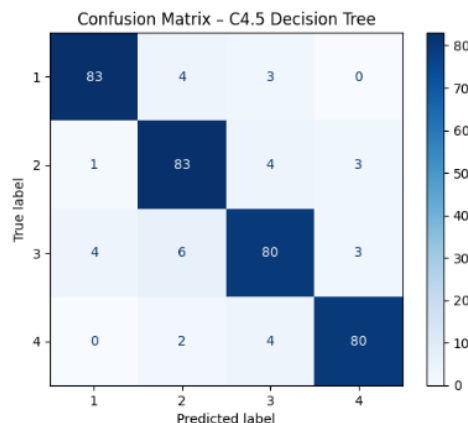
A. Hasil Pengujian Algoritma C4.5

Pengujian tahap pertama dilakukan menggunakan algoritma *Decision Tree C4.5* sebagai *baseline*. Berdasarkan validasi silang 10 lipatan (*stratified 10-fold cross-validation*), algoritma C4.5 memberikan rata-rata akurasi sebesar 90.56%. Hasil detail per-fold dapat dilihat pada Tabel 3.

TABEL 3
HASIL RATA-RATA STRATIFIED 10-FOLD CROSS-VALIDATION ALGORITMA C4.5

Algoritma	Fold	Akurasi %
C4.5	1	94,44
	2	91,67
	3	88,89
	4	86,11
	5	86,11
	6	86,11
	7	94,44
	8	94,44
	9	94,44
	10	88,89
Rata-rata		90,56
Std. Deviasi		3,56

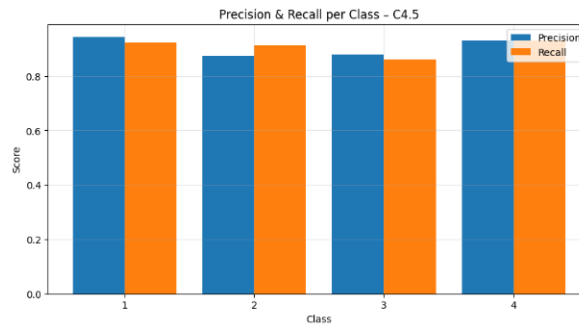
Meskipun mencapai akurasi rata-rata di atas 90%, terlihat adanya fluktuasi performa yang cukup signifikan antar *fold*. Pada *fold* ke-4, 5, dan 6, akurasi turun hingga 86,11%, sementara pada *fold* lain mencapai 94,44%. Hal ini tercermin dari nilai standar deviasi sebesar 3.56%, yang mengindikasikan bahwa model pohon keputusan tunggal (C4.5) cukup sensitif terhadap variasi data latih (*high variance*).



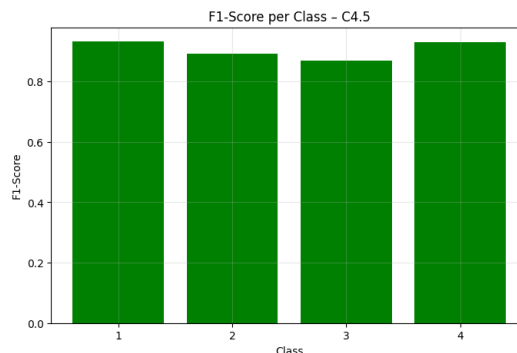
Gambar 6. Confusion Matrix Algoritma C4.5 Decision Tree

Berdasarkan Gambar 6, terlihat bahwa mayoritas data telah terklasifikasi dengan benar pada diagonal utama (kotak biru tua). Namun, terdapat misklasifikasi yang cukup mencolok pada Kelas 3. Sebanyak 13 data dari Kelas 3 salah diprediksi (4 data diprediksi sebagai Kelas 1, 6 data sebagai Kelas 2, dan 3 data sebagai Kelas 4). Kesalahan ini menunjukkan bahwa algoritma C4.5 mengalami kesulitan dalam membedakan karakteristik gas pada tembakau dengan kondisi pengeringan fase transisi (Kelas 2 dan 3), yang memiliki kemiripan pola data sensor.

Untuk memvalidasi distribusi kesalahan tersebut, performa klasifikasi per kelas dianalisis lebih lanjut menggunakan metrik Presisi, Recall, dan *F1-Score* yang disajikan pada Gambar 7 dan Gambar 8.



Gambar 7. Grafik Precision dan Recall per kelas – C4.5

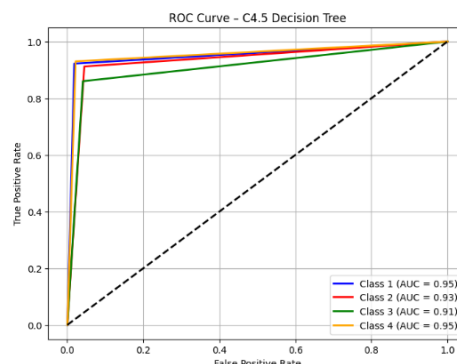


Gambar 8. Grafik F1-Score per kelas – C4.5

Berdasarkan Gambar 7 dan 8, terlihat jelas bahwa Kelas 3 memiliki performa terendah dibandingkan kelas lainnya. Penurunan nilai *Recall* pada Kelas 3 menunjukkan bahwa algoritma C4.5 sering gagal mendeteksi sampel kelas ini (banyak *False Negative*), sedangkan penurunan *Precision* mengindikasikan bahwa banyak sampel dari kelas lain yang salah diprediksi masuk ke Kelas 3 (*False Positive*). Hal ini mengonfirmasi temuan pada *confusion matrix* bahwa Kelas 3 adalah kelas yang paling sulit dipisahkan oleh *decision tree* tunggal.

Kinerja C4.5 yang lebih rendah dibandingkan Random Forest disebabkan karena algoritma ini membangun model menggunakan satu *decision tree* sehingga sangat dipengaruhi oleh distribusi data latih. Pada data *electronic nose*, pola respons sensor antar kelas memiliki kemiripan, khususnya pada Kelas 2 dan Kelas 3 yang merupakan kondisi pengeringan transisi. Kemiripan pola tersebut menyebabkan proses pemisahan data menjadi kurang optimal sehingga meningkatkan kemungkinan terjadinya kesalahan klasifikasi. Selain itu, penggunaan satu pohon keputusan membuat C4.5 lebih sensitif terhadap *noise* dan variasi data.

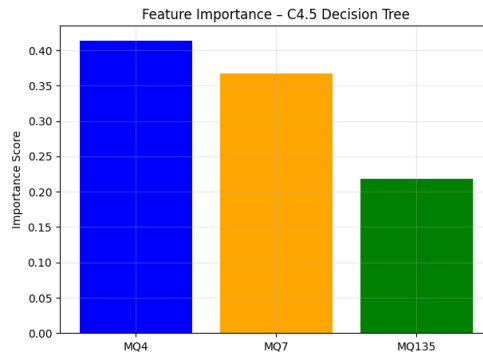
Selanjutnya, kemampuan separabilitas model dievaluasi menggunakan kurva ROC (*Receiver Operating Characteristic*) pada Gambar 9.



Gambar 9. Kurva ROC Multiclass – C4.5

Kurva ROC menunjukkan area di bawah kurva yang merepresentasikan kemampuan model membedakan antar kelas. Kelas 1 dan 4 memiliki nilai AUC sempurna atau mendekati sempurna (0.95), sedangkan Kelas 3 memiliki nilai AUC terendah yaitu 0.91. Meskipun nilai 0.91 masih tergolong kategori "Sangat Baik", selisih ini menunjukkan adanya *overlap* fitur yang signifikan pada data Kelas 3 yang tidak dapat dipisahkan secara linear oleh C4.5.

Selanjutnya, analisis *feature importance* dilakukan untuk mengetahui kontribusi setiap sensor gas dalam pembentukan pohon keputusan C4.5.



Gambar 10. Feature Importance – C4.5

Sebagaimana ditunjukkan pada Gambar 10, sensor MQ4 (pendeteksi Metana/Natural Gas) memiliki kontribusi terbesar dengan skor di atas 0.40, diikuti oleh MQ7 (Karbon Monoksida) dan MQ135 (Amonia/Benzena). Dominasi sensor MQ4 menunjukkan bahwa sensor tersebut memberikan kontribusi yang besar dalam membedakan pola respons gas antar kelas tembakau. Hal ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa MQ4 dapat digunakan sebagai bagian dari sensor array pada *E-Nose* untuk membentuk pola respons terhadap campuran senyawa volatil [22].

Selain metrik akurasi dan *F1-Score*, evaluasi lanjutan dilakukan menggunakan *Cohen's Kappa*, *Matthews Correlation Coefficient (MCC)*, *Balanced Accuracy*, dan *Cross-Entropy Log Loss*. Pada algoritma C4.5, nilai *Cohen's Kappa* sebesar 0.8740 dan MCC sebesar 0.8742 menunjukkan tingkat kesepakatan yang sangat tinggi antara prediksi dan label sebenarnya. Nilai MCC: 0.8742 mengindikasikan performa prediktif yang stabil pada kondisi kelas imbang maupun tidak imbang. Metrik *Balanced Accuracy* sebesar 0.9062 menunjukkan bahwa model mampu mengenali tiap kelas secara seimbang. Sementara itu, *Cross-Entropy Log Loss* sebesar 3.4041 mencerminkan kualitas probabilitas prediksi yang dihasilkan model.

TABEL 4
METRIK EVALUASI TAMBAHAN MODEL C4.5

Metrik Evaluasi	Nilai
Cohen's Kappa	0.8740
MCC	0.8742
Balanced Accuracy	0.9062
Log Loss	3.4041

Nilai *Cohen's Kappa* dan *MCC* yang tinggi menunjukkan bahwa model tidak hanya bergantung pada distribusi kelas, tetapi benar-benar mampu memprediksi pola data tembakau dengan stabil. *Balanced Accuracy* yang mencapai 0.9062 memperlihatkan bahwa performa model konsisten pada semua kelas kualitas tembakau, termasuk kelas transisi yang sebelumnya sulit dibedakan. *Log Loss* yang relatif rendah menunjukkan kualitas probabilitas prediksi yang baik, terutama pada tingkat kedewasaan aroma tembakau yang lebih kompleks.

B. Hasil Pengujian Algoritma Random Forest

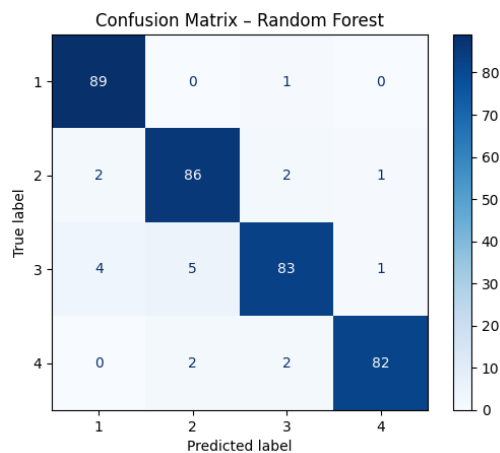
Pengujian tahap selanjutnya menerapkan algoritma Random Forest yang merupakan metode *ensemble learning*. Dengan konfigurasi parameter yang sama dan metode validasi *stratified 10-fold cross-validation*, Random Forest menunjukkan peningkatan performa yang konsisten dibandingkan C4.5. Rata-rata akurasi meningkat menjadi 94,44%. Detail hasil pengujian untuk setiap iterasi disajikan pada Tabel 5.

TABEL 5
HASIL RATA-RATA STRATIFIED 10-FOLD CROSS-VALIDATION ALGORITMA RANDOM FOREST

Algoritma	Fold	Akurasi %
Random Forest	1	94,44
	2	91,67
	3	94,44
	4	94,44
	5	88,89
	6	94,44
	7	94,44
	8	97,22
	9	97,22
	10	97,22
Rata-rata		94,44
Std. Deviasi		2,48

Berbeda dengan hasil C4.5 yang menunjukkan fluktuasi performa cukup signifikan, Random Forest menunjukkan stabilitas yang lebih baik. Hal ini tercermin dari penurunan nilai standar deviasi menjadi 2,48% (dibandingkan 3,56% pada C4.5). Sebagian besar *fold* mencapai akurasi tinggi 94,44% hingga 97,22%, dengan hanya satu penurunan di bawah 90% pada *fold* ke-5. Konsistensi ini mengindikasikan bahwa metode *ensemble* efektif dalam mereduksi varians yang menjadi kelemahan *decision tree* tunggal.

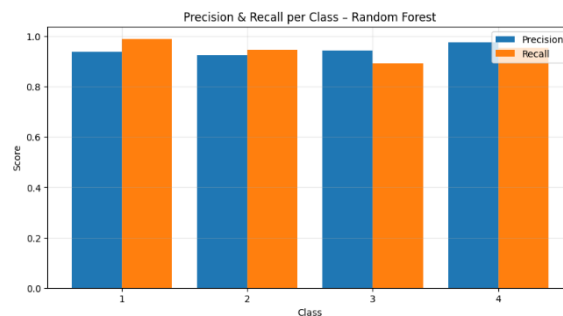
Random Forest mampu mengatasi kelemahan tersebut melalui pendekatan *ensemble learning* yang membangun banyak *decision tree* menggunakan proses *bootstrap sampling* dan pemilihan subset fitur secara acak pada setiap *node*. Strategi ini menghasilkan model yang lebih beragam sehingga kesalahan prediksi pada satu pohon dapat dikompensasi oleh pohon lainnya melalui mekanisme *majority voting*.



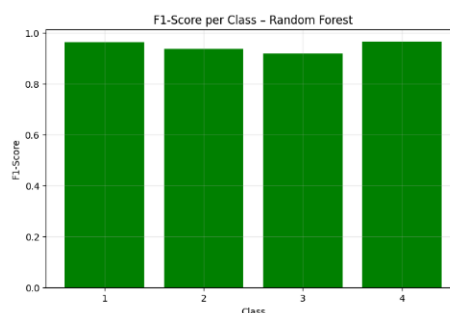
Gambar 11. Confusion Matrix Algoritma Random Forest

Perbandingan antara Gambar 11 (Random Forest) dan Gambar 6 (C4.5) menunjukkan pengurangan jumlah misklasifikasi yang signifikan. Pada Kelas 3, jumlah kesalahan prediksi berkurang dari 13 data (pada C4.5) menjadi 10 data (4 diprediksi sebagai Kelas 1, 5 sebagai Kelas 2, dan 1 sebagai Kelas 4). Selain itu, pada Kelas 1, Random Forest hampir mencapai prediksi sempurna dengan hanya 1 kesalahan.

Peningkatan performa ini juga terlihat pada metrik evaluasi per kelas yang disajikan pada Gambar 12 dan Gambar 13.

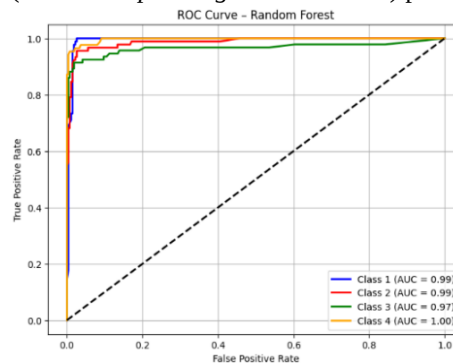


Gambar 12. Grafik Precision dan Recall per kelas – Random Forest



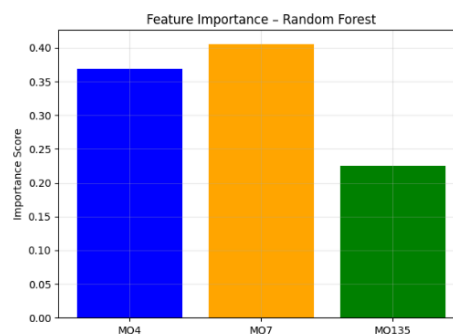
Gambar 13. Grafik F1-Score per kelas – Random Forest

Berdasarkan grafik di atas, terlihat bahwa nilai *F1-Score* untuk Kelas 3 mengalami peningkatan dari sekitar 0.87 (pada C4.5) menjadi 0.92 (pada Random Forest). Hal ini membuktikan bahwa metode *bagging* (*Bootstrap Aggregating*) pada Random Forest efektif dalam menangani *noise* dan varians data pada kelas transisi, sehingga kemampuan model dalam mengenali pola gas tembakau kualitas menengah menjadi lebih baik. Validasi lebih lanjut dilakukan menggunakan kurva ROC (*Receiver Operating Characteristic*) pada Gambar 14.



Gambar 14. Kurva ROC Multiclass – Random Forest

Kurva ROC Random Forest menunjukkan performa yang sangat superior. Area di bawah kurva untuk Kelas 4 mencapai nilai sempurna 1.00, sedangkan Kelas 1 dan 2 mencapai 0.99. Bahkan untuk Kelas 3 yang paling sulit, nilai AUC meningkat drastis menjadi 0.97 (dibandingkan 0.91 pada C4.5). Nilai AUC yang mendekati 1.00 ini menegaskan bahwa Random Forest memiliki tingkat sensitivitas dan spesifisitas yang sangat tinggi dalam membedakan seluruh kategori kualitas tembakau. Terakhir, analisis kontribusi fitur sensor gas ditampilkan pada Gambar 15.



Gambar 15. Feature Importance – Random Forest

Pada Gambar 15, pola distribusi *feature importance* pada Random Forest terlihat sedikit berbeda dan lebih seimbang dibandingkan C4.5. Meskipun sensor MQ7 kini menjadi fitur yang paling dominan dengan skor di atas 0.40, kontribusi sensor MQ4 tetap sangat signifikan (0.37). Pergeseran dominasi fitur ini menunjukkan bahwa Random Forest mampu mengeksplorasi informasi dari kombinasi sensor secara lebih optimal, tidak hanya bergantung pada satu sensor saja seperti yang cenderung dilakukan oleh *single decision tree*.

Evaluasi lanjutan menggunakan metrik tambahan menunjukkan performa yang sangat baik pada model Random Forest. Nilai *Cohen's Kappa* sebesar 0.9259 dan *MCC* sebesar 0.9262 menegaskan konsistensi dan reliabilitas prediksi model. Selain itu, *Balanced Accuracy* sebesar 0.9450 menunjukkan kemampuan model dalam mengenali seluruh kelas secara seimbang. Adapun *Log Loss* sebesar 0.4291 mengindikasikan bahwa probabilitas prediksi yang dihasilkan model relatif akurat dan stabil.

TABEL 6
METRIK EVALUASI TAMBAHAN MODEL RANDOM FOREST

Metrik Evaluasi	Nilai
Cohen's Kappa	0.9259
MCC	0.9262
Balanced Accuracy	0.9450
Log Loss	0.4291

Perbandingan performa kedua algoritma dilakukan berdasarkan beberapa indikator evaluasi, yaitu akurasi, *Cohen's Kappa*, *Matthews Correlation Coefficient* (*MCC*), *balanced accuracy*, dan *cross-entropy log loss*. Selain itu, dilakukan analisis menggunakan *confusion matrix*, kurva ROC, serta *feature importance* untuk melihat kemampuan model dalam membedakan setiap kelas dan kontribusi masing-masing sensor terhadap proses klasifikasi.

IV. SIMPULAN

Berdasarkan hasil pengujian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa algoritma Random Forest memiliki performa yang lebih baik dibandingkan C4.5 dalam mengklasifikasikan kualitas tembakau berdasarkan pola aroma yang diperoleh dari perangkat *E-Nose*. Perbandingan dilakukan menggunakan beberapa parameter evaluasi, yaitu akurasi, *Cohen's Kappa*, *Matthews Correlation Coefficient (MCC)*, *balanced accuracy*, *cross-entropy log loss*, *confusion matrix*, kurva ROC, dan *feature importance*. Hasil pengujian menunjukkan bahwa Random Forest memperoleh akurasi sebesar 94,44%, lebih tinggi dibandingkan C4.5 yang memperoleh akurasi 90,56%. Performa C4.5 yang lebih rendah disebabkan karena algoritma ini hanya membangun satu *decision tree* sehingga lebih sensitif terhadap variasi data latih, *noise*, dan kemiripan pola respons sensor antar kelas, khususnya pada kelas dengan kondisi pengeringan transisi. Sebaliknya, Random Forest mampu mengatasi kelemahan tersebut melalui pendekatan *ensemble learning* yang membangun banyak *decision tree* menggunakan *bootstrap sampling* dan pemilihan subset fitur secara acak. Strategi tersebut menghasilkan model yang lebih stabil, mampu mereduksi varians, serta meningkatkan kemampuan generalisasi sehingga memberikan performa klasifikasi yang lebih baik. Dengan demikian, Random Forest dapat menjadi metode yang lebih efektif untuk klasifikasi kualitas tembakau berdasarkan data pola aroma yang diperoleh menggunakan *electronic nose*. Untuk penelitian selanjutnya, disarankan untuk menambah variasi sensor gas yang lebih spesifik terhadap senyawa organik volatil tembakau atau menerapkan metode *deep learning* untuk analisis data.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Nahdlatul Ulama Sunan Giri atas dukungan dan fasilitas yang diberikan selama pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada dosen pembimbing atas arahan, bimbingan, serta masukan yang konstruktif sehingga penelitian dan penulisan artikel ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] H. Rivai, S. Nanda, A. Alifo Yusuf, dan P. Akademi Pimpinan Perusahaan Jakarta, "Dampak Besar Industri Hasil Tembakau Bagi Ekonomi Indonesia," *J. Ekon. Revolusioner*, vol. 7, no. 1, hal. 130–135, 2024.
- [2] A. Bella *et al.*, "Macroeconomic impact of tobacco taxation in Indonesia," *Tob. Control*, vol. 33, hal. S108–S114, 2024, doi: 10.1136/tc-2022-057735.
- [3] N. Nur Salsabila, J. Tarik Ibrahim, dan G. Mumpuni Ningsih, "Analysis The Competitiveness of Indonesian Tobacco Exports to International Markets 2003–2022," *Int. J. Sci. Res. Manag.*, vol. 12, no. 01, hal. 479–487, 2024, doi: 10.18535/ijssm/v12i01.ah01.
- [4] L. Ramadhanti, B. Kusmanadhi, D. Wulanjari, dan S. B. Patricia SM, "Kualitas dan Beberapa Karakteristik Tembakau (*Nicotiana tabacum* L.) Rajangan Varietas Maesan 1 Akibat Teknik dan Lama Pengeringan Yang Berbeda," *Technologica*, vol. 2, no. 1, hal. 1–13, 2023, doi: 10.55043/technologica.v2i1.49.
- [5] M. Zhang *et al.*, "Integrated characterization of filler tobacco leaves: HS–SPME–GC–MS, E-nose, and microbiome analysis across different origins," *Bioresour. Bioprocess.*, vol. 11, no. 1, 2024, doi: 10.1186/s40643-024-00728-w.
- [6] M. Misbah, N. Arif, dan Y. A. Suryo, "Deteksi Kualitas Tembakau Berdasarkan Aroma Berbasis Electronic Nose," *E-Link J. Tek. Elektro dan Inform.*, vol. 16, no. 1, hal. 37, 2023, doi: 10.30587/e-link.v16i1.2096.
- [7] C. H. A. Esteves, B. A. Iglesias, T. Ogawa, K. Araki, L. Hoehne, dan J. Gruber, "Identification of Tobacco Types and Cigarette Brands Using an Electronic Nose Based on Conductive Polymer/Porphyrin Composite Sensors," *ACS Omega*, vol. 3, no. 6, hal. 6476–6482, 2018, doi: 10.1021/acsomega.8b00403.
- [8] Iqbal H. Sarker, "20 Machine Learning Algorithms, Real-World Applications and Research," 2021, *SN Computer Science*. doi: <https://doi.org/10.1007/s42979-021-00592-x>.
- [9] S. N. Bardab, T. M. Ahmed, T. Abdalkarim, dan A. Mohammed, "International Journal of Advanced and Applied Sciences Data mining classification algorithms: An overview," *Int. J. Adv. Appl. Sci.*, vol. 8, no. 2, hal. 1–5, 2021, doi: <https://doi.org/10.21833/ijaas.2021.02.001>.
- [10] M. Drogkoula, K. Kokkinos, dan N. Samaras, "applied sciences A Comprehensive Survey of Machine Learning Methodologies with Emphasis in Water Resources Management," *Appl. Sci.*, vol. 13, no. 22, hal. 1–44, 2023, doi: <https://doi.org/10.3390/app132212147>.
- [11] J. Ross Quinlan, "C4.5: Programs for Machine Learning," *Mach. Learn.*, vol. 7, no. 1, hal. 55–77, 1993, doi: 10.1016/S0019-9958(64)90259-1.
- [12] C. Paramita, F. A. Rafrastara, dan L. I. Kencana, "Pengembangan Sistem Klasifikasi Karakteristik Siswa Berbasis Website dengan menggunakan Algoritma C4 . 5," *J. Inform. J. Pengemb. IT*, vol. 8, no. 1, hal. 17–21, 2023, doi: <https://doi.org/10.30591/jpit.v8i1.4678>.
- [13] K. Lu dan M. Chen, "Research on Application of C4.5 Algorithm in Performance Analysis," *Proc. Int. Conf. Electron. Mech. Cult. Med.*, vol. 45, no. Emcm 2015, hal. 290–294, 2016, doi: 10.2991/emcm-15.2016.140.
- [14] T. B. Nugroho dan E. Sugiharti, "The Improvement of C4.5 Algorithm Accuracy in Predicting Forest Fires Using Discretization and AdaBoost," *J. Adv. Inf. Syst. Technol.*, vol. 3, no. 1, hal. 43–52, 2021, doi: 10.15294/jaist.v3i1.49094.
- [15] M. A. S. Mula Agung Barata, Edi Noersasongko, Purwanto, "Improving the Accuracy of C4.5 Algorithm with Chi-Square Method on Pure Tea Classification Using Electronic Nose," *Resti*, vol. 7, no. 2, hal. 226–235, 2023, doi: 10.29207/resti.v7i2.4687.
- [16] D. R. Wijaya, F. Afianti, A. Arifianto, D. Rahmawati, dan V. S. Kodogiannis, "Ensemble machine learning approach for electronic nose signal processing," *Sens. Bio-Sensing Res.*, vol. 36, no. April, hal. 100495, 2022, doi: 10.1016/j.sbsr.2022.100495.
- [17] Y. Ren, X. Zhu, K. Bai, R. Zhang, dan S. Member, "A New Random Forest Ensemble of Intuitionistic Fuzzy Decision Trees," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 5, hal. 1729–1741, 2022, doi: 10.1109/TFUZZ.2022.3215725.
- [18] H. K. Argretya, "Implementasi Random Oversampling pada Algoritma Random Forest untuk Klasifikasi Indeks Standar Udara di DKI Jakarta," *J. Inform. J. Pengemb. IT*, vol. 10, no. 4, hal. 1189–1198, 2025, doi: 10.30591/jpit.v10i4.9466.
- [19] J. S. Rhodes, A. Cutler, dan K. R. Moon, "Geometry- and Accuracy-Preserving Random Forest Proximities," *IEEE Trans. Pattern Anal.*

- Mach. Intell.*, vol. 45, no. 9, hal. 10947–10959, 2023, doi: <https://doi.org/10.1109/TPAMI.2023.3263774>.
- [20] E. Ismanto dan M. Novalia, “Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas,” *Techno.com*, vol. 20, no. 3, hal. 400–410, 2021, doi: <https://doi.org/10.33633/tc.v20i3.4576>.
- [21] S. Tirtosastro *et al.*, “Kandungan Kimia Tembakau dan Rokok,” *Bul. Tanam. Tembakau, Serat Miny. Ind.*, vol. 2, no. 1, hal. 33–43, 2010.
- [22] S. Wakhid, R. Sarno, S. I. Sabilla, dan D. B. Maghfira, “Detection and classification of indonesian civet and non-civet coffee based on statistical analysis comparison using E-Nose,” *Int. J. Intell. Eng. Syst.*, vol. 13, no. 4, hal. 56–65, 2021, doi: [10.22266/IJIES2020.0831.06](https://doi.org/10.22266/IJIES2020.0831.06).
- [23] J. V. M. Antony, P. Ramani, A. Ramasubramanian, dan G. Sukumaran, “Particle size penetration rate and effects of smoke and smokeless tobacco products – An invitro analysis,” *Heliyon*, vol. 7, no. 3, hal. e06455, 2021, doi: [10.1016/j.heliyon.2021.e06455](https://doi.org/10.1016/j.heliyon.2021.e06455).
- [24] M. Pilat-Rożek, E. Łazuka, D. Majerek, B. Szeląg, S. Duda-Saturnus, dan G. Łagód, “Application of Machine Learning Methods for an Analysis of E-Nose Multidimensional Signals in Wastewater Treatment,” *Sensors*, vol. 23, no. 1, 2023, doi: [10.3390/s23010487](https://doi.org/10.3390/s23010487).
- [25] Yixu Huang, Iyell-Joon Doh, Euiwon Bae, “Design and Validation of a Portable Machine Learning-Based Electronic Nose,” *Sensors (Switzerland)*, 2021, doi: <https://doi.org/10.3390/s21113923>.