

**ANALISIS KOMPARATIF ALGORITMA RANDOM
FOREST DAN DECISION TREE DALAM
KLASIFIKASI KUALITAS UDARA
DI DKI JAKARTA**

SKRIPSI SARJANA KOMPUTER

Oleh

Shinta Dwi Rahayu

217006516035



PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL

2025

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Shinta Dwi Rahayu
NPM : 217006516045
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : 24 Februari 2026

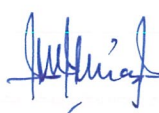
JUDUL DALAM BAHASA INDONESIA :

ANALISIS KOMPARATIF ALGORITMA RANDOM FOREST DAN DECISION TREE DALAM KLASIFIKASI KUALITAS UDARA DI DKI JAKARTA

JUDUL DALAM BAHASA INGGRIS :

COMPARATIF ANALYSIS OF RANDOM FOREST AND DECISION TREE ALGORITMA IN AIR QUALITY CLASSIFICATION IN DKI JAKARTA

TANDA TANGAN DAN TANGGAL

Pembimbing 1	Ka. Prodi	Mahasiswa
TGL :	TGL : 2/3 2026	TGL : 02 MARET 2026
		

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

ANALISIS KOMPARATIF ALGORITMA RANDOM FOREST DAN DECISION TREE DALAM KLASIFIKASI KUALITAS UDARA DI DKI JAKARTA

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2025-2026 pada tanggal 24 Februari Tahun 2026.

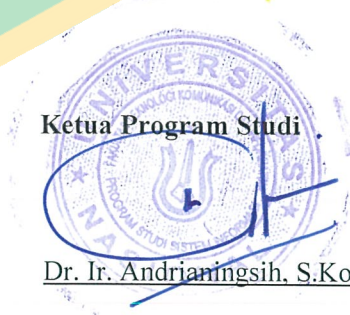
Dosen Pembimbing 1



Sari Ningsih, S.Si., M.M.

NIDN 0302066701

Ketua Program Studi



Dr. Ir. Andrianingsih, S.Kom., MMSI

NIDN 0303097902

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Program Studi Sistem Informasi, Universitas Nasional, Fakultas Teknologi Komunikasi dan Informatika, saya yang bertanda tangan di bawah ini:

Nama : Shinta Dwi Rahayu

NIM : 217006516035

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Sekolah Universitas Nasional, Fakultas Teknologi Komunikasi dan Informatika, Hak Bebas Royalti Noneksklusif (Non-exclusive Royalti Free Right) atas karya ilmiah saya yang berjudul:

Analisis Komparatif Algoritma Random Forest dan Decision Tree dalam Klasifikasi Kualitas Udara di DKI Jakarta.

Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Universitas Nasional berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 2 Maret 2026

Yang menyatakan



(Shinta Dwi Rahayu)

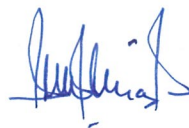
HALAMAN PENGESAHAN

TUGAS AKHIR

ANALISIS KOMPARATIF ALGORITMA RANDOM FOREST DAN
DECISION TREE DALAM KLASIFIKASI KUALITAS UDARA DI DKI
JAKARTA



Dosen Pembimbing 1



(Sari Ningsih, S.Si., M.M.)

HALAMAN PENGESAHAN TUGAS SARJANA

ANALISIS KOMPARATIF ALGORITMA RANDOM FOREST DAN DECISION TREE DALAM KLASIFIKASI KUALITAS UDARA DI DKI JAKARTA



Pembimbing I

(Sari Ningsih, S.Si., M.M.)

NIP: 0302066701

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

ANALISIS KOMPARATIF ALGORITMA RANDOM FOREST DAN DECISION TREE DALAM KLASIFIKASI KUALITAS UDARA DI DKI JAKARTA

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 27 Februari 2026



Shinta Dwi Rahayu

Npm: 217006516035

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan etika dan kaidah penulisan karya ilmiah yang berlaku, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Yang menyatakan,

Nama : Shinta Dwi Rahayu

NIM : 217006516035

Tanda Tangan :

Tanggal : 15 Desember 2025

Mengetahui

Pembimbing I : Sari Ningsih, S.Si., M.M. (



ABSTRAK

Kualitas udara adalah salah satu dari indikator dalam menentukan kondisi lingkungan suatu wilayah. Tingkat polusi udara yang signifikan di wilayah DKI Jakarta memerlukan sistem monitoring yang dapat menyajikan informasi secara responsif dan akurat. Dalam penelitian ini dilakukan proses klasifikasi kualitas udara di wilayah Jakarta berdasarkan Indeks Standar Pencemaran Udara (ISPU) dengan memanfaatkan dua algoritma Machine Learning, yaitu Decision Tree dan Random Forest. Sumber data dalam penelitian ini berasal dari dataset ISPU Jakarta yang diakses melalui portal Open Data DKI Jakarta, dengan lima parameter polutan yang digunakan sebagai variabel pengukuran, yaitu PM10, SO₂, CO, O₃, dan NO₂. Dalam implementasinya, penelitian ini menggunakan perangkat kerja CRISP-DM (Cross Industry Standard Process for Data Mining) yang mencakup lima tahapan utama dalam proses analisis data, yaitu data understanding, data cleaning, data preparation, modeling, dan evaluation. Dataset yang digunakan pada penelitian ini selanjutnya dipisahkan menjadi dua bagian, yaitu 80% sebagai data pelatihan (training data) dan 20% sebagai data pengujian (testing data). Untuk menilai performa model klasifikasi yang dihasilkan, proses evaluasi dilakukan dengan memanfaatkan confusion matrix serta beberapa metrik pengukuran kinerja, yaitu accuracy, precision, recall, dan F1-score.

Hasil pengujian menunjukkan bahwa algoritma Decision Tree memperoleh tingkat akurasi sebesar 80,75%, sedangkan algoritma Random Forest menghasilkan akurasi 85,17%. Perbandingan tersebut mengindikasikan bahwa algoritma Random Forest memiliki performa klasifikasi yang lebih baik dibandingkan Decision Tree dengan selisih nilai akurasi sebesar 4,41%. Hasil klasifikasi yang dihasilkan selanjutnya diimplementasikan dalam bentuk sistem informasi berbasis web yang digunakan sebagai sarana penyajian informasi hasil klasifikasi, sehingga dapat dimanfaatkan sebagai media peringatan dini berdasarkan hasil prediksi model.

Berdasarkan rangkaian proses analisis dan evaluasi yang telah dilakukan, penelitian ini menunjukkan bahwa Random Forest memiliki tingkat efektivitas yang lebih tinggi dalam proses klasifikasi kualitas udara di wilayah Jakarta dibandingkan dengan algoritma Decision Tree.

Kata kunci: Klasifikasi, Kualitas Udara, ISPU, *Decision Tree*, *Random Forest*, Machine Learning.

ABSTRACT

Air quality is one of the indicators in determining the environmental condition of an area. Significant levels of air pollution in the DKI Jakarta area require a monitoring system that can present information in a responsive and accurate manner. In this study, the process of classification of air quality in the Jakarta area was carried out based on the Air Pollution Standard Index (ISPU) by utilizing two Machine Learning algorithms, namely Decision Tree and Random Forest. The data source in this study comes from the ISPU Jakarta dataset accessed through the DKI Jakarta Open Data portal, with five pollutant parameters used as measurement variables, namely PM10, SO₂, CO, O₃, and NO₂.

In its implementation, this study uses the CRISP-DM (Cross Industry Standard Process for Data Mining) work tool which includes five main stages in the data analysis process, namely data understanding, data cleaning, data preparation, modeling, and evaluation. The dataset used in this study was then separated into two parts, namely 80% as training data and 20% as testing data. To assess the performance of the resulting classification model, the evaluation process is carried out using a confusion matrix and several performance measurement metrics, namely accuracy, precision, recall, and F1-score.

The test results showed that the Decision Tree algorithm achieved an accuracy rate of 80.75%, while the Random Forest algorithm achieved an accuracy of 85.17%. The comparison indicates that the Random Forest algorithm has a better classification performance than Decision Tree with a difference in accuracy value of 4.41%. The resulting classification results are then implemented in the form of a web-based information system that is used as a means of presenting classified information, so that it can be used as an early warning medium based on the model's prediction results.

Based on a series of analysis and evaluation processes that have been carried out, this study shows that Random Forest has a higher level of effectiveness in the air quality classification process in the Jakarta area compared to the Decision Tree algorithm.

Keywords: Classification, Air Quality, ISPU, Decision Tree, Random Forest, Machine Learning.

KATA PENGANTAR

Penulis mengucapkan syukur kepada Allah SWT atas izin, rahmat, dan nikmat-Nya sehingga saya sebagai penulis dapat menyelesaikan skripsi ini dengan waktu yang telah ditentukan. Skripsi berjudul **Analisis Komparatif Algoritma Random Forest dan Decision Tree dalam Klasifikasi Kualitas Udara di DKI Jakarta.**

Skripsi ini ditulis sebagai salah satu persyaratan akademik untuk memperoleh gelar Sarjana Komputer dari Fakultas Teknologi Komunikasi dan Informatika di Universitas Nasional. Penulis menerima dukungan dari berbagai sumber selama penelitian dan penulisan tesis. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya, terutama kepada pembimbing tesis, Ibu Sari Ningsih, S.Si., M.M. yang telah memberikan bimbingan, bantuan praktis, waktu, pengetahuan, motivasi, dan kesabaran selama proses penelitian dan penulisan. Penulis juga ingin mengucapkan terima kasih kepada individu-individu berikut:

1. Ayah tercinta saya Sukiyanto dan Ibu tercinta saya, Lasmini. Meskipun dengan segala keterbatasan, selalu mencurahkan kasih sayang, dukungan, dan doa terbaik mereka kepada saya.
2. Dr. El Amry Bermawi Putera, M.A., Rektor Universitas Nasional yang memberi penulis kesempatan untuk belajar di Program Sarjana Sistem Informasi.
3. Dr. Agung Triayudi, S.Kom., M.Kom., Dekan Fakultas Komunikasi dan Teknologi Informasi, dan Dr. Andrianingsih, S.Kom., M.M.S.I., Kepala Program Studi Sistem Informasi.
4. Seluruh dosen di Program Studi Sistem Informasi di Fakultas Komunikasi dan Teknologi Informasi, serta dosen dari Program Studi lain yang telah berbagai ilmu dengan saya selama studi saya.

5. Dan yang terakhir, saya mengucapkan terima kasih kepada diri saya yang telah berjuang keras hingga saat ini, semoga pencapaian ini membuka pintu kesempatan baru di kemudian hari.



DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN TUGAS SARJANA	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	6
2.1 Penelitian Terdahulu	6
2.2 Studi Literatur	8
2.3 Landasan Teori.....	9
2.3.1 Kualitas Udara	9
2.3.2 Indeks Standar Pencemaran Udara (ISPU / AQI).....	9
2.3.3 Data Mining	10
2.3.4 Machine Learning	10
2.3.5 Algoritma Decision Tree.....	10
2.3.6 Algoritma Random Forest.....	11
2.3.7 Evaluasi Model	11
BAB III METODE PENELITIAN.....	13
3.1 Jenis dan Metode Penelitian.....	13
3.2 Alur Penelitian	14
3.3 Sumber dan Jenis Data	15
3.4 Variabel Penelitian.....	16
3.5 Tahapan Pengolahan Data.....	16
3.5.1 Data Understanding	16
3.5.2 Data Cleaning	16
3.5.3 Data Transformation.....	17
3.5.4 Labeling Kategori ISPU	17
3.5.5 Data Selection	17
3.5.6 Data Splitting.....	17
3.6 Metode Klasifikasi	18
3.6.1 Algoritma Decision Tree	18
3.6.2 Algoritma.....	19
3.7 Metode Pengujian Kinerja Model	19
3.7.1 Confusion Matrix	19

3.7.2 Accuracy.....	20
3.7.3 Precision, Recall, F1-Score	21
3.7.4 Hasil Akhir	22
3.8 Perhitungan Manual Algoritma.....	22
3.8.1 Data Sampel Perhitungan Manual	22
3.8.2 Perhitungan Manual Algoritma Decision Tree.....	23
3.8.3 Perhitungan Manual Algoritma Random Forest.....	25
3.9 Analisis Perbandingan	29
3.9.1 Perbandingan Nilai Akurasi.....	29
3.9.2 Perbandingan Precision & Recall	29
3.9.3 Analisis Confusion Matrix.....	30
BAB IV HASIL DAN PEMBAHASAN	32
4.1 Hasil Pengumpulan Data.....	32
4.2 Hasil Preprocessing Data	32
4.2.1 Data Cleaning	33
4.2.2 Data Transformation.....	33
4.2.3 Data Selection	33
4.2.4 Data Splitting.....	34
4.3 Hasil Data Preparation	34
4.4 Hasil Klasifikasi.....	34
4.4.1 Hasil Evaluasi Akurasi	34
4.4.2 Analisis Confusion Matrix Decision Tree	35
4.4.3 Analisis Confusion Matrix Random Forest	37
4.4.4 Analisis Precision, Recall, dan F1-Score.....	38
4.5 Perbandingan Model Decision Tree dan Random Forest	39
4.6 Tampilan Visualisasi.....	40
4.7 Pembahasan.....	41
4.7.1 Analisis Kinerja Model Secara Kuantitatif	41
4.7.2 Analisis Kesalahan Klasifikasi (Confusion Matrix).....	42
4.7.3 Analisis Perbandingan dengan Penelitian Terdahulu.....	42
4.7.2 Mengapa Random Forest Lebih Unggul di Kasus Jakarta	43
BAB V KESIMPULAN DAN SARAN.....	45
5.1 Kesimpulan.....	45
5.2 Saran.....	46
DAFTAR PUSTAKA.....	48
LAMPIRAN.....	50

DAFTAR GAMBAR

Gambar 3. 1	Gambar Tahapan Analisis Data.....	13
Gambar 3. 2	Flow Chart Alur Penelitian.....	15
Gambar 3. 8. 1	Data Sampel Perhitungan Manual	23
Gambar 4. 4. 1	Hasil Evaluasi Akurasi	35
Gambar 4. 4. 2	Analisis Confusion Matrix Decision Tree	35
Gambar 4. 4. 3	Analisis Confusion Matrix Random Forest	37
Gambar 4. 4. 4	Decision Tree	38
Gambar 4. 4. 5	Random Forest	39
Gambar 4. 5	Perbandingan Model Decision Tree dan Random Forest	39



DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	6
Tabel 3. 7. 1 Confusion Matrix	19
Tabel 3. 7. 4 Hasil Akhir	22
Tabel 3. 8. 3 Perhitungan Manual Algoritma Random Forest	25
Tabel 3. 8. 3. 1 Bootstrap Sampling.....	26
Tabel 3. 8. 3. 2 Data Random.....	26
Tabel 3. 8. 3. 3 Data Random.....	27
Tabel 3. 8. 3. 4 Split $PM_{10} \leq 50$	28
Tabel 3. 8. 3. 5 Split $PM_{10} \leq 50$	28
Tabel 3. 9. 2 Perbandingan Precision dan Recall	29

