



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



**RANCANG BANGUN APLIKASI KLASIFIKASI
PENCEMARAN UDARA MENGGUNAKAN
ALGORITMA *XGBOOST* BERBASIS WEB**

SKRIPSI

LANANG PRASETYO 2007411042

**POLITEKNIK
NEGERI
JAKARTA**

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN
KOMPUTER**

POLITEKNIK NEGERI JAKARTA

2026



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



**RANCANG BANGUN APLIKASI KLASIFIKASI
PENCEMARAN UDARA MENGGUNAKAN
ALGORITMA *XGBOOST* BERBASIS WEB**

SKRIPSI

**Dibuat untuk Melengkapi Syarat-syarat yang Diperlukan untuk Memperoleh
Diploma Empat Politeknik**

**POLITEKNIK
NEGERI
JAKARTA**

20074110042

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN
KOMPUTER**

POLITEKNIK NEGERI JAKARTA

2026



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

SURAT PERNYATAAN BEBAS PLAGIARISME

Saya yang bertanda tangan di bawah ini:

Nama : Lanang Prasetyo
NIM : 2007411042
Jurusan/Program Studi : T. Informatika dan Komputer / Teknik Informatika
Judul skripsi : RANCANG BANGUN APLIKASI KLASIFIKASI
PENCEMARAN UDARA MENGGUNAKAN
ALGORITMA XGBOOST BERBASIS WEB

Menyatakan dengan sebenarnya bahwa skripsi ini benar-benar merupakan hasil karya saya sendiri, bebas dari peniruan terhadap karya dari orang lain. Kutipan pendapat dan tulisan orang lain ditunjuk sesuai dengan cara-cara penulisan karya ilmiah yang berlaku.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa dalam skripsi ini terkandung cirri-ciri plagiat dan bentuk-bentuk peniruan lain yang dianggap melanggar peraturan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Bekasi, 26 Mei 2026

Yang membuat pernyataan,



Lanang Prasetyo

NIM 2007411042



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

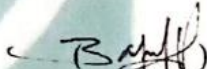
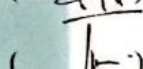
LEMBAR PENGESAHAN

Saya yang bertanda tangan di bawah ini :

Nama : Lanang Prasetyo
NIM : 2007411042
Program Studi : Teknik Informatika
Judul Skripsi : RANCANG BANGUN APLIKASI KLASIFIKASI
PENCEMARAN UDARA MENGGUNAKAN
ALGORITMA XGBOOST BERBASIS WEB

Telah diuji oleh tim penguji dalam Sidan Skripsi pada hari Rabu Tanggal 17, Bulan Juni, Tahun 2026, dan dinyatakan LULUS.

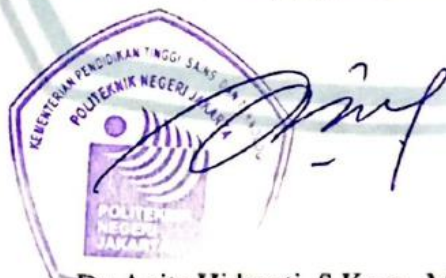
Disahkan oleh

Pembimbing I : Anggi Mardiyono, S.Kom., M.Kom. ()
Penguji I : Rizki Elisa Nalawati, S.T., M.T. ()
Penguji II : Bambang Warsuta, S.Kom., M.T.I. ()
Penguji III : Iik Muhamad Malik Matin, S.Kom., M.T. ()

Mengetahui :

Jurusan Teknik Informatika dan Komputer

Ketua



Dr. Anita Hidayati, S.Kom., M.Kom.

NIP. 197908032003122003



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

KATA PENGANTAR

Syukur Alhamdulillah, penulis panjatkan atas kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Rancang Bangun Aplikasi Klasifikasi Pencemaran Udara Menggunakan Algoritma XGBoost Berbasis Web” sebagai salah satu persyaratan untuk mencapai gelar Sarjana Terapan Program Studi Teknik Informatika di Politeknik Negeri Jakarta.

Penulis menyadari bahwa telah banyak pihak yang memberikan dukungan dan bantuan dalam proses penyelesaian studi skripsi ini. Dengan demikian, sudah sepantasnya penulis mengungkapkan rasa terimakasih yang tulus dan berdoa agar Allah memberikan balasan yang terbaik. Secara khusus penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Ibu Dr. Anita Hidayati, S.Kom, M.Kom selaku Ketua Jurusan Teknik Informatika dan Komputer.
2. Ibu Euis Oktavianti, S.Si., M.TI., selaku Kepala Program Studi Jurusan Teknik Informatika.
3. Pak Anggi Mardiyono, S.Kom., M.Kom. selaku dosen pembimbing yang telah menyediakan waktu dan memberikan bantuan yang sangat berarti dalam penyusunan skripsi ini.
4. Kedua orang tua tercinta yang selalu mendukung, mendoakan, memberi semangat, serta kasih sayang yang tiada habisnya kepada penulis.
5. Adik-adik yang tersayang yang telah sepenuhnya memberikan dukungan dan doanya.
6. Sahabat penulis yang selalu memberi semangat.



© Hak Cipta milik Politeknik Negeri Jakarta

Akhir kata, semoga skripsi penulis ini dapat bermanfaat bagi semua pihak, baik dari segi pendidikan maupun penelitian. Penulis sadar bahwasanya skripsi ini masih jauh dari kata sempurna, penulis mengucapkan permohonan maaf atas kekurangan dan keterbatasan dalam skripsi ini. Oleh karena itu, penulis mengharapkan segala bentuk kritik, saran, dan masukan yang dapat membantu memperbaiki serta menyempurnakan skripsi ini.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Politeknik Negeri Jakarta, saya bertanda tangan di bawah ini :

Nama : Lanang Prasetyo

NIM : 2007411042

Jurusan/Program Studi : T.Informatika dan Komputer / Teknik Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Politeknik Negeri Jakarta Hak Bebas Royalti Non-Eksklusif atas karya ilmiah saya yang berjudul :

**RANCANG BANGUN APLIKASI KLASIFIKASI PENCEMARAN UDARA
DENGAN MENGGUNAKAN ALGORITMA XGBOOST BERBASIS WEB**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksklusif ini Politeknik Negeri Jakarta berhak menyimpan, mengalihmediakan/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.. Demikian pernyataan ini saya buat dengan sebenarnya.

Bekasi, 17 Juli 2026

Yang Membuat Pernyataan,



Lanang Prasetyo

NIM 2007411042



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

RANCANG BANGUN APLIKASI KLASIFIKASI PENCEMARAN UDARA MENGUNAKAN ALGORITMA XGBOOST BERBASIS WEB

ABSTRAK

DKI Jakarta merupakan salah satu wilayah dengan tingkat pencemaran udara yang menjadi perhatian karena tingginya aktivitas transportasi, industri, dan kepadatan penduduk yang berkontribusi terhadap peningkatan konsentrasi polutan udara. Penelitian ini bertujuan mengembangkan aplikasi klasifikasi pencemaran udara berbasis web yang mengintegrasikan model machine learning sehingga pengguna dapat melakukan klasifikasi data Indeks Standar Pencemar Udara (ISPU) secara otomatis. Aplikasi dikembangkan menggunakan metode Waterfall dengan framework Laravel dan bahasa pemrograman PHP, sedangkan pengembangan model machine learning mengacu pada tahapan CRISP-DM. Dataset yang digunakan merupakan data ISPU DKI Jakarta periode 2010–2024 yang bersumber dari Portal Satu Data Jakarta. Model klasifikasi dibangun menggunakan algoritma XGBoost Classifier dengan penerapan Stratified Train-Test Split, Hyperparameter Tuning menggunakan RandomizedSearchCV, Stratified K-Fold Cross Validation, serta perbandingan tiga skenario penanganan ketidakseimbangan data, yaitu tanpa penanganan, SMOTE, dan Class Weight. Hasil pengujian menunjukkan bahwa penerapan Class Weight dengan rasio split data 80:20 menghasilkan performa terbaik dengan nilai accuracy sebesar 99,89%. Model yang dihasilkan kemudian berhasil diintegrasikan ke dalam sistem aplikasi melalui modul Python sehingga sistem mampu melakukan klasifikasi data ISPU secara konsisten sesuai dengan model machine learning.

Kata kunci : Klasifikasi Pencemaran Udara, Class Weight, Laravel, Machine Learning, XGBoost.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

DAFTAR ISI

SURAT PERNYATAAN BEBAS PLAGIARISME	iii
LEMBAR PENGESAHAN	i
KATA PENGANTAR	vi
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	ii
ABSTRAK	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat	3
1.5 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu.....	5
2.2 Website	9
2.3 CRISP-DM.....	10
2.4 Klasifikasi	12
2.5 Class Weight.....	12
2.6 Stratified K-Fold Cross Validation.....	13
2.7 Randomized Search Cross Validation	13
2.8 XGBoost.....	14



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

2.9 Waterfall	20
2.10 Python	23
2.11 Laravel.....	23
2.12 UML	24
2.13 Use Case Diagram.....	25
2.14 ISPU	26
BAB III METODE PENELITIAN	27
3.1 Rancangan Penelitian.....	27
3.2 Tahap Penelitian	28
3.3 Tahapan Pembangunan Model Machine Learning.....	28
3.4 Tahap Pembuatan Aplikasi Web.....	33
3.5 Objek Penelitian	33
BAB IV HASIL DAN PEMBAHASAN	33
4.1 Analisa Kebutuhan	33
4.1.1 Kebutuhan Pengembangan Model	33
4.1.2 Kebutuhan Pengembangan Aplikasi Web	35
4.1.3 Kebutuhan Perangkat Keras & Perangkat Lunak.....	40
4.2 Perancangan Sistem	42
4.2.1 Perancangan Model.....	42
4.3 Implementasi Sistem	51
4.3.1 Implementasi Model <i>Machine Learning</i>	51
4.3.2 Implementasi Web.....	70
4.4 Pengujian.....	72
4.4.1 Deskripsi Pengujian	72
4.4.2 Prosedur Pengujian	81



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

4.4.3 Data Hasil Pengujian.....	83
4.4.4 Evaluasi Pengujian	82
BAB V PENUTUP.....	85
5.1 Kesimpulan	85
5.2 Saran.....	91
DAFTAR PUTAKA.....	92
DAFTAR RIWAYAT HIDUP PENULIS	93
LAMPIRAN	94





Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

DAFTAR GAMBAR

Gambar 2. 1 Tahapan dalam CRISP-DM.....	10
Gambar 2. 2 Alur Stratified K-Fold Cross Validation.....	13
Gambar 2. 3 Cara Kerja RandomizedSearchCV.....	14
Gambar 2. 4 Algoritma XGBoost	15
Gambar 2. 5 Alur Metode Waterfall.....	21
Gambar 2. 6 Contoh Diagram Use Case	25
Gambar 3. 1 Flowchart Diagram Model Machine Learning.....	29
Gambar 3. 2 Flowchart Diagram Pembuatan Aplikasi Web	33
Gambar 4. 1 Activity Diagram Proses Klasifikasi	42
Gambar 4. 2 Activity Diagram Mengunduh Hasil Klasifikasi.....	43
Gambar 4. 3 Activity Diagram Unduh Hasil Evaluasi.....	44
Gambar 4. 4 Wireframe Halaman Upload.....	45
Gambar 4. 5 Wireframe Halaman Hasil Klasifikasi.....	50
Gambar 4. 6 Distribusi Data Pencemar.....	52
Gambar 4. 7 Informasi Umum Dataset	53
Gambar 4. 8 Informasi Data Non Null.....	50
Gambar 4. 9 Total Missing Value.....	50
Gambar 4. 10 Informasi Distribusi Kelas	51
Gambar 4. 11 Deteksi Outlier	52
Gambar 4. 12 Boxplot Outlier.....	53
Gambar 4. 13 Tahap Pembersihan Data	53
Gambar 4. 14 Tahap Pembersihan Nilai Negatif.....	54
Gambar 4. 15 Tahap Encoding Label.....	55
Gambar 4. 16 Analisis Temporal PM ₁₀	55
Gambar 4. 17 Analisis Temporal PM ₂₅	60
Gambar 4. 18 Analisis Temporal SO ₂	60
Gambar 4. 19 Analisis Temporal NO ₂	60
Gambar 4. 20 Analisis Temporal SO ₂	60



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Gambar 4. 21 Analisis Temporal O ₃	61
Gambar 4. 22 Analisis Korelasi Antar Polutan	61
Gambar 4. 23 Tahap Data Splitting.....	63
Gambar 4. 24 Tanpa Penerapan Metode Imbalance Data Handling	60
Gambar 4. 25 Penerapan Teknik Class Weight	60
Gambar 4. 26 Penerapan Teknik SMOTE.....	63
Gambar 4. 27 Tahap Hyperparameter Tuning.....	63
Gambar 4. 28 Stratified K-Fold Cross Validation.....	64
Gambar 4. 29 RandomizedSearchCV	65
Gambar 4. 30 Tahap Evaluasi	70
Gambar 4. 31 Diagram Confusion Matrix	72
Gambar 4. 32 Tahap Evaluasi	73
Gambar 4. 33 Tampilan Halaman Upload.....	70
Gambar 4. 34 Tampilan Halaman Hasil Klasifikasi.....	71
Gambar 4. 35 Tampilan Diagram Hasil Klasifikasi	71



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu.....	5
Tabel 4. 1 Tabel Kebutuhan Dataset.....	33
Tabel 4. 2 Tabel Kebutuhan Model	34
Tabel 4. 3 Tabel Kebutuhan Fungsional.....	35
Tabel 4. 4 Tabel Kebutuhan Non Fungsional.....	40
Tabel 4. 5 Tabel Kebutuhan Perangkat Keras	41
Tabel 4. 6 Tabel Kebutuhan Perangkat Lunak	41
Tabel 4. 7 Distribusi Kelas	52
Tabel 4. 8 Tabel Distribusi Kelas Data.....	61
Tabel 4. 9 Ringkasan Metrik Model.....	71
Tabel 4. 10 Hasil Classsification Report	71
Tabel 4. 11 Rancangan Black Box Testing 1.....	73
Tabel 4. 12 Rancangan Black Box Testing 2	74
Tabel 4. 13 Tabel Model Performance Testing 1.....	80
Tabel 4. 14 Tabel Uji Model Performance Testing 2.....	80
Tabel 4. 15 Rancangan Model Performance Testing 3.....	81
Tabel 4. 16 Hasil Black Box Testing 1	83
Tabel 4. 17 Hasil Black Box Testing 2.....	80
Tabel 4. 18 Hasil Pengujian Variasi Persentase Split Data	81
Tabel 4. 19 Hasil Performance Testing	81



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pencemaran udara merupakan salah satu permasalahan lingkungan yang serius. Seiring dengan meningkatnya aktivitas manusia, seperti aktivitas industri, penggunaan kendaraan berbahan bakar minyak bumi, pembakaran lahan, hingga pertumbuhan populasi di kawasan perkotaan, mempunyai andil yang sangat besar dalam memperburuk tingkat pencemaran udara (Maharani & Aryanta, 2023). Kondisi ini berdampak langsung pada kesehatan masyarakat. Paparan udara yang terkontaminasi polusi dapat menyebabkan berbagai penyakit, seperti penyakit jantung, stroke, penyakit paru-paru obstruktif kronis, kanker paru-paru, infeksi saluran pernapasan, bahkan menurunkan harapan hidup (Rahmawati, siti dan pratama, 2023). Wilayah perkotaan dengan tingkat pencemaran udara yang tinggi salah satunya adalah provinsi DKI Jakarta, yang memiliki kepadatan penduduk dan tingkat mobilitas masyarakat yang tinggi, dengan sekitar 71% penduduk berada pada usia produktif, sehingga mendorong peningkatan jumlah kendaraan bermotor yang menurut Badan Pusat Statistik didominasi oleh sepeda motor dan menjadi penyumbang utama polutan udara seperti NO_x, CO, dan HC (Tantarani et al., 2024)

Terdapat beberapa penelitian yang melakukan klasifikasi terhadap pencemaran udara menggunakan beberapa algoritma *machine learning*. Beberapa di antaranya menerapkan metode seperti *Random Forest*, *K-Nearest Neighbors*, *Support Vector Machine*, hingga *Artificial Neural Network* untuk mengelompokkan tingkat pencemaran udara di berbagai wilayah, termasuk Jakarta. Namun, penelitian-penelitian tersebut umumnya menggunakan dataset dengan rentang waktu terbatas, parameter yang beragam, penerapan teknik yang berbeda pada model yang dibangun, serta belum adanya integrasi antara model *machine learning* ke dalam aplikasi berbasis web.



Di sisi lain, penelitian terkait penerapan XGBoost menunjukkan bahwa algoritma ini mampu memberikan kinerja yang lebih unggul dibandingkan dengan metode lain pada berbagai studi klasifikasi, selain itu penerapan algoritma XGBoost pada data ISPU Jakarta dan mengintegrasikannya ke dalam sistem aplikasi berbasis web masih relatif jarang dilakukan. Kesenjangan ini menunjukkan perlunya penelitian lanjutan yang mengembangkan model klasifikasi pencemaran udara dengan algoritma yang lebih akurat dan dataset yang lebih kompleks rentang waktunya yang akan diintegrasikan ke dalam sistem aplikasi berbasis web.

Meskipun data Indeks Standar Pencemar Udara (ISPU) telah tersedia dan dapat dimanfaatkan sebagai dasar pemantauan kualitas udara, pemanfaatannya untuk proses klasifikasi secara otomatis masih memerlukan dukungan sistem yang mampu mengolah data tersebut menggunakan pendekatan *machine learning*. Berdasarkan permasalahan tersebut, diperlukan suatu pengembangan sistem yang mampu melakukan proses klasifikasi secara otomatis terhadap data ISPU yang diinputkan oleh pengguna. Oleh karena itu, penelitian ini bertujuan untuk merancang dan membangun sebuah aplikasi berbasis web menggunakan *framework Laravel* yang mengintegrasikan algoritma XGBoost untuk melakukan klasifikasi tingkat pencemaran udara dengan memanfaatkan dataset ISPU DKI Jakarta periode 2010 hingga 2024, sehingga diharapkan mampu menghasilkan model yang lebih akurat dan aplikatif.

1.2 Rumusan Masalah

Berdasarkan rumusan masalah yang telah dijabarkan, rumusan masalah utama adalah “Bagaimana merancang dan membangun aplikasi berbasis web yang mampu melakukan klasifikasi tingkat pencemaran udara di wilayah DKI Jakarta menggunakan algoritma XGBoost berdasarkan data ISPU yang diinputkan oleh pengguna ?”.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

1.3 Batasan Masalah

Berdasarkan rumusan masalah yang telah dirumuskan, berikut merupakan Batasan-batasan masalah yang digunakan dalam penelitian ini :

1. Aplikasi klasifikasi pencemaran udara yang dikembangkan berbasis web.
2. *Framework* yang digunakan untuk membuat aplikasi ini adalah Laravel.
3. Dataset yang digunakan adalah data ISPU wilayah DKI Jakarta periode 2010 – 2024.
4. Algoritma yang digunakan untuk klasifikasi adalah XGBoost.
5. Input data dilakukan melalui file CSV yang diunggah oleh pengguna.
6. Output sistem berupa hasil klasifikasi tingkat pencemaran udara.

1.4 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah Merancang dan membangun aplikasi berbasis web yang mampu melakukan klasifikasi tingkat pencemaran udara di wilayah DKI Jakarta menggunakan algoritma XGBoost berdasarkan data ISPU. Sedangkan manfaat teoritis yaitu memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang *machine learning*, khususnya dalam penerapan algoritma XGBoost untuk klasifikasi pencemaran udara, selain itu manfaat praktisnya adalah menyediakan aplikasi berbasis web yang dapat digunakan oleh para ahli untuk mengolah dan mengklasifikasikan data ISPU secara otomatis, sehingga membantu dalam memahami kondisi kualitas udara secara lebih cepat dan informatif.



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

1.5 Sistematika Penulisan

Dalam membuat laporan skripsi terdapat sistematika penulisan yang terdiri dari beberapa bagian, diantaranya adalah :

BAB I PENDAHULUAN

BAB I menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat, sistematika penulisan.

BAB II TINJAUAN PUSTAKA

BAB II berisi tentang penelitian-penelitian terdahulu serta teori-teori yang mendukung penelitian, perancangan dan pembuatan aplikasi.

BAB III METODE PENELITIAN

BAB III berisi uraian metode yang digunakan dalam penelitian, meliputi rancangan penelitian, tahapan penelitian, serta objek penelitian.

BAB IV HASIL DAN PEMBAHASAN

BAB IV berisi pembahasan terkait hasil penelitian mulai dari analisis kebutuhan, perancangan, implementasi, dan pengujian.

BAB V PENUTUP

BAB V menjelaskan simpulan hasil penelitian dan saran terkait penelitian yang telah dilakukan.

DAFTAR PUSTAKA

DAFTAR PUSTAKA melampirkan seluruh sumber ilmiah yang digunakan.

LAMPIRAN

LAMPIRAN melampirkan hal-hal lain yang mendukung penelitian.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini telah berhasil mengembangkan aplikasi klasifikasi pencemaran udara berbasis web dengan menggunakan algoritma *XGBoost Classifier* dengan memanfaatkan dataset Indeks Standar Pencemar Udara (ISPU). Penelitian dilaksanakan melalui tahapan pengumpulan dan prapemrosesan data, pembangunan model *machine learning*, optimasi model, evaluasi performa, hingga implementasi model ke dalam aplikasi berbasis web menggunakan *framework Laravel*

Pada tahap pembangunan model, dilakukan proses data *preprocessing* yang meliputi normalisasi label, penanganan nilai kosong menggunakan *median imputation*, penyesuaian atribut, serta analisis kualitas data. Selanjutnya dilakukan pengujian terhadap tiga variasi persentase pembagian data menggunakan metode *Stratified Train-Test Split*, yaitu 70:30, 80:20, dan 90:10. Berdasarkan hasil pengujian, pembagian data 80:20 memberikan performa terbaik karena mampu menghasilkan keseimbangan antara data latih dan data uji dengan tetap mempertahankan proporsi setiap kelas.

Penelitian ini juga melakukan perbandingan terhadap tiga metode penanganan ketidakseimbangan data (*imbalanced data handling*), yaitu tanpa penanganan *imbalance*, *Class Weight*, dan *Synthetic Minority Oversampling Technique* (SMOTE). Hasil pengujian menunjukkan bahwa metode *Class Weight* memberikan performa terbaik dibandingkan kedua metode lainnya. Model yang menggunakan *Class Weight* memperoleh nilai *Accuracy* sebesar 99,89%, *Precision (Macro)* sebesar 99,56%, *Recall (Macro)* sebesar 99,95%, *F1-Score (Macro)* sebesar 99,75%, *Log Loss* sebesar 0,0049, serta *ROC-AUC* sebesar 1,0000. Hasil tersebut menunjukkan bahwa pemberian bobot pada kelas minoritas mampu meningkatkan keseimbangan performa model tanpa mengubah distribusi data asli sehingga model mampu mengklasifikasikan seluruh kategori kualitas udara dengan sangat baik.



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Model klasifikasi yang telah dibangun kemudian berhasil diintegrasikan ke dalam aplikasi web berbasis Laravel. Aplikasi ini dilengkapi dengan beberapa fitur utama, yaitu:

- Fitur unggah data ISPU, yang memungkinkan pengguna mengunggah file CSV berisi data kualitas udara.
- Fitur klasifikasi kualitas udara, yang menampilkan hasil prediksi kategori kualitas udara berdasarkan data yang diunggah.
- Fitur unduh hasil klasifikasi, yang memungkinkan pengguna menyimpan hasil klasifikasi ke dalam format CSV.
- Fitur unduh hasil evaluasi, yang menyajikan ringkasan hasil evaluasi model.
- Fitur visualisasi data, yang menampilkan grafik hasil klasifikasi sehingga distribusi kategori kualitas udara dapat dipahami dengan lebih mudah.

Berdasarkan hasil penelitian tersebut dapat disimpulkan bahwa algoritma *XGBoost* yang dioptimasi menggunakan *Hyperparameter Tuning*, *Stratified Train-Test Split*, dan *Class Weight* mampu memberikan performa klasifikasi terbaik pada dataset ISPU yang digunakan dalam penelitian ini. Selain itu, aplikasi web yang dikembangkan berhasil mengimplementasikan model *machine learning* sehingga dapat dimanfaatkan sebagai media pendukung dalam proses klasifikasi dan pemantauan kualitas udara secara lebih efektif.

**POLITEKNIK
NEGERI
JAKARTA**



5.2 Saran

Berdasarkan hasil penelitian dan pengujian model klasifikasi pencemaran udara menggunakan algoritma XGBoost yang telah diintegrasikan ke dalam aplikasi berbasis web, berikut beberapa saran untuk pengembangan lebih lanjut :

1. Peningkatan Kualitas Model

Untuk meningkatkan kualitas prediksi, disarankan dilakukan optimasi model dengan teknik *hyperparameter tuning* yang lebih mendalam serta membandingkan performa XGBoost dengan algoritma lain seperti *Random Forest* dan model lainnya. Dengan pendekatan ini, diharapkan model dapat menghasilkan prediksi yang lebih akurat dan *robust* terhadap variasi data kualitas udara di berbagai kondisi.

2. Peningkatan Kualitas Dataset

Dataset ISPU yang digunakan dapat ditingkatkan dengan memperbanyak jumlah data, baik dari segi cakupan waktu, maupun cakupan geografis. Hal ini akan memungkinkan model untuk menangkap pola pencemaran udara yang lebih kompleks dan meningkatkan kemampuan generalisasi saat diterapkan pada data baru.

3. Pengayaan Fitur Aplikasi

Aplikasi web dapat dikembangkan lebih lanjut dengan menambahkan fitur notifikasi otomatis apabila kualitas udara berada pada kategori berbahaya, serta memperluas opsi ekspor hasil klasifikasi. Hal ini akan meningkatkan kegunaan aplikasi bagi pengguna akhir.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

DAFTAR PUSTAKA

- Abdulloh, R. (2022). *7 Materi Pemrograman Web untuk Pemula*. Elex Media Komputindo. <https://books.google.co.id/books?id=E7WAEAAAQBAJ>
- Ahmad, N., Krisnanik, E., Rupilele, F. G. J., Muliawati, A., Syamsiyah, N., Kraugusteeliana, K., Cahyono, B. D., Sriyeni, Y., Kristanto, T., Irwanto, I., & others. (2022). *ANALISA \& PERANCANGAN SISTEM INFORMASI BERORIENTASI OBJEK*. Penerbit Widina. <https://books.google.co.id/books?id=wSSFEAAAQBAJ>
- Anggrawan, A., Datya, A. I., Angellia, F., Kuntarto, G. P., Hidayat, S., Nurfatma, M., Umami, I., Miranda, E., Gunawan, I. P., & others. (2023). *KOMPUTER \& MASYARAKAT: Teori dan Penerapan Komputer Masyarakat Era Industri 4.0 dan Society 5.0*. PT. Sonpedia Publishing Indonesia. <https://books.google.co.id/books?id=Q6rLEAAAQBAJ>
- Arifa, R. A., Suarna, N., Bahtiar, A., Rahaningsih, N., & Prihartono, W. (n.d.). (*S I N T E K*) *MODEL MACHINE LEARNING UNTUK PREDIKSI RISIKO PENYAKIT*. VI(1), 39–45.
- Ayu, J. M., Dachi, S., & Sitompul, P. (2023). Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit. *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam (JURRIMIPA)*, 2(2), 87–103. <https://prin.or.id/index.php/JURRIMIPA/article/view/1470>
- Dr. Abdiansah, S. K. M. C. S. (2022). *PEMROGRAMAN DASAR PYTHON 3.0: TEORI \& IMPLEMENTASI*. Bening Media Publishing. <https://books.google.co.id/books?id=AkaEEAAAQBAJ>
- Febrianti, W. D., Kholifahtus, N., & Najah, J. (n.d.). *Analisis Klasifikasi Kepuasan Penumpang Maskapai Penerbangan Menggunakan Metode Support Vector Machine , Decision Tree , Dan Random Forest*. xx(x), 1–7. <https://doi.org/10.12962/j27213862.vxix.xxxx>



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Firdaus, D., Sumardi, I., & Chazar, C. (2025). *Deteksi Serangan Pada Jaringan Internet Of Things Medis Menggunakan Machine Learning Dengan Algoritma XGBoost Medical of Things Attack Detection On Internet Medical Of Things Using Machine Learning With Xgboost Algorithm*. 8(1), 34–42.

Giantara, W., Tugiyono, T., Setiawan, A., & Susanto, G. N. (2021). Air Quality and Noise Study in Settlements and Metal Scrap Melting Factory. *Jurnal Ilmiah Biologi Eksperimen Dan Keanekaragaman Hayati (J-BEKH)*, 8(2), 1–7. <https://doi.org/10.23960/jbekh.v8i1.159>

Illaihiyah, R., & Abdurrahman, G. (2026). *Implementation of the XGBoost Classifier Algorithm for Classifying Crowds of Hajj and Umrah Pilgrims Implementasi Algoritma XGBoost Classifier untuk Klasifikasi Kerumunan Jemaah Haji dan Umrah*. 6(January), 106–115.

Kurniawan, D., & Yasir, M. (2022). Optimization Sentimen Analysis using CRISP-DM and Naive Bayes Methods Implemented on Social Media. *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, 6(2), 74. <https://doi.org/10.22373/cj.v6i2.12793>

Kurniawan, H., Syafa'at, F., Budihartono, E., Lorosae, T. A., Apriana, D., Marisa, M., Carudin, C., Adhicandra, I., Syaddad, H. N., Ikhsan, M., & others. (2023). *BELAJAR WEB PROGRAMMING : Referensi Pengenalan Dasar Tahapan Belajar Pemrograman Web Untuk Pemula*. PT. Sonpedia Publishing Indonesia. <https://books.google.co.id/books?id=gs3OEAAAQBAJ>

Maharani, S., & Aryanta, W. R. (2023). Dampak Buruk Polusi Udara Bagi Kesehatan Dan Cara Meminimalkan Risikonya. *Jurnal Ecocentrism*, 3(2), 47–58. <https://doi.org/10.36733/jeco.v3i2.7035>

Manullang, A. H., Aritonang, M., & Purba, M. J. (2021). Sistem Informasi Bimbingan Belajar Number One Medan Berbasis Web. *TAMIKA: Jurnal Tugas Akhir Manajemen Informatika & Komputerisasi Akuntansi*, 1(1), 44–49. <https://doi.org/10.46880/tamika.vol1no1.pp44-49>



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Maulani, G., Hasan, F. N., Setiawan, D., Bowo, I. T., Ardhana, V. Y. P., Ramdhani, Y., Inayah, I., Ardiantoro, L., Sugianto, C. A., Chandra, R., & others. (2025). *Machine Learning*. MEGA PRESS NUSANTARA. <https://books.google.co.id/books?id=RblPEQAAQBAJ>

Muhammad Ibnu Sa'ad, S. K. M. K., Cresensia Devi, S. K. M. K., & Riski Galang Rahmadani, S. K. M. K. (2025). *Teori Pengembangan Perangkat Lunak Zaman Now*. Nas Media Pustaka. <https://books.google.co.id/books?id=V-FNEQAAQBAJ>

Nugraha, A. C., & Irawan, M. I. (2023). Komparasi Deteksi Kecurangan pada Data Klaim Asuransi Pelayanan Kesehatan Menggunakan Metode Support Vector Machine (SVM) dan Extreme Gradient Boosting (XGBoost). *Jurnal Sains Dan Seni ITS*, 12(1). <https://doi.org/10.12962/j23373520.v12i1.107032>

Nurfadilah, S. (2022). Perancangan Website Company Profile pada UMKM Orazio Multiusaha Indonesia. *Applied Business and Administration Journal*, 1(3), 1–8. <https://doi.org/10.62201/ha4se080>

Oriana, L., Dwi, A., Wati, S., Ramahdani, A. P., Safira, N. N., & Ismanto, E. (2025). *Peningkatan Kinerja Model Random Forest untuk Deteksi Kecurangan Kartu Kredit Menggunakan RandomizedSearchCV*. 15(2), 229–237.

Rachmi, H. (2023). Klasifikasi Pencemaran Udara Di DKI Jakarta Menggunakan Metode Naïve Bayes. *Jurnal Publikasi Ilmu Komputer Dan Multimedia (JUPIKOM)*, 2(2), 86–93. <https://doi.org/10.55606/jupikom.v2i2.18578>

Rahmawati, siti dan pratama, inka nusamuda. (2023). Pengaruh Penggunaan Transportasi Berkelanjutan Terhadap Kualitas Udara Dan Kesejahteraan Masyarakat. *Journal of Enviromental Policy and Technology*, 1(2), 95.

Rychlicki, M., Kasprzyk, Z., & Stawowy, M. (2025). *Applying Machine Learning to Preselective Weighing of Moving Vehicles*. 1–20.

Seno Aji, B. A., Setiawan, Y., & Anggraini, S. D. (2025). Analisis Perbandingan



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

- Algoritma Decision Tree, Random Forest, dan XGBoost untuk Klasifikasi Penyakit Infeksi Gigi dan Mulut. *INTEGER: Journal of Information Technology*, 10(1), 135–148. <https://doi.org/10.31284/j.integer.2024.v10i1.7501>
- Singgalen, Y. A. (2023). Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree (DT). *Jurnal Sistem Komputer Dan Informatika (JSON)*, 5(2), 237. <https://doi.org/10.30865/json.v5i2.7081>
- Subandi, M., Suryawati, D., Widnyani, I. A. P. S., Jaya, H. W., Negoro, A. H. S., Yudawisastra, H. G., & Hidayatulloh, W. (2025). *MANAJEMEN ADMINISTRASI PEMBANGUNAN*. Penerbit Widina. <https://books.google.co.id/books?id=cIl3EQAAQBAJ>
- Tantarani, A., Wiarsih, K., Nurbarokah, R., Apriani, S., & Fahlapi, R. (2024). Jumlah Penduduk DKI Jakarta Berdasarkan Kelompok Umur Tahun 2020. *Jurnal Humaniora, Sosial, Dan Bisnis*, 2(Table 10), 4–6.
- Tarigan, I. A., & Kurniawan, A. (2022). Prototipe Pendeteksi dan Pengenalan Wajah Berbasis Web Menggunakan Algoritma Local Binary Pattern Histogram untuk Absensi. *Multinetics*, 8(1), 77–86. <https://doi.org/10.32722/multinetics.v8i1.4591>
- Wafa, M. T., Puspaningrum, E. Y., & Rahajoe, A. D. (2025). *Analisis Kinerja XGBoost pada Data Seimbang dan tidak Seimbang dalam Klasifikasi Tingkat Keparahan Pasien Hemodialisis*. 8, 13928–13936.
- Wayahdi, M. R., & Ruziq, F. (2023). *Pemodelan Sistem Penerimaan Anggota Baru dengan Unified Modeling Language (UML) (Studi Kasus : Programmer Association of Battuta)*. 12, 1514–1521.
- Werdana, A. P., Setyowati, C., & Fitri, E. K. (2026). *Klasifikasi Diabetes pada Data Tidak Seimbang Menggunakan Algoritma HistGradientBoosting dengan Class Weighting dan Hyperparameter Tuning*. 2(2), 74–83.



© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Zen, M, & Rizki, M. (2024). *Pencatatan Kreatif Siswa Berbasis Android*. Serasi Media Teknologi. <https://books.google.co.id/books?id=prk0EQAAQBAJ>

Zen, Muhammad, Rizal, C., & Eka, M. (2022). *Perancangan Sistem Informasi Desa Tomuan Holbung Menggunakan Metode Waterfall*. 9(2), 274–280. <https://doi.org/10.30865/jurikom.v9i2.3986>





Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

DAFTAR RIWAYAT HIDUP PENULIS



Lanang Prasetyo

Lahir di Jakarta, 08 September 2002. Anak pertama dari tiga bersaudara. Lulus dari SDN Jatiwaringin IX pada tahun 2014, SMPN 6 Bekasi pada tahun 2017, dan SMAN 21 Bekasi pada tahun 2020. Saat ini sedang menempuh Pendidikan Sarjana Terapan pada Program Studi Teknik Informatika di Politeknik Negeri Jakarta. Tertarik pada bidang *web programming*.

**POLITEKNIK
NEGERI
JAKARTA**



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian , penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

LAMPIRAN

Lampiran 1 Dataset ISPU Jakarta

1.1 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini merupakan data Indeks Pencemaran Udara (ISPU) DKI Jakarta periode tahun 2010 – 2024 yang diambil dari portal data *Satu Data Jakarta* . Dataset ini memiliki jumlah data sebanyak 4626 baris data serta jumlah atribut yaitu sebanyak 11 kolom atribut.

1.2 Struktur Atribut Dataset

No	Nama Atribut	Tipe Data	Keterangan
1.	Tanggal	Object	Tanggal diambilnya data
2.	Stasiun	Object	Lokasi stasiun pemantau kualitas udara
3.	Pm10	Float	Konsentrasi PM10
4.	Pm25	Float	Konsentrasi PM2.5
5.	SO2	Float	Konsentrasi SO2
6.	CO	Float	Konsentrasi CO
7.	O3	Float	Konsentrasi O3
8.	NO2	Float	Konsentrasi NO2
9.	Max	Float	Nilai maksimum parameter
10.	Critical	object	Parameter pencemar dominan



1.3 Contoh Data

tanggal,stasiun,pm10,pm25,so2,co,o3,no2,max,critical,categori	
2010-01-01,DKI1 (Bunderan HI),60.0,,4.0,73.0,27.0,14.0,73.0,CO,SEDANG	
2010-01-02,DKI1 (Bunderan HI),32.0,,2.0,16.0,33.0,9.0,33.0,O3,BAIK	
2010-01-03,DKI1 (Bunderan HI),27.0,,2.0,19.0,20.0,9.0,27.0,PM10,BAIK	
2010-01-04,DKI1 (Bunderan HI),22.0,,2.0,16.0,15.0,6.0,22.0,PM10,BAIK	
2010-01-05,DKI1 (Bunderan HI),25.0,,2.0,17.0,15.0,8.0,25.0,PM10,BAIK	
2010-01-06,DKI1 (Bunderan HI),30.0,,3.0,22.0,19.0,11.0,30.0,PM10,BAIK	
2010-01-07,DKI1 (Bunderan HI),41.0,,4.0,24.0,27.0,13.0,41.0,PM10,BAIK	
2010-01-08,DKI1 (Bunderan HI),64.0,,8.0,51.0,19.0,15.0,64.0,PM10,SEDANG	
2010-01-09,DKI1 (Bunderan HI),55.0,,5.0,39.0,23.0,16.0,55.0,PM10,SEDANG	
2010-01-10,DKI1 (Bunderan HI),34.0,,5.0,23.0,26.0,12.0,34.0,PM10,BAIK	
2010-01-11,DKI1 (Bunderan HI),55.0,,10.0,32.0,23.0,16.0,55.0,PM10,SEDANG	
2010-01-12,DKI1 (Bunderan HI),45.0,,4.0,33.0,9.0,17.0,45.0,PM10,BAIK	
2010-01-13,DKI1 (Bunderan HI),28.0,,5.0,29.0,10.0,12.0,29.0,CO,BAIK	
2010-01-14,DKI1 (Bunderan HI),38.0,,6.0,28.0,15.0,10.0,38.0,PM10,BAIK	

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

POLITEKNIK
NEGERI
JAKARTA



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Lampiran 2 Source Code Model XGBoost

2.1 Source Code Import Library

```
import seaborn as sns
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.impute import SimpleImputer
from sklearn.metrics import classification_report, accuracy_score,
confusion_matrix, log_loss, roc_auc_score
from sklearn.pipeline import Pipeline
from xgboost import XGBClassifier
from collections import Counter
import joblib
import matplotlib.pyplot as plt
import os
```

Penjelasan :

Kode tersebut digunakan untuk mengimpor pustaka (*library*) yang dibutuhkan dalam proses pengolahan data, pembangunan model XGBoost, evaluasi model, visualisasi data, serta penyimpanan model hasil pelatihan.

2.2 Source Code Membaca Dataset

```
file_path = r'C:/Users/USER/SKRIPSI/archive/ispu_dki_all.csv'
df = pd.read_csv(file_path)
```




Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Penjelasan :

Kode digunakan untuk membaca dataset ISPU yang akan digunakan dalam proses pelatihan model klasifikasi.

2.3 Source Code Preprocessing Data

```
expected_columns = ['pm10', 'so2', 'co', 'o3', 'no2', 'pm25']

numerical_columns = [col for col in expected_columns if col in
df.columns]

for col in numerical_columns:

    df[col] = pd.to_numeric(df[col], errors='coerce')

    df.loc[df[col] < 0, col] = np.nan

df[numerical_columns] = df[numerical_columns].replace(0, np.nan)
```

Penjelasan :

Tahap *preprocessing* dilakukan dengan mengubah nilai negatif dan nilai nol menjadi NaN agar dapat diproses menggunakan imputasi median.

2.4 Source Code Encoding Target

```
le_target = LabelEncoder()

df['kategori'] = le_target.fit_transform(df['kategori'])
```

Penjelasan :

Label kategori kualitas udara dikonversi menjadi bentuk numerik agar dapat diproses oleh algoritma XGBoost.

2.5 Source Code Pembagian Data Training dan Testing

```
df["tanggal"] = pd.to_datetime(df["tanggal"], errors='coerce')

train_df = df[df["tanggal"] < "2022-01-01"]

test_df = df[df["tanggal"] >= "2022-01-01"]
```

Penjelasan :



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Dataset dibagi berdasarkan waktu untuk menghindari kebocoran data (*data leakage*)

2.6 Source Code Pembentukan Model XGBoost

```
base_model = XGBClassifier(  
  
    n_estimators=200,  
  
    learning_rate=0.1,  
  
    max_depth=6,  
  
    subsample=0.8,  
  
    colsample_bytree=0.8,  
  
    random_state=42,  
  
    eval_metric="mlogloss"  
)
```

Penjelasan :

Model XGBoost dibentuk menggunakan parameter terbaik yang digunakan dalam penelitian.

2.7 Source Code Pembuatan Pipeline

```
pipeline = Pipeline(steps=[  
  
    ("imputer", SimpleImputer(strategy="median")),  
  
    ("classifier", base_model)  
)
```

Penjelasan :

Pipeline digunakan untuk menggabungkan proses imputasi data dan proses klasifikasi ke dalam satu alur pemrosesan.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

2.8 Source Code Training Model

```
pipeline.fit(X_train,y_train,  
classifier__sample_weight=sample_weight)
```

Penjelasan :

Model dilatih menggunakan data training dengan menerapkan pembobotan kelas untuk mengatasi ketidakseimbangan data.

2.9 Source Evaluasi Model

```
y_pred = pipeline.predict(X_test)  
y_pred_proba = pipeline.predict_proba(X_test)  
accuracy = accuracy_score(y_test, y_pred)  
report = classification_report(  
    y_test, y_pred, target_names=le_target.classes_,  
    output_dict=True  
)  
cm = confusion_matrix(y_test, y_pred)  
logloss = log_loss(y_test, y_pred_proba)  
roc_auc = roc_auc_score(y_test_bin, y_pred_proba,  
multi_class='ovr', average='macro')  
macro_avg = report["macro avg"]  
weighted_avg = report["weighted avg"]
```

Penjelasan :

Evaluasi model dilakukan menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

2.10 Source Code Penyimpanan Model

```
joblib.dump(pipeline, "C:/Users/USER/SKRIPSI/Model Machine  
Learning/XGBOOST/xgboost_model.pkl")
```

```
joblib.dump(le_target, "C:/Users/USER/SKRIPSI/Model Machine  
Learning/XGBOOST/encoder_target.pkl")
```

Penjelasan :

Model yang telah dilatih disimpan dalam format PKL agar dapat digunakan kembali pada aplikasi web klasifikasi pencemaran udara.

