

Uji Keakuratan Kinerja Tiga *Tree-Based* Model Berbasis *Classification* dalam Memprediksi Kualitas Udara di Daerah Khusus Jakarta

Muhizzadin Abdul Adzan¹, Binsar Hakim Aritonang¹, Agung Hari Saputra¹

¹Program Studi Meteorologi, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika,
Kota Tangerang, Indonesia

Email: muhizzadin12@gmail.com¹, hakimarios01@gmail.com¹, agung.hs@stmkg.ac.id¹

Received 10 Mei 2025 | Revised 20 Mei 2025 | Accepted 25 Mei 2025

ABSTRAK

Beberapa dekade terakhir, isu kualitas udara menjadi masalah di kota-kota besar seperti Daerah Khusus Jakarta. Studi ini bertujuan untuk mengevaluasi performa tiga *tree-based* model random forest (RF), *extra trees* (ET), dan *gradient boosting classifier* (GBC) dalam memprediksi serta mencari faktor terpenting yang memengaruhi kualitas udara. Model menggunakan data historis Automatic Weather Station dan ISPU Daerah Khusus Jakarta selama 2018 – 2021 yang dikembangkan untuk memprediksi 4 kategori kualitas udara (0-3). Model dilatih menggunakan 90% data sebagai set pelatihan dan 10% data sebagai set pengujian. Berdasarkan *feature importance plot* dari ketiga model tersebut menunjukkan bahwa parameter ISPU O₃ merupakan parameter yang sangat memengaruhi kualitas udara, serta arah dan kecepatan angin merupakan faktor meteorologis yang paling memengaruhi kualitas udara. Berdasarkan pengujian keakuratan model tersebut menunjukkan nilai yang hampir sama antara ketiga *tree-based* model tersebut dengan nilai *accuracy* 0.8646–0.8821, *precision* 0.8755–0.8870, *recall* 0.8646–0.8821, *f1-score* 0.8674–0.8816 serta *AUC* 0.9286–0.9441. Berdasarkan ketiga *tree-based* model tersebut, model RF menunjukkan kemampuan terbaik dibandingkan dengan dua model lainnya dengan nilai *accuracy* 0.8821, *precision* 0.8870, *recall* 0.8821, *f1-score* 0.8816, dan *AUC* 0.9425. Temuan ini diharapkan dapat mendorong penggunaan model RF sebagai sistem peringatan dini kualitas udara serta mendorong pentingnya pengendalian O₃ dalam kebijakan mitigasi.

Kata kunci: random forest, extra trees, gradient boosting classifier, kualitas udara, meteorologi

ABSTRACT

In recent decades, air quality has become an issue in large cities such as the Special Capital Region of Jakarta. This study aims to evaluate the performance of three *tree-based* models—random forest (RF), *extra trees* (ET), and *gradient boosting classifier* (GBC)—in predicting and identifying the most important factors affecting air quality. The models utilize historical data from Automatic Weather Stations and the Jakarta Special Region's Air Quality Index (AQI) from 2018 to 2021, developed to predict four air quality categories (0–3). The models were trained using 90% of the data as the training set and 10% as the testing set. Based on the *feature importance plots* of the three models, it was found that the ISPU O₃ parameter is a parameter that significantly influences air quality, while wind direction and speed are the meteorological factors that most influence air quality. Based on the accuracy testing of the models, the results show nearly identical values among the three *tree-based* models, with accuracy ranging from 0.8646 to 0.8821, precision from 0.8755 to 0.8870, recall from 0.8646 to 0.8821, F1-score 0.8674–0.8816, and AUC 0.9286–0.9441. Based on these three *tree-based* models, the RF model showed the best performance compared to the other two models with accuracy values of 0.8821, precision of 0.8870, recall of 0.8821, F1-score of 0.8816, and AUC of 0.9425. These findings are expected to encourage the use of the RF model as an early warning system for air quality and highlight the importance of O₃ control in mitigation policies.

Keywords: random forest, extra trees, gradient boosting classifier, air quality, meteorology

1. PENDAHULUAN

Beberapa dekade terakhir, isu pencemaran udara kian mencuat karena sangat memengaruhi kualitas udara dan kehidupan manusia serta lingkungan sekitarnya [1]. Pencemaran udara merupakan penyebab terjadinya penurunan kualitas udara akibat masuknya polutan-polutan ke udara yang disebabkan oleh aktivitas manusia maupun secara alami [2]. Pencemaran udara dapat berasal dari berbagai faktor, terutama akibat aktivitas manusia seperti penggunaan kendaraan berbahan bakar fosil, pembangkit listrik, aktivitas industri, aktivitas rumah tangga, pertambangan, dan sebagainya [3]. Pencemaran udara di suatu wilayah dapat ditentukan berdasarkan nilai ISPU (Indeks Standar Pencemaran Udara) yang merupakan pengukuran terhadap indeks kondisi mutu udara di suatu wilayah [4].

Parameter ISPU yang digunakan dalam menentukan kualitas udara suatu wilayah yaitu partikulat berukuran $10\ \mu\text{m}$ (PM_{10}), CO , SO_2 , dan NO_2 [5]. Selain itu, terdapat faktor-faktor parameter meteorologi yang turut memengaruhi tingkat kualitas udara di suatu wilayah [6]. Parameter-parameter meteorologi tersebut terdiri dari suhu udara, tekanan udara, angin, perawanan, kelembaban, radiasi matahari, dan sebagainya [7]. Semua parameter meteorologi tersebut dapat memengaruhi kualitas udara di suatu wilayah dengan tingkat korelasi yang berbeda-beda antar parameternya.

Dalam menggambarkan tingkat kualitas udara suatu wilayah, digunakan model yang memiliki tingkat keakuratan yang terbaik untuk menggambarkan tingkat kualitas udara di suatu wilayah. Penggunaan pemodelan dalam memprediksi suatu masalah karena pemodelan dapat menyelesaikan berbagai permasalahan serta dapat menghasilkan prediksi berdasarkan data-data sebelumnya [8]. *Tree-based* model menjadi salah satu model algoritma yang paling populer dan sukses dalam prakteknya karena memiliki keunggulan yang lebih baik dalam mempelajari hubungan yang kompleks dalam data [9]. Beberapa *tree-based* model yang populer dalam melakukan tugas klasifikasi dan regresi dengan membangun struktur seperti pohon untuk menentukan kelas atau nilai variabel target sesuai dengan fitur adalah *model random forest (RF)*, *extra trees (ET)*, dan *gradient boosting classifier (GBC)* [10].

Random forest adalah metode pembelajaran mesin terintegrasi yang menggunakan teknik pengambilan sampel acak *bootstrap* dan teknik pemisahan simpul acak untuk membangun beberapa pohon keputusan [11]. *Gradient boosting* merupakan metode pengajaran yang berbasis pohon keputusan dengan metode pembelajaran berbasis pohon klasifikasi awal dan terus menyesuaikan pohon baru dengan mengurangi fungsi kerugian [12]. Selain itu, *extra trees* memiliki kesamaan dengan *random forest* dalam membangun serangkaian pohon keputusan. Namun, keacakan model *extra trees* berasal dari pemisahan pengacakan pada semua pengamatan dan bukan berasal dari *bootstrap* data. Ketiga *tree-based* model tersebut memiliki kelebihan dan kekurangan masing-masing dalam melakukan prediksi.

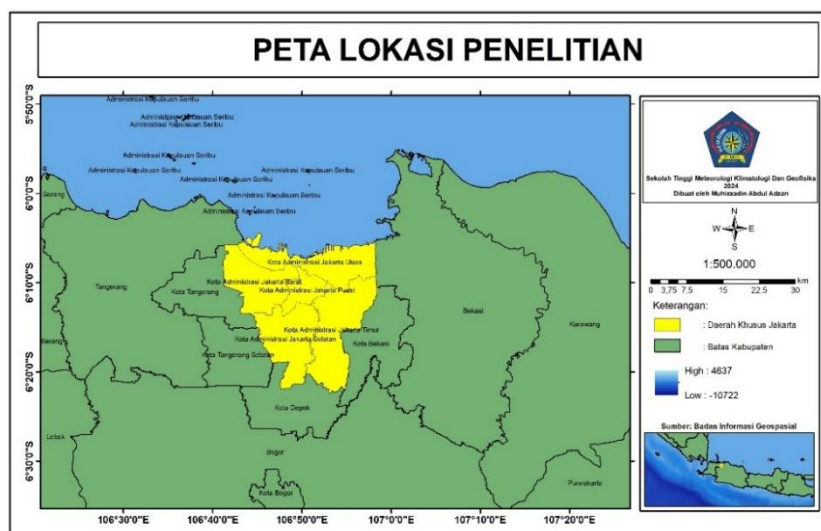
Pada penelitian ini, Daerah Khusus Jakarta dipilih sebagai lokasi penelitian dikarenakan Daerah Khusus Jakarta merupakan salah satu kota dengan tingkat kualitas udara terburuk di dunia [13]. Dalam penelitian ini, ketiga *tree-based* model *random forest (RF)*, *extra trees (ET)*, dan *gradient boosting classifier (GBC)* akan dibandingkan keakuratannya dalam memprediksi kualitas udara di Daerah Khusus Jakarta. Perbandingan keakuratan dari ketiga *tree-based* model ini menjadi penting karena dapat menentukan model terbaik yang dapat digunakan dalam memprediksi kualitas udara di suatu wilayah, terutama di Daerah Khusus Jakarta.

2. METODOLOGI

2.1 Lokasi dan data penelitian

Lokus penelitian ini dilakukan di wilayah Daerah Khusus Jakarta yang merupakan salah satu kota dengan tingkat pencemaran udara terburuk di Indonesia. Secara astronomi, Daerah Khusus Jakarta terletak sekitar $5,32^\circ - 6,3983^\circ$ LS dan $106,3783^\circ - 106,9717^\circ$ BT seperti yang digambarkan pada Gambar 1. Selain itu, penelitian ini berfokus pada periode 2018 hingga 2021 dalam menguji kinerja

ketiga *tree-based* model dalam memprediksi kualitas udara dan menghasilkan analisis yang komprehensif.



Gambar 1. Lokasi Penelitian

Dalam penelitian ini, *dataset* yang digunakan yaitu parameter Indeks Standar Pencemaran Udara (ISPU) Daerah Khusus Jakarta dan parameter meteorologi dengan periode waktu selama 4 tahun (2018 - 2021). Data ISPU Daerah Khusus Jakarta yang digunakan yaitu PM₁₀, CO, NO₂, O₃, dan SO₂ yang dapat diperoleh melalui web Jakarta Open Data (<https://data.jakarta.go.id/dataset>). Selain itu, parameter meteorologis yang digunakan yaitu suhu udara maksimum dan minimum (*tt_air_max* dan *tt_air_min*), tekanan udara (*pp_air*), arah dan kecepatan angin (*wd_avg* dan *ws_avg*), kelembaban (*rh_avg*), curah hujan (*rr*), dan radiasi matahari rata-rata dan maksimal (*sr_avg* dan *sr_max*) yang dapat diperoleh melalui data AWS BMKG (<https://awscenter.bmkg.go.id/accessdata>).

Data yang telah diperoleh kemudian disatukan ke dalam format *excel* (*.xlsx) yang kemudian data akan diperbaiki dengan *reprocessing* terhadap data yang ganda, hilang, atau tidak [14]. Selanjutnya, akan dilakukan teknik rekayasa fitur untuk mengekstraksi fitur-fitur yang relevan dalam proses pemodelan seperti waktu, ISPU, suhu, kelembaban, tekanan, angin, hujan, dan perawanan [15]. Selain itu, teknik ini juga berguna untuk memilih, memanipulasi, dan mengubah data mentah menjadi fitur yang dapat digunakan dalam pembelajaran model serta menentukan target dari data yang ada. Target sarannya adalah kualitas udara wilayah Jakarta dengan 4 kategori yaitu sehat (0), sedang (1), dan tidak sehat (2), sangat tidak sehat (3) berdasarkan Tabel 1 [5].

Tabel 1. Target Sasaran Kualitas Udara dalam *Machine Learning*

Klasifikasi Kualitas Udara	Target (<i>binary numbers</i>)
Sehat	0
Sedang	1
Tidak sehat	2
Sangat tidak sehat	3

Data yang telah diproses sebelumnya akan dibagi menjadi *set* pelatihan dan pengujian untuk menentukan keakuratan model dan pembelajaran model [16]. Dalam penelitian ini, *set* pelatihan model yang digunakan yaitu 90% dari total data serta *set* pengujian model sebesar 10% [17]. *Set* pelatihan digunakan untuk melatih ketiga *tree-based* model, sedangkan *set* pengujian digunakan untuk mengevaluasi performa model [15]. Ketiga *tree-based* model *random forest* (RF), *extra trees* (ET), dan

gradient boosting classifier (GBC) akan mempelajari *dataset* pelatihan sebelumnya dan mengevaluasi performa model dari *dataset* pengujian.

2.2 Penentuan Konfigurasi Hypermaters

Penentuan konfigurasi *hyperparameters* menjadi hal yang sangat krusial dalam sebuah pemodelan karena konfigurasi *hyperparameter* berperan dalam penentuan parameter terbaik yang akan mengatur proses pembelajaran dan pengujian pada sebuah pemodelan [18]. Dalam penelitian ini, konfigurasi *hyperparameters* yang digunakan yaitu pada Tabel 2.

Tabel 2. Tabel *Hypermaters*

Description	Value
Session id	123
Target Type	Multiclass
Preprocess	True
Fix imbalance	True
Normalize	True
Use GPU	False
Log Experiment	False
Fix imbalance method	SMOTE

2.3 Analisis Parameter Penting

Pada ketiga *tree-based* model dapat dianalisis parameter-parameter yang memengaruhi kategori kualitas udara di Daerah Khusus Jakarta. Analisis fitur yang memengaruhi kualitas udara di wilayah Daerah Khusus Jakarta dilakukan berdasarkan *feature importance plot* [19]. Dari hasil *feature importance plot* tersebut dapat diketahui parameter ISPU dan meteorologi yang memiliki kontribusi yang besar dalam memengaruhi kualitas udara di wilayah Daerah Khusus Jakarta.

2.4 Evaluasi Performa

Performa pemodelan dari *tree-based* model dapat dievaluasi menggunakan *set* pengujian berdasarkan nilai *accuracy*, *precision*, *recall*, *F1-score*, *confusion metrics*, *receiver operating characteristic* (ROC) dan *area under the curve* (AUC) [15] (Saputra & Oktarina, 2022). Nilai *accuracy*, *precision*, *recall*, *confusion metrics*, dan *F1-score* dapat dijelaskan melalui Tabel 3 dan Persamaan 1 – 4 berikut [20 - 22]. Selain itu, Menurut Florin Gorunescu (2011) [23], pedoman klasifikasi keakuratan pengujian AUC dijabarkan melalui Tabel 4.

Tabel 3. Tabel *Confusion Matrix*

Confusion Matrix		Actual Class	
		Positive	Negative
Predicted Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Tabel 4. Tabel klarifikasi keakuratan pengujian AUC

Skala	Keterangan
0.90 -1.00	Excellent Classification
0.80 -0.90	Good Classification
0.70 -0.80	Fair Classification
0.60 -0.70	Poor Classification
0.50 -0.60	Failure

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Sensitivity/Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right) \quad (4)$$

Keterangan:

TP : *True Positive* (data positif yang diidentifikasi positif)

TN : *True Negative* (data negatif yang diidentifikasi negatif)

FP : *False Positive* (data negatif yang diidentifikasi positif)

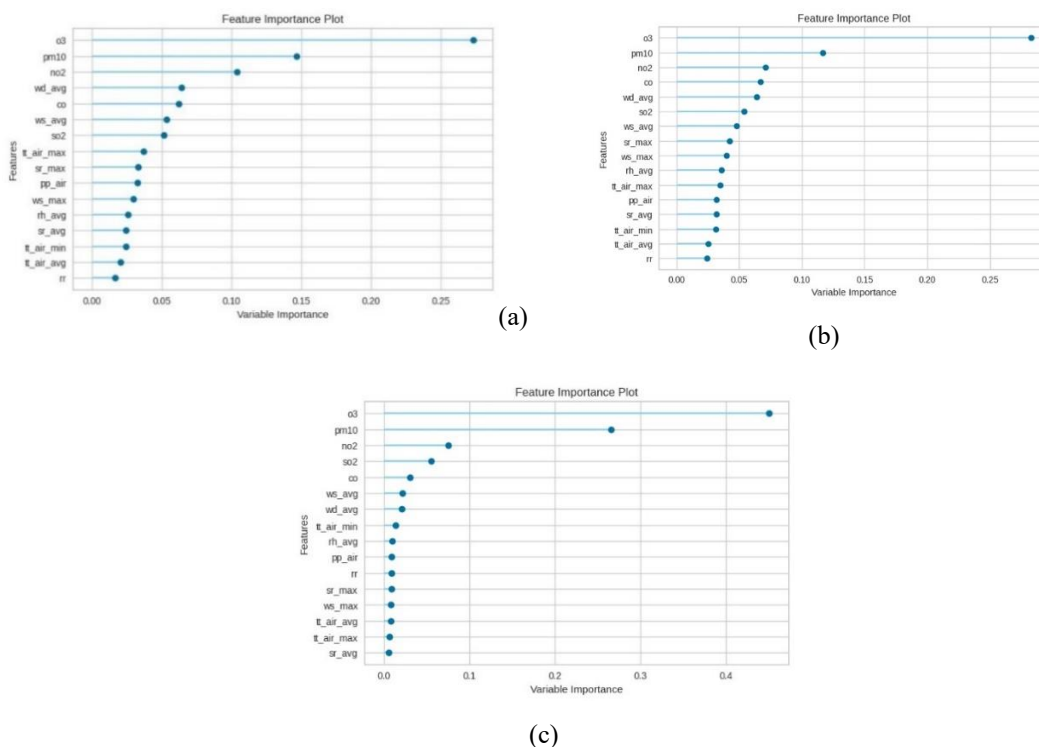
FN : *False Negative* (data positif yang diidentifikasi negatif)

3. HASIL DAN PEMBAHASAN

3.1 Important Features

Analisis fitur yang memengaruhi kualitas udara di wilayah Daerah Khusus Jakarta dilakukan berdasarkan *feature importance plot* [19]. Gambar 2 merupakan *feature importance plot* dari model *machine learning*. *Feature importance* digunakan dalam menentukan parameter-parameter yang paling memengaruhi atau memiliki kontribusi yang semakin tinggi terhadap sebuah variabel target yang dituju [24]. Fitur terpenting dalam grafik terbagi dalam 2 kategori yaitu parameter ISPU dan meteorologis.

Gambar 2(a) merupakan model *random forest* dimana pada pemodelan ini, O_3 merupakan parameter yang paling penting serta arah angin merupakan parameter meteorologis yang paling memengaruhi kualitas udara di Daerah Khusus Jakarta. Pada Gambar 2(b), model *extra trees classifier* menunjukkan bahwa parameter O_3 merupakan parameter terpenting yang memengaruhi kualitas udara serta arah angin merupakan parameter meteorologi yang paling memengaruhi kualitas udara di wilayah Daerah Khusus Jakarta. Selain itu, pada Gambar 2(c), model *gradient boosting classifier* juga memberikan informasi bahwa O_3 merupakan fitur terpenting serta arah dan kecepatan angin merupakan parameter meteorologi yang memengaruhi kualitas udara di wilayah Daerah Khusus Jakarta. Dari ketiga model tersebut, dapat dilihat bahwa parameter yang paling memengaruhi kualitas udara di Daerah Khusus Jakarta yakni O_3 . Dari segi meteorologi, parameter arah dan kecepatan angin menjadi parameter yang turut berkontribusi dalam memengaruhi tingkat kualitas udara di Daerah Khusus Jakarta.



Gambar 2. *Feature Importance Plot* dari ketiga model (a) *Random Forest*, (b) *Extra Trees Classifier*, dan (c) *Gradient Boosting Classifier*

3.2 Uji keakuratan model

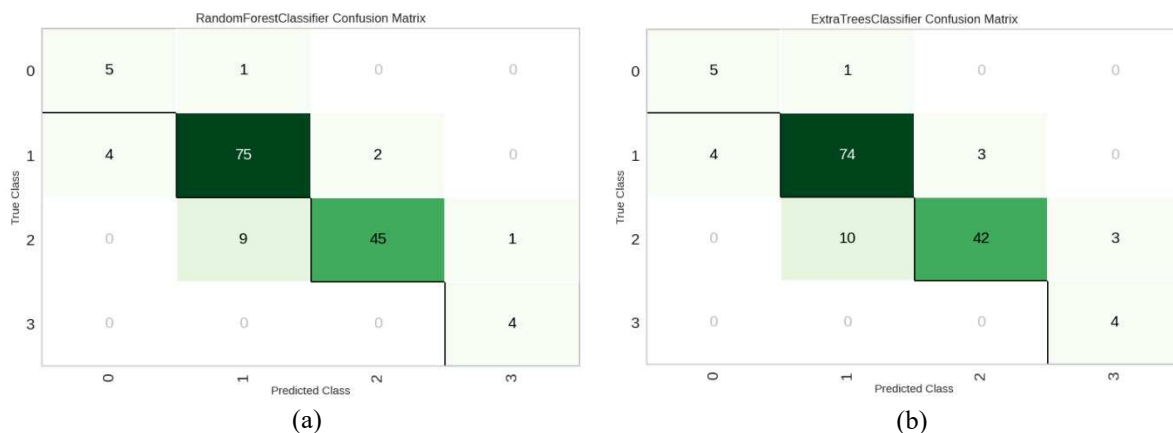
Perbandingan performa ketiga model ditampilkan pada Tabel 5 berikut. Masing-masing model dapat dievaluasi dengan membandingkan nilai yang tertinggi. Beberapa metrik umum dalam pembelajaran *machine learning* diantaranya, yakni *accuracy*, *recall*, *precision*, *area under the curve* (AUC), dan F1 [25]. Dari data yang ditampilkan, dapat dilihat bahwa *gradient boosting classifier* memiliki nilai AUC lebih tinggi daripada model-model lainnya. Hal tersebut menunjukkan bahwa performa *gradient boosting classifier* lebih baik membedakan antara kelas positif dan negatif dibandingkan model lainnya. Selain itu, berdasarkan nilai *accuracy*, *recall*, *precision*, dan F1 dapat dilihat bahwa nilai pemodelan *random forest classifier* yang lebih tinggi dibandingkan dengan model-model lainnya. Akan tetapi, secara umum, ketiga model tersebut menunjukkan tingkat keakuratan yang baik, dimana memiliki nilai AUC di atas 0,9 yang termasuk *excellent classification* [23].

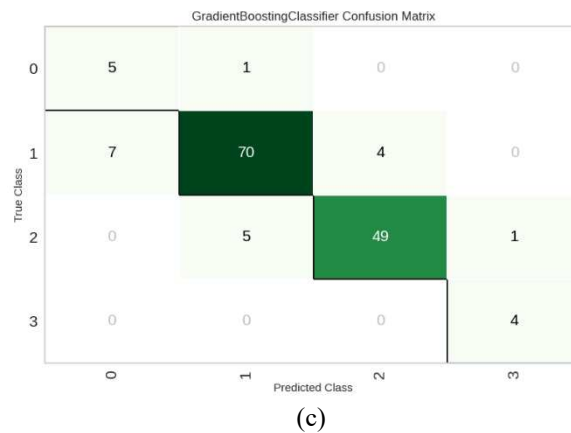
Tabel 5. Performa tiga model oleh matrik

<i>Model</i>	<i>Acc.</i>	<i>Auc.</i>	<i>Rec.</i>	<i>Prec.</i>	<i>F1</i>
<i>Random Forest Classifier (rf)</i>	0.8821	0.9425	0.8821	0.8870	0.8816
<i>Extra Trees Classifier (et)</i>	0.8669	0.9286	0.8669	0.8719	0.8648
<i>Gradient Boosting Classifier (gbc)</i>	0.8646	0.9441	0.8646	0.8755	0.8674

Selain menggunakan *accuracy*, *recall*, *precision*, *area under the curve* (AUC), dan F1 dalam menguji keakuratan ketiga model tersebut, pengujian kinerja model juga dapat dilakukan menggunakan *confusion matrix* seperti pada Gambar 3 berikut. Matriks kebingungan (*confusion matrix*) merupakan metode yang efektif untuk menilai kinerja pengklasifikasi serta dapat memberikan analisis terhadap kesalahan khusus pada parameter yang diuji seperti mencakup kesalahan *false positive* (kesalahan inklusi) dan *false negative* (kesalahan eksklusi) [26]. Pada ketiga gambar *confusion matrix* menunjukkan empat kelas (0,1,2,3) dengan memuat diagonal atas sebelah kiri hingga ke kanan bahwa menunjukkan jumlah prediksi yang benar setiap kelas.

Pada gambar 3(a) model *random forest* memuat data dari kelas 0, 1, 2, dan 3 dengan memberikan *prediction class* yang sama dengan *true class* masing-masing sebanyak 5, 75, 45, dan 4 data. Terdapat kesalahan di mana model menghasilkan *prediction class* yang berbeda dengan *true class* pada masing-masing kelas dengan total kesalahan kelas sekitar 17 data. Gambar 3(b) model *extra trees classifier* menunjukkan bahwa pada kelas 0, 1, 2, dan 3 memberikan *prediction class* yang sama dengan *true class* masing-masing sebanyak 5, 74, 42, dan 4 data. Kesalahan pada model tersebut menghasilkan *prediction class* yang berbeda dengan *true class* pada masing-masing kelas dengan total kelas yang berbeda sekitar 21 data. Selain itu, pada Gambar 3(c) model *gradient boosting classifier* menjelaskan bahwa pada kelas 0, 1, 2, dan 3 memberikan *prediction class* yang sama dengan *true class* masing-masing sebanyak 5, 70, 49, dan 4 data. Kesalahan pada model tersebut menghasilkan *prediction class* yang berbeda dengan *true class* pada masing-masing kelas dengan total kelas yang berbeda sekitar 18 data.





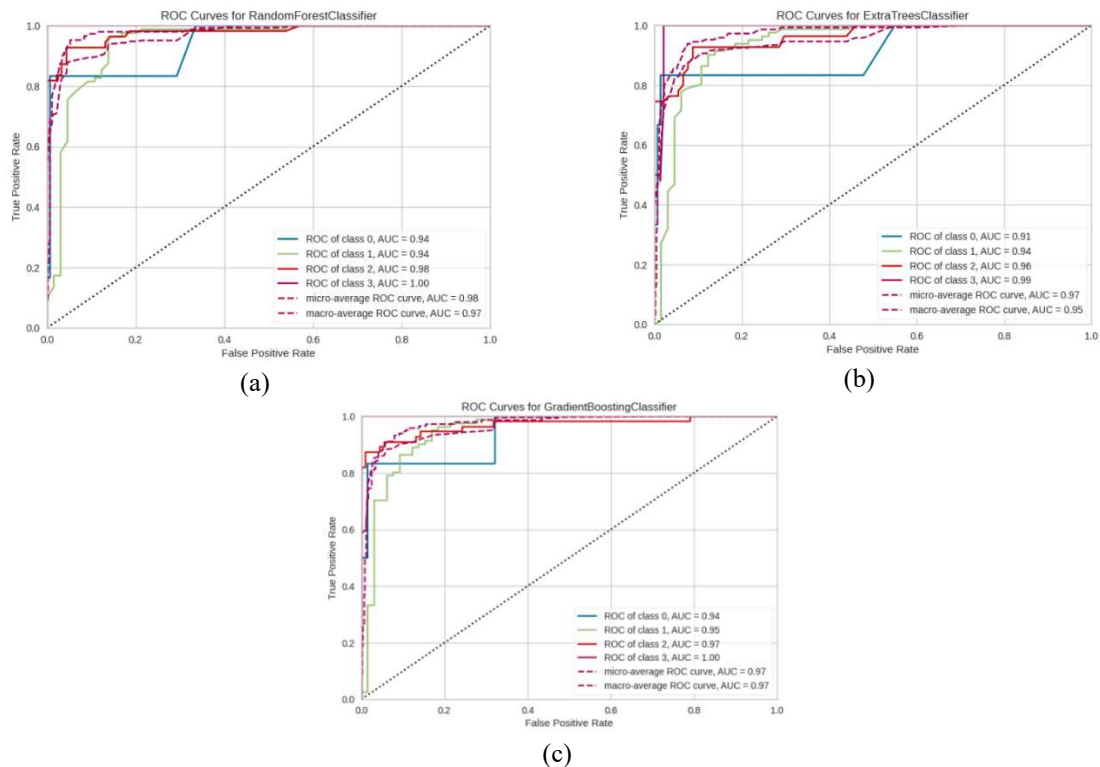
Gambar 3. Confusion Matrix dari tiga model (a) Random Forest, (b) Extra Trees Classifier, dan (c) Gradient Boosting Classifier

Gambar 4 merupakan grafik *receiver operating characteristic* (ROC). Kurva AUC - ROC merupakan kurva yang digunakan dalam menentukan permasalahan klasifikasi pengaturan pada nilai ambang batas berdasarkan hasil pengukuran kinerja. ROC adalah kurva kemungkinan serta AUC merupakan nilai yang menyatakan nilai atau derajat yang mana semakin mendekati 1, maka semakin baik model dalam membedakan kelas [27]. Klasifikasi nilai AUC dapat dilihat pada Tabel 4 di atas berdasarkan Gorunescu (2011) [23].

Gambar 4(a) model *random forest classifier* memberikan informasi nilai bahwa kelas 0 dan 1 memiliki AUC sebesar 0,94 yang menunjukkan kriteria yang *excellent classification*. Kemudian kelas 2 memiliki nilai sebesar 0,98 yang mendekati klasifikasi sempurna. Kelas 3 memiliki AUC sebesar 1,00 yang menunjukkan klasifikasi sempurna. Selain itu, gambar tersebut memberikan informasi bahwa model *random forest classifier* menghasilkan kurva ROC rata-rata mikro (AUC = 0,98) dan kurva ROC rata-rata makro (AUC = 0,97).

Pada Gambar 4(b) model *extra trees classifier* memberikan informasi bahwa nilai AUC pada kelas 0 sebesar 0,91 yang menunjukkan model *excellent classification*. Kelas 1 memiliki nilai 0,94 yang menunjukkan kinerja *excellent classification*. Kelas 2 memiliki nilai AUC sebesar 0,96 yang memberikan informasi bahwa model ini memiliki kinerja *excellent classification*. Pada kelas 3 memiliki nilai AUC sebesar 0,99 menunjukkan kinerja yang hampir sempurna. Selain itu, model *extra trees classifier* menghasilkan kurva ROC rata-rata mikro (AUC = 0,97) dan kurva ROC rata-rata makro (AUC = 0,95).

Gambar 4(c) model *gradient boosting classifier* memberikan informasi nilai AUC kelas 0 sebesar 0,94 yang menunjukkan bahwa model memiliki kemampuan yang *excellent classification*. Kelas 1 memiliki nilai AUC sebesar 0,95 menunjukkan klasifikasi kinerja yang *excellent classification*. Pada kelas 2 memiliki nilai 0,97 menunjukkan kinerja yang *excellent classification* serta kelas 3 dengan nilai AUC 1,00 yang menunjukkan kinerja yang sempurna. Selain itu, model tersebut menghasilkan kurva ROC rata-rata mikro dan kurva ROC rata-rata makro masing-masing AUC = 0,97. Berdasarkan nilai AUC dari ketiga model di atas, rata-rata AUC dari ketiga model berada di atas 0,9 yang menunjukkan *excellent classification*.



Gambar 4. Grafik receiver operating characteristic (ROC) dari ketiga model (a) Random Forest (b) Extra Trees Classifier (c) Gradient Boosting Classifier

4. KESIMPULAN

Penelitian di atas menunjukkan performa dari tiga *tree-based* model *random forest* (RF), *extra trees* (ET), dan *gradient boosting classifier* (GBC) dalam memprediksi kualitas udara di wilayah Daerah Khusus Jakarta. Data yang digunakan dalam penelitian ini berasal dari data AWS BMKG dan ISPU Daerah Khusus Jakarta selama periode 2018 – 2021 dengan parameter yang digunakan yaitu parameter ISPU dan meteorologi. Berdasarkan *feature importance plot* dari ketiga model tersebut menunjukkan bahwa parameter ISPU O_3 merupakan parameter yang sangat memengaruhi kualitas udara di Daerah Khusus Jakarta, serta arah dan kecepatan angin merupakan faktor meteorologis yang paling memengaruhi kualitas udara. Di samping itu, berdasarkan uji keakuratan ketiga *tree-based* model tersebut, model *random forest* (RF) menunjukkan kemampuan terbaik dalam memprediksi kualitas udara di Daerah Khusus Jakarta dengan nilai *accuracy* 0.8821, *precision* 0.8870, *recall* 0.8821, *f1-score* 0.8816 serta AUC 0.9425. Selain itu, berdasarkan model *extra trees* (ET) dan *gradient boosting classifier* (GBC) secara berurutan memiliki nilai *accuracy* 0.8669 dan 0.8646, *precision* 0.8719 dan 0.8755, *recall* 0.8669 dan 0.8646, *f1-score* 0.8648 dan 0.8674, serta AUC 0.9286 dan 0.9441. Berdasarkan *confusion matrix*, model RF, ET, dan GBC dapat memberikan *prediction class* yang sama dengan *true class* masing-masing sebanyak 129, 125, dan 128 data. Berdasarkan AUC dari ketiga model tersebut, rata-rata AUC ketiga model tersebut berada di atas 0,9 yang menunjukkan *excellent classification*.

UCAPAN TERIMA KASIH

Terima kasih sebesar-besarnya diberikan kepada Program Studi Meteorologi, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika (STMKG) serta para dosen STMKG yang telah memberikan ilmu, dukungan, dan bantuan dalam penelitian ini. Ucapan terima kasih juga diberikan kepada BMKG

dan Dinas Komunikasi, Informatika, dan Statistik Provinsi DKI Jakarta yang telah menyediakan data AWS BMKG serta Indeks Standar Pencemaran Udara (ISPU) Jakarta yang menjadi data utama dalam penelitian ini. Kontribusi dari semua pihak tersebut menjadi kunci dalam menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] Wahab, M. I. (2021). Health impacts of air pollution. In Environmental Sustainability-Preparing for Tomorrow. IntechOpen.
- [2] Kiswandono, A. A. (2017). Kajian Indeks Standar Polusi Udara (ISPU) Nitrogen Dioksida (NO₂) Di Tiga Lokasi Kota Bandar Lampung. Analit: Analytical and Environmental Chemistry, 2(1), 42-51.
- [3] Lelieveld, J., Haines, A., Burnett, R., Tonne, C., Klingmüller, K., Münzel, T., & Pozzer, A. (2023). Air pollution deaths attributable to fossil fuels: observational and modelling study. *bmj*, 383. doi: 10.1136/bmj-2023-077784
- [4] Kurniawan, A. (2018). Pengukuran parameter kualitas udara (CO, NO₂, SO₂, O₃ dan PM₁₀) di Bukit Kototabang berbasis ISPU. Jurnal Teknosains, 7(1), 1-13. doi: 10.22146/teknosains.34658
- [5] Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia. (2020). Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor P.14/MENLHK/SETJEN/KUM.1/7/2020 tentang Indeks Standar Pencemar Udara. <https://peraturan.bpk.go.id/Details/163466/permen-lhk-no-14-tahun-2020>
- [6] Istirokhatun, T., Agustini, I. T., & Sudarno, S. (2016). Investigasi pengaruh kondisi lalu lintas dan aspek meteorologi terhadap konsentrasi pencemar SO₂ di Kota Semarang. *Jurnal Presipitasi: Media Komunikasi dan Pengembangan Teknik Lingkungan*, 13(1), 21-27.
- [7] Serlina, Y. (2020). Pengaruh Faktor Meteorologi Terhadap Konsentrasi NO₂ di Udara Ambien (Studi Kasus Bundaran Hotel Indonesia Daerah Khusus Jakarta). *Jurnal Serambi Engineering*, 5(3). doi: 10.32672/jse.v5i3.2146
- [8] Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan machine learning dalam berbagai bidang. *Jurnal Khatulistiwa Informatika*, 5(1), 490845.
- [9] Wojciech, S. (2020). Learning with explainable trees. *Nature Machine Intelligence*, 2(1), 16-17. doi: 10.1038/s42256-019-0142-0
- [10] Alkahfi, C., Santoso, Z. R., Fitrianto, A., & Oktarin, S. D. (2022). Variable Importance Kesehatan dan Pendidikan dalam Pembentukan IPM dengan Algoritme Machine Learning. *Sains Dan Informatika: Research Of Science And InformatiC*, 8(2), 77-85. doi: 10.22216/jsi.v8i2.1623
- [11] Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- [12] Suryana, S. E., Warsito, B., & Suparti, S. (2021). Penerapan Gradient Boosting Dengan Hyperopt Untuk Memprediksi Keberhasilan Telemarketing Bank. *Jurnal Gaussian*, 10(4), 617-623. doi: 10.14710/j.gauss.10.4.617-623
- [13] Umri, S. S. A. (2021). Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Daerah Khusus Jakarta. *JIKO (Jurnal Informatika dan Komputer)*, 4(2), 98-104. doi: 10.33387/jiko.v4i2.2871
- [14] Halim, K., Herwindiati, D. E., & Sutrisno, T. (2023). Penerapan Metode Decision Tree Untuk Prakiraan Cuaca Kota Bekasi. *Jurnal Ilmu Komputer dan Sistem Informasi*, 11(2). doi: 10.24912/jiksi.v11i2.26026
- [15] Saputra, A. H., & Oktarina, S. (2022). Comparing the Performance of Three Decision Tree Models for Precipitation Prediction in Cengkareng Soekarno Hatta. *Jurnal Aplikasi Meteorologi*, 1(2). doi: 10.36754/jam.v1i2.318
- [16] Vrigazova, B. (2021). The proportion for splitting data into training and test set for the bootstrap in classification problems. *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy*, 12(1), 228-242.
- [17] Oktaviani, C., & Afdal, A. (2014). Prediksi curah hujan bulanan menggunakan jaringan syaraf tiruan dengan beberapa fungsi pelatihan backpropagation (studi kasus: stasiun meteorologi tabing padang, tahun 2001-2012). *J. Fis. Unand*, 2(4), 228-237. doi: 10.25077/jfu.2.4.%25p.2013

- [18] Vincent, A. M., & Jidesh, P. (2023). An improved hyperparameter optimization framework for AutoML systems using evolutionary algorithms. *Scientific Reports*, 13(1), 4737.
- [19] Sihombing, C., Saputra, A. H., Sari, F. P., & Mulya, A. (2022). Prediksi Curah Hujan di Wilayah DKI Jakarta dengan Model NeuralProphet. *Jurnal Aplikasi Meteorologi*, 1(2). doi: 10.36754/jam.v1i2.317
- [20] Andika, L. A., Azizah, P. A. N., & Respatiwan, R. (2019). Analisis sentimen masyarakat terhadap hasil quick count pemilihan presiden indonesia 2019 pada media sosial twitter menggunakan metode naive bayes classifier. *Indonesian Journal of Applied Statistics*, 2(1), 34-41. doi: 10.13057/ijas.v2i1.29998
- [21] Prasetyawan, P., Ahmad, I., Borman, R. I., Pahlevi, Y. A., & Kurniawan, D. E. (2018, October). Classification of the period undergraduate study using back-propagation neural network. In *2018 International Conference on Applied Engineering (ICAE)* (pp. 1-5). IEEE. doi: 10.1109/INCAE.2018.8579389
- [22] Mulyanto, A., Susanti, E., Rossi, F., Wajiran, W., & Borman, R. I. (2021). Penerapan Convolutional Neural Network (CNN) pada Pengenalan Aksara Lampung Berbasis Optical Character Recognition (OCR). *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 7(1), 52-57. doi: 10.26418/jp.v7i1.44133
- [23] Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media.
- [24] Pambudi, A. (2023). Penerapan Crisp-Dm Menggunakan Mlr K-Fold Pada Data Saham Pt. Telkom Indonesia (Persero) Tbk (Tlkm)(Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022). *Jurnal Data Mining dan Sistem Informasi*, 4(1), 1-14. doi: 10.33365/jdmsi.v4i1.2462
- [25] Intan, I., Rismayani, R., Nurdin, N., & Koswara, A. T. (2021). Performance Analysis of Weather Forecasting using Machine Learning Algorithms (Analisis Performansi Prakiraan Cuaca Menggunakan Algoritma Machine Learning). *Pekommas*, 6(2), 1-8. doi: 10.30818/jpkm.2021.2060221
- [26] FADLI, M., & Saputra, R. A. (2023). Klasifikasi Dan Evaluasi Performa Model Random Forest Untuk Prediksi Stroke. *Jurnal Teknik*, 12(2).
- [27] Hendrawan, A., Huizen, L. M., Pinem, A. P. R., & Wicaksana, D. A. (2021, November). Implementasi Pemilihan Fitur Metode Wrapper dan Embedded dalam Prediksi Ketepatan Kelulusan Mahasiswa. In *Seminar Nasional Penelitian dan Pengabdian Kepada Masyarakat* (pp. 330-335). doi: 0000-0002-0732-6447