

SKRIPSI

**IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK
KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA
(ISPU) DI PROVINSI DKI JAKARTA**



ELANG PUTRA KARUNIAWAN

1512622023

PROGRAM STUDI

PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER

FAKULTAS TEKNIK

UNIVERSITAS NEGERI JAKARTA

2026



KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
PERPUSTAKAAN DAN KEARSIPAN

Jalan Rawamangun Muka Jakarta 13220

Telepon/Faksimili: 021-4894221

Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Elang Putra Karuniawan
NIM : 1512622023
Fakultas/Prodi : Teknik / Pendidikan Teknik Informatika dan Komputer
Alamat email : elangputrakaruniawan@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan dan Kearsipan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain (.....)

yang berjudul :

**IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK KLASIFIKASI INDEKS
STANDAR PENCEMARAN UDARA (ISPU) DI PROVINSI DKI JAKARTA**

Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan dan Kearsipan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

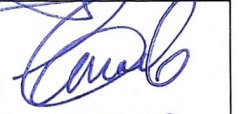
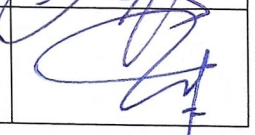
Jakarta, 5 Agustus 2026

(Elang Putra Karuniawan)

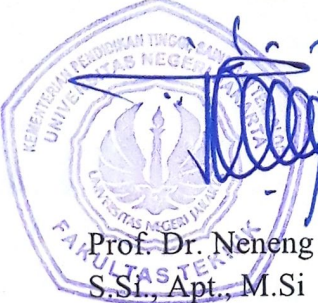
LEMBAR PENGESAHAN TUGAS AKHIR
BENTUK : SKRIPSI

Nama Mahasiswa : Elang Putra Karuniawan
NIM : 1512622023
Program Studi/Fakultas : Pendidikan Teknik Informatika dan Komputer / Teknik
Judul Penelitian : Implementasi Algoritma *Random Forest* untuk Klasifikasi Indeks Standar Pencemaran Udara (ISPU) di Provinsi DKI Jakarta

ini telah dipertahankan dihadapan dewan penguji dan dinyatakan **Lulus** / ~~Tidak Lulus~~ * Ujian Tugas Akhir oleh Tim Penguji pada Tanggal 13 Juli 2026
Tim Penguji,

Ketua Penguji	Ressy Dwitias Sari, S.T., M.T.I. NIDN. 0115098902	
Anggota Penguji (Penguji Ahli 1)	Fitrah Izul Falaq, M.Pd NIDN. 0729019801	
Anggota Penguji (Penguji Ahli 2)	Bambang Satrio, S. Pd., M. Pd. NUPTK. 474777567613022	
Anggota Penguji (Pembimbing 1)	Dr. Widodo, S. Kom., M. Kom. NIDN. 0025037206	
Anggota Penguji (Pembimbing 2)	Neng Ayu Herawati, S. Pd., M. T. NUPTK. 125277667723015	

Mengetahui,
Dekan Fakultas Teknik



Prof. Dr. Neneng Siti Silfi Ambarwati,
S.Si., Apt., M.Si
NIDN. 0029027204

Koordinator Program Studi
Pendidikan Teknik Informatika
dan Komputer



Muchammad Ficky Duskarnaen,
M.Sc.
NIDN. 0024097304

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Skripsi ini merupakan Karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 30 Juni 2026

Yang membuat pernyataan



Elang Putra Karuniawan

NIM. 1512622023

KATA PENGANTAR

Dengan menyebut nama Allah yang maha pengasih lagi maha penyayang. Segala puji bagi Allah Rabb semesta alam yang telah memberikan limpahan Rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Implementasi Algoritma Random Forest untuk Klasifikasi Indeks Standar Pencemaran Udara (ISPU) di Provinsi DKI Jakarta” dengan baik. Shalawat beserta salam tak lupa tucurahkan kepada Nabi Muhammad SAW yang menjadi suri tauladan bagi seluruh umat manusia.

Penyusunan skripsi ini dilakukan sebagai salah satu syarat akademis untuk memperoleh gelar sarjana pada Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta. Penulis menyadari bahwa dalam proses penyusunan skripsi ini tidak terlepas dari doa, bimbingan, dukungan, dan motivasi dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ayah Iwan Heri Purwanto dan Ibu Misyanti, selaku kedua orang tua yang tiada henti memberikan doa, limpahan kasih sayang, serta dukungan yang luar biasa kepada penulis.
2. Bayu Ardiawan, selaku saudara penulis, yang turut berjuang memberikan banyak bantuan kepada penulis selaku adiknya agar bisa meraih gelar sarjana.
3. Bapak Muhammad Ficky Duskarnaen, ST, M.Sc., selaku Koordinator Program Studi Pendidikan Teknik Informatika dan Komputer serta Dosen Pembimbing Akademik. Terima kasih atas bimbingan, arahan, serta dukungannya sehingga proses penyusunan skripsi ini dapat berjalan dengan lancar.
4. Bapak Dr. Widodo, M.Kom., selaku Dosen Pembimbing I yang telah meluangkan waktu dan tenaga, dan pikirannya dalam memberikan bimbingan, arahan, dan masukan yang berharga kepada penulis selama proses penyusunan skripsi.

5. Ibu Neng Ayu Herawati, S.Pd., M.T., selaku Dosen Pembimbing II yang turut membimbing serta memberikan arahan dan masukan yang sangat membantu penulis dalam menyelesaikan setiap tahapan penelitian ini.
6. Seluruh jajaran Dosen dan Staf Program Studi Pendidikan Teknik Informatika dan Komputer, atas ilmu pengetahuan dan bantuan yang telah diberikan selama penulis menempuh masa perkuliahan.
7. Teman-teman seperjuangan Program Studi Pendidikan Teknik Informatika dan Komputer angkatan 2022, yang telah memberikan semangat, bantuan, dan kebersamaan selama masa perkuliahan.
8. Serta seluruh pihak yang telah membantu dalam penyusunan skripsi ini, baik secara langsung maupun tidak langsung, yang tidak dapat penulis sebutkan satu per satu, tanpa mengurangi rasa hormat.

Penulis menyadari bahwa penyusunan skripsi ini masih jauh dari kata sempurna dan memiliki banyak kekurangan. Oleh karena itu, segala bentuk kritik dan saran yang membangun sangat diharapkan oleh penulis demi penyempurnaan di masa mendatang. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang *machine learning*.

Jakarta, 30 Juni 2026

Penulis,



Elang Putra Karuniawan

NIM. 1512622023

Intelligentia - Dignitas

ABSTRAK

Elang Putra Karuniawan, Implementasi Algoritma *Random Forest* untuk Klasifikasi Indeks Standar Pencemaran Udara (ISPU) di Provinsi DKI Jakarta. Dosen Pembimbing: Dr. Widodo, M.Kom. dan Neng Ayu Herawati, S.Pd., M.T. Program Studi Pendidikan Teknik Informatika dan Komputer. Fakultas Teknik. Universitas Negeri Jakarta. 2026.

Pencemaran udara di wilayah perkotaan, khususnya Provinsi DKI Jakarta, memerlukan sistem klasifikasi kualitas udara yang mampu menganalisis hubungan kompleks antar parameter polutan secara lebih adaptif dibandingkan pendekatan konvensional. Penelitian ini bertujuan membangun model klasifikasi kualitas udara berbasis *machine learning* menggunakan algoritma *Random Forest*. Dataset yang digunakan adalah data historis Indeks Standar Pencemaran Udara (ISPU) DKI Jakarta periode Januari 2024–November 2025. Tahapan penelitian meliputi data *cleaning*, data *transformation*, pembagian data latih dan data uji, optimasi *hyperparameter* menggunakan *GridSearchCV*, serta pemodelan *Random Forest* melalui dua skenario, yaitu *baseline* dan optimasi menggunakan SMOTE. Evaluasi model dilakukan menggunakan *Confusion Matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengujian menunjukkan kedua skenario menghasilkan nilai akurasi global yang stabil sebesar 0.9871 (99%). Namun, penerapan SMOTE berhasil meningkatkan sensitivitas model pada kelas minoritas, ditunjukkan dengan melonjaknya nilai *recall* kategori "Baik" dari 0.91 menjadi 0.94, serta memperkecil celah selisih (*gap accuracy*) latih-uji menjadi sangat ideal di angka 0.49%. Selain itu, analisis *Feature Importance* mencatatkan parameter partikulat halus PM2.5 sebagai fitur paling dominan dengan bobot kontribusi mencapai 67.2%. Dapat disimpulkan bahwa kombinasi *Random Forest* mampu menghasilkan model klasifikasi ISPU yang akurat, stabil, dan memiliki kemampuan generalisasi yang baik, sedangkan penerapan SMOTE efektif dalam meningkatkan keseimbangan kemampuan klasifikasi terhadap kelas minoritas tanpa mengurangi performa model secara keseluruhan.

Kata Kunci : ISPU, *Machine Learning*, Klasifikasi, *Random Forest*, SMOTE.

ABSTRACT

Elang Putra Karuniawan, *Implementation of the Random Forest Algorithm for Classification of the Air Pollution Standard Index (ISPU) in DKI Jakarta Province*. Supervisors: Dr. Widodo, M.Kom. and Neng Ayu Herawati, S.Pd., M.T. Study Program of Informatics and Computer Engineering Education, Faculty of Engineering, State University of Jakarta. 2026.

Air pollution in urban areas, particularly in the Special Capital Region of Jakarta, requires an air quality classification system capable of analyzing the complex relationships among pollutant parameters in a more adaptive manner than conventional approaches. This study aims to develop a machine learning-based air quality classification model using the Random Forest algorithm. The dataset used consists of historical Air Pollution Standard Index (ISPU) data for DKI Jakarta from January 2024 to November 2025. The research stages included data cleaning, data transformation, splitting the data into training and test sets, hyperparameter optimization using GridSearchCV, and Random Forest modeling through two scenarios: a baseline and an optimization using SMOTE. Model evaluation was performed using a Confusion Matrix with the metrics accuracy, precision, recall, and F1-score. The test results show that both scenarios produced a stable global accuracy of 0.9871 (99%). However, the application of SMOTE successfully improved the model's sensitivity for the minority class, as evidenced by a surge in the recall value for the "Good" category from 0.91 to 0.94, as well as a reduction in the training-test accuracy gap to an ideal 0.49%. Furthermore, the Feature Importance analysis identified the PM_{2.5} fine particulate matter parameter as the most dominant feature, with a contribution weight of 67.2%. It can be concluded that the Random Forest combination is capable of producing an accurate, stable, and well-generalizing ISPU classification model, while the application of SMOTE is effective in improving the balance of classification performance for the minority class without compromising the model's overall performance.

Keywords: ISPU, Machine Learning, Classification, Random Forest, SMOTE.

DAFTAR ISI

LEMBAR PENGESAHAN TUGAS AKHIR	ii
LEMBAR PERNYATAAN	iii
KATA PENGANTAR.....	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah.....	4
1.3. Batasan Masalah	4
1.4. Rumusan Masalah.....	5
1.5. Tujuan Penelitian	5
1.6. Manfaat Penelitian	5
1.6.1. Manfaat Teoritis	6
1.6.2. Manfaat Praktis	6
BAB II KAJIAN PUSTAKA	7
2.1. Landasan Teori.....	7
2.1.1. Indeks Standar Pencemaran Udara (ISPU)	7
2.1.2. <i>Machine Learning</i>	9
2.1.3. Klasifikasi	11
2.1.4. <i>Ensemble Learning</i>	11
2.1.5. <i>Random Forest</i>	12
2.1.6. <i>Preprocessing</i>	16
2.1.7. <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	18
2.1.8. <i>Feature Importance</i>	19
2.1.9. Evaluasi Model.....	20
2.2. Penelitian Terdahulu	22
2.3. Kerangka Berpikir.....	29

BAB III METODOLOGI PENELITIAN	32
3.1. Tempat dan Waktu Penelitian.....	32
3.2. Alat dan Bahan Penelitian.....	32
3.2.1. Alat Penelitian.....	32
3.2.2. Bahan Penelitian.....	33
3.3. Diagram Alir Penelitian	34
3.4. Prosedur Penelitian	35
3.4.1. Pengumpulan Data Historis ISPU	35
3.4.2. <i>Data Cleaning</i>	36
3.4.3. <i>Data Transformation</i>	37
3.4.4. <i>Modelling</i>	38
3.4.5. Analisis dan Evaluasi	41
3.5. Teknik Pengumpulan Data	43
3.6. Teknik Analisis Data	43
BAB IV PEMBAHASAN.....	45
4.1. Hasil Penelitian	45
4.1.1. Pengumpulan Data Historis ISPU	45
4.1.2. <i>Data Cleaning</i>	47
4.1.2.1. <i>Feature Selection</i>	47
4.1.2.2. <i>Handling Missing Values</i>	48
4.1.3. <i>Data Transformation</i>	49
4.1.3.1. <i>Feature Scaling</i>	49
4.1.3.2. <i>Label Encoding</i>	51
4.1.4. <i>Modelling</i>	52
4.1.4.1. <i>Data Splitting</i>	52
4.1.4.2. Pelatihan dan Pengujian Skenario <i>Baseline</i>	53
4.1.4.3. Pelatihan dan Pengujian Skenario Optimasi (<i>SMOTE</i>)... 55	
4.2. Analisis Hasil Penelitian	57
4.2.1. Analisis <i>Feature Importance</i>	58
4.2.2. Evaluasi Skenario <i>Baseline</i>	59
4.2.3. Evaluasi Skenario Optimasi (<i>SMOTE</i>).....	62
4.2.4. Perbandingan Kinerja Pada Kedua Skenario	64

4.3. Pembahasan.....	67
4.3.1. Karakteristik dan Distribusi Data ISPU	67
4.3.2. Analisis Kinerja <i>Random Forest</i> dalam Klasifikasi ISPU	68
4.3.3. Pengaruh Penerapan <i>SMOTE</i> Terhadap Kinerja Model.....	69
4.3.4. Analisis Parameter Polutan Berdasarkan <i>Feature Importance</i>	71
4.3.5. Perbandingan Hasil Penelitian dengan Penelitian Terdahulu.....	72
4.3.6. Keterbatasan Penelitian.....	74
BAB V PENUTUP.....	76
5.1. Kesimpulan	76
5.2. Saran	76
DAFTAR PUSTAKA.....	78
LAMPIRAN.....	82



Intelligentia - Dignitas

DAFTAR TABEL

Tabel 2.1 Konversi Nilai Konsentrasi Parameter ISPU	9
Tabel 2.2 <i>Confusion Matrix</i>	20
Tabel 2.3 Tabel Ringkasan Penelitian Terdahulu	25
Tabel 3.1 Spesifikasi Perangkat Penelitian	32
Tabel 3.2 Sebaran Jumlah Sampel Kategori ISPU	33
Tabel 3.3 Pemetaan Label Kategori ISPU.....	38
Tabel 3.4 Skenario Uji Coba <i>Hyperparameter Tuning</i>	40
Tabel 4.1 Struktur Atribut Data Historis ISPU.....	46
Tabel 4.2 Jumlah <i>Missing Values</i> pada Dataset.....	48
Tabel 4.3 Hasil <i>Label Encoding</i>	52
Tabel 4.4 Hasil <i>Data Splitting</i>	53
Tabel 4.5 Hasil <i>Hyperparameter</i> Terbaik (Skenario <i>Baseline</i>).....	54
Tabel 4.6 Kinerja <i>Cross-Validation</i> Parameter Terbaik Skenario <i>Baseline</i>	55
Tabel 4.7 Hasil <i>Hyperparameter</i> Terbaik (Skenario Optimasi)	56
Tabel 4.8 Kinerja <i>Cross-Validation</i> Parameter Terbaik Skenario Optimasi.....	57
Tabel 4.9 <i>Classification Report</i> Skenario <i>Baseline</i>	61
Tabel 4.10 <i>Classification Report</i> Skenario Optimasi.....	63
Tabel 4.11 Ringkasan Komparatif Metrik Evaluasi Kedua Skenario	66
Tabel 4.12 Ringkasan Komparasi dengan Penelitian Terdahulu.....	72

Intelligentia - Dignitas

DAFTAR GAMBAR

Gambar 2.1 Diagram Arsitektur Random Forest (Chen et al., 2019)	13
Gambar 2.2 Distribusi Sampel Sintentis Teknik SMOTE (Pertiwi et al., 2020)... 18	
Gambar 2.3 Kerangka Berpikir	31
Gambar 3.1 Diagram Alir Penelitian.....	34
Gambar 4.1 Dataset ISPU DKI Jakarta.....	46
Gambar 4.2 Dataset Hasil Seleksi Fitur	47
Gambar 4.3 Distribusi Nilai Konsentrasi Polutan Sebelum <i>Scaling</i>	50
Gambar 4.4 Distribusi Nilai Konsentrasi Polutan Setelah <i>Scaling</i>	51
Gambar 4.5 Rekayasa Fitur Terhadap Kategori Sangat Tidak Sehat	51
Gambar 4.6 Kode Program Model <i>Baseline</i>	54
Gambar 4.7 Distribusi Data Latih Sebelum dan Sesudah SMOTE	56
Gambar 4.8 <i>Feature Importance</i> Parameter Polutan Berdasarkan Skor MDI	58
Gambar 4.9 <i>Heatmap</i> Korelasi Antar Parameter Polutan	59
Gambar 4.10 <i>Confusion Matrix</i> Skenario <i>Baseline</i>	60
Gambar 4.11 <i>Confusion Matrix</i> Skenario Optimasi.....	62
Gambar 4.12 <i>Perbandingan Akurasi Model Baseline vs. Optimasi SMOTE</i>	65

Intelligentia - Dignitas

DAFTAR LAMPIRAN

Lampiran 1. Lembar Pernyataan Kelayakan Judul/Tema Skripsi	82
Lampiran 2. Surat Tugas	83
Lampiran 3. Lembar Konsultasi Proposal Skripsi	84
Lampiran 4. Kartu Kehadiran Seminar Proposal	88
Lampiran 5. Biodata Peserta Ujian Skripsi	90
Lampiran 6. Surat Permohonan Ujian Skripsi	94
Lampiran 7. Surat Persetujuan Dosen Pembimbing Skripsi	95
Lampiran 8. Surat Persetujuan Dosen Pembimbing Akademik	97
Lampiran 9. Pra-Transkrip Akademik.....	98
Lampiran 10. Surat Pernyataan Menyelesaikan Mata Kuliah Wajib	101
Lampiran 11. Dataset Penelitian	102
Lampiran 12. Dataset Setelah <i>Preprocessing</i>	103
Lampiran 13. Dataset Setelah <i>SMOTE</i>	104
Lampiran 14. Kode Program.....	105

Intelligentia - Dignitas