



© 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

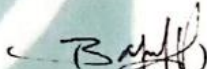
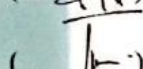
LEMBAR PENGESAHAN

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 :

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 II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

No	Judul Penelitian	Metode	Hasil	Perbedaan
1.	Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan XGBoost untuk Klasifikasi Penyakit Infeksi Gigi dan Mulut (Seno Aji et al., 2025).	(<i>Decision Tree</i> , <i>Random Forest</i> , <i>XGBoost</i>) + <i>SMOTE</i>	Algoritma XGBoost memberikan performa terbaik dari algoritma lainnya dengan nilai akurasi sebesar 93.1 % tanpa SMOTE dan 97.2% dengan SMOTE	Penelitian ini membandingkan metode <i>imbalance data handling</i> antara <i>class weight</i> , <i>smote</i> , dan tanpa penerapan <i>imbalance data handling</i> .

© 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

No	Judul Penelitian	Metode	Hasil	Perbedaan
2.	Analisis Perbandingan Algoritma <i>XGBoost</i> dan Algoritma <i>Random Forest Ensemble Learning</i> pada Klasifikasi Keputusan Kredit (Ayu et al., 2023).	<i>XGBoost</i> , <i>Random Forest</i> + <i>Random Search</i>	<i>XGBoost</i> memberikan performa yang sangat baik dengan nilai akurasi yaitu 100% pada dataset 10.000 dan dataset 100.000	Penelitian ini tidak menerapkan <i>imbalance data handling</i> .
3.	Klasifikasi Pencemaran Udara di DKI Jakarta Menggunakan Metode <i>Naïve Bayes</i> (Rachmi, 2023).	<i>Naïve Bayes</i>	Algoritma <i>Naïve Bayes</i> mendapatkan nilai akurasi sebesar 91,96%.	Penelitian ini menggunakan algoritma <i>Naïve Bayes</i> , selain itu hanya mengklasifikasikan data hanya ke dalam 3 kategori yaitu SEDANG, BAIK, TIDAK SEHAT

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 menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

No	Judul Penelitian	Metode	Hasil	Perbedaan
4.	Analisis Kinerja XGBoost pada Data Seimbang dan Tidak Seimbang dalam Klasifikasi Tingkat Keparahan Pasien Hemodialisis (Wafa et al., 2025)	<i>XGBoost</i> , <i>SMOTE</i> , <i>Stratified Train-Test Split</i>	Distribusi kelas mempengaruhi kinerja XGBoost, penerapan teknik SMOTE mampu meningkatkan kinerja model pada rasio data 60:40 dan 70:30	Penelitian ini hanya menggunakan <i>SMOTE</i> untuk <i>imbalance data handling</i> , selain itu dataset hanya dibagi pada rasio 60:40 dan rasio 70:30.
5.	Klasifikasi Diabetes pada Data Tidak Seimbang Menggunakan Algoritma HistGradientBoosting dengan Class Weighting dan Hyperparameter Tuning (Werdana et al., 2026)	<i>HistGradientBoosting</i> , <i>Class Weight</i> , <i>Grid Search Cross Validation</i>	Optimasi model mampu meningkatkan performa <i>HistGradientBoosting</i> dalam mengklasifikasi dataset diabetes yang imbalance	Penelitian ini menggunakan algoritma <i>HistGradientBoosting</i> dan menerapkan metode <i>Grid Search Cross Validation</i> untuk mencari kombinasi <i>Hyperparameter</i> terbaik.

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

No	Judul Penelitian	Metode	Hasil	Perbedaan
6.	Implementasi Algoritma XGBoost Classifier Untuk Klasifikasi Kerumunan Jamaah Haji dan Umrah (Illaihiyah & Abdurrahman, 2026)	<i>XGBoost, Class Weight, Randomized Search Cross Validation, Stratified K-Fold Cross Validation</i>	Optimasi yang dilakukan mampu memberikan performa yang baik pada algoritma <i>XGBoost</i> dalam melakukan klasifikasi pada jamaah haji dan umrah	Penelitian ini hanya menggunakan metode <i>class weight</i> tanpa membandingkan dengan metode lainnya, selain itu, penelitian ini membandingkan hasil yang diperoleh dari penerapan metode <i>Stratified K-Fold Cross Validation</i> yaitu dengan menggunakan 5 Fold dan 10 Fold.

Berdasarkan tabel penelitian terdahulu, dapat diketahui bahwa berbagai penelitian telah menunjukkan bahwa algoritma *XGBoost* mampu menghasilkan performa klasifikasi yang lebih baik dibandingkan beberapa algoritma lainnya, seperti *Decision Tree*, *Random Forest*, maupun *Naïve Bayes*. Berbagai metode optimasi juga telah diterapkan untuk meningkatkan performa model, di antaranya *SMOTE*, *Class Weight*, *Hyperparameter Tuning*, *Randomized Search*, serta *Stratified K-Fold Cross Validation*. Penelitian Wafa et al. (2025) menunjukkan bahwa penerapan *SMOTE* mampu meningkatkan kinerja *XGBoost* pada dataset tidak seimbang, sedangkan Werdana et al.(2026) membuktikan bahwa *Class Weight* dan *Hyperparameter Tuning* dapat meningkatkan performa model pada klasifikasi data imbalance. Selain itu, peneltiian Implementasi Algoritma *XGBoost Classifier* untuk Klasifikasi Kerumunan Jamaah Haji dan Umrah menerapkan kombinasi *Class Weight*, *Randomized Search Cross Validation*, dan *Stratified K-Fold Cross*



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

Validation sebagai upaya optimasi model. Meskipun demikian, sebagian besar penelitian tersebut hanya berfokus pada satu teknik penanganan data tidak seimbang atau satu pendekatan optimasi tertentu serta dilakukan pada domain permasalahan yang berbeda.

Berdasarkan hasil kajian tersebut, masih terdapat kesenjangan penelitian (*research gap*), yaitu belum banyak penelitian yang melakukan perbandingan cara komprehensif antara tanpa penanganan *imbalance data*, *Class Weight*, dan *SMOTE* pada algoritma *XGBoost*, sekaligus menganalisis pengaruh variasi *Stratified Train-Test Split*, *Randomized Search Cross Validation*, dan *Hyperparameter Tuning* dalam satu kerangka eksperimen. Oleh karena itu, penelitian ini mengusulkan penerapan kombinasi metode tersebut pada algoritma *XGBoost* untuk memperoleh model klasifikasi kualitas udara yang lebih optimal serta mengimplementasikannya ke dalam aplikasi berbasis web sebagai sarana klasifikasi data ISPU.

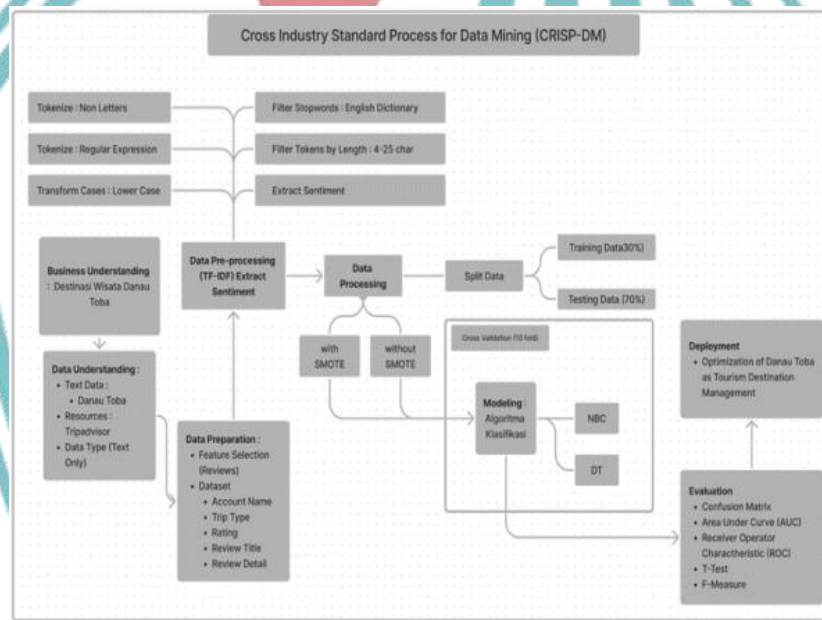
2.2 Website

Merupakan sekumpulan informasi yang menyimpan berbagai informasi yang berupa data digital baik berupa gambar, teks, video, audio ataupun animasi yang disediakan dengan menggunakan koneksi internet (Nurfadilah, 2022). Website dibedakan menjadi 2 yaitu website bersifat statis dan website bersifat dinamis. Website yang bersifat statis merupakan website yang isi kontennya tetap atau tidak berubah-ubah, hanya pemilik website yang bisa mengubah isi kontennya. Website dinamis merupakan website yang isi kontennya berubah-ubah, selain pemilik website pengguna juga dapat merubah isi konten dari website tersebut (Manullang et al., 2021).



2.3 CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan sebuah model proses industry-independen yang biasa digunakan untuk keperluan *data mining*. Metode ini terdiri dari 6 fase berulang, yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment* (Singgalen, 2023). Dalam siklus pengembangan metode ini merupakan metodologi yang sangat lengkap untuk pemenuhan kebutuhan proyek industri, selain itu di bidang *data mining* dan *data science* penggunaan metode CRISP-DM sudah digunakan secara luas untuk keperluan penelitian analitik. Tahapan yang ada pada metode CRISP-DM dapat dilihat pada Gambar 2.1 berikut ini



Gambar 2. 1 Tahapan dalam CRISP-DM

Sumber : (Singgalen, 2023)

Berdasarkan gambar berikut setiap proses atau fase pada metode CRISP-DM memiliki fungsi nya masing-masing. Tanda panah menunjukkan keterikatan setiap fase, namun, urutan setiap fase tidak selalu bersifat kaku, ada kalanya urutannya bergantung pada hasil pada setiap fase. Tujuan dari metode CRISP-DM ini adalah membuat proyek penambangan data dengan skala besar memiliki biaya yang lebih



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

murah, memiliki kinerja yang lebih cepat, serta memiliki performa yang lebih andal.

Fase pertama pada metode CRISP-DM adalah *business understanding*. Pada fase ini tujuannya adalah menganalisis dan memahami tujuan dari proyek dari perspektif bisnis. Informasi yang diperoleh dari fase ini akan dijadikan landasan dalam perencanaan proyek guna mencapai tujuan proyek yang telah didefinisikan.

Setelah masalah dan tujuan proyek telah didefinisikan fase selanjutnya adalah *data understanding*. Fase ini dimulai dengan proses pengumpulan data, mengidentifikasi data, mengidentifikasi masalah kualitas data, serta menemukan informasi penting yang terkandung di dalam data. Setelah informasi terkait data yang dibutuhkan dalam pengerjaan proyek berhasil didapatkan, proses berlanjut ke fase selanjutnya yaitu *data preparation*. Fase ini mencakup semua kegiatan yang dilakukan untuk membangun data mentah menjadi dataset akhir yang akan digunakan dalam pengerjaan proyek. Kegiatan seperti pembersihan data, transformasi data, serta pemilihan atribut dilakukan guna mendapatkan dataset yang siap digunakan.

Setelah dataset siap digunakan, maka fase selanjutnya adalah fase *modelling*. Pada fase ini dilakukan berbagai teknik pemodelan yang dipilih untuk mengolah dataset. Model dipilih harus sesuai dengan tujuan, permasalahan proyek serta format dataset yang didapatkan guna mendapatkan hasil yang optimal. Selanjutnya adalah tahap *evaluation*. Pada tahap ini model yang telah dipilih mulai dibangun untuk menghasilkan analisis data yang memiliki kualitas yang optimal. Evaluasi secara menyeluruh pada model sangat penting karena untuk memastikan bahwa model mengatasi masalah dan mencapai tujuan akhir dari proyek. Fase terakhir adalah *deployment*, semua hasil dan informasi yang didapatkan dari proses pengerjaan proyek diatur dan disajikan sedemikian rupa agar dapat diimplementasikan secara langsung di lingkungan terkait (D. Kurniawan & Yasir, 2022).



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

2.4 Klasifikasi

Klasifikasi merupakan kegiatan pengelompokan data dimana data yang digunakan memiliki label kelas atau target. Klasifikasi terbagi menjadi 2, yaitu klasifikasi biner dan klasifikasi multi kelas. Klasifikasi biner merupakan klasifikasi yang mengklasifikasikan data ke dalam dua kategori atau kelas. Sedangkan klasifikasi multi kelas yaitu klasifikasi yang mengklasifikasikan data ke dalam lebih dari dua kategori atau kelas (Maulani et al., 2025).

2.5 Class Weight

Class Weight adalah metode penanganan data tidak seimbang (*imbalance data*) yang bekerja dengan memberikan tingkat kepentingan yang berbeda pada setiap kelas selama proses pembelajaran model. Dalam metode ini, kelas minoritas memperoleh bobot yang lebih tinggi dibandingkan kelas mayoritas agar model tidak cenderung bias terhadap kelas yang memiliki jumlah data lebih banyak. Salah satu metode penentuan bobot yang banyak digunakan *inverse class frequency*, yaitu menghitung bobot setiap kelas berdasarkan rasio antara jumlah total data (N) dengan jumlah sampel pada kelas ke- k , yang kemudian dinormalisasi terhadap jumlah kelas (K). Nilai bobot tersebut selanjutnya diintegrasikan ke dalam *loss function*, sehingga kesalahan klasifikasi pada kelas minoritas akan memberikan kontribusi yang lebih besar terhadap proses optimasi model (Illaihiyah & Abdurrahman, 2026). Berikut formula dari metode *class weight* (Illaihiyah & Abdurrahman, 2026)

$$\omega_i = \frac{N}{K \times n_i}$$

Keterangan :

ω_i : Bobot kelas ke- i

N : Jumlah seluruh data latih

K : Jumlah kelas

n_i : Jumlah data pada kelas ke- i



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

2.6 Stratified K-Fold Cross Validation

K-Fold Cross Validation merupakan salah metode validasi berbasis teknik *resampling* yang digunakan untuk mengukur kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya serta mengurangi potensi *overfitting*. Metode ini bekerja dengan membagi dataset menjadi k bagian (*fold*) yang berukuran relative sama.

Selanjutnya dilakukan proses pelatihan dan pengujian sebanyak k kali, di mana pada setiap iterasi satu *fold* digunakan sebagai data pengujian, sedangkan *fold* lainnya digunakan sebagai data pelatihan. Pendekatan ini memungkinkan setiap data digunakan sebagai data latih maupun data uji secara bergantian sehingga menghasilkan estimasi performa model yang lebih akurat dan konsisten (Febrianti et al., n.d.). Alur kerja dari metode *Stratified K-Fold Cross Validation* diilustrasikan pada Gambar 2.2.



Gambar 2. 2 Alur Stratified K-Fold Cross Validation

(Sumber : (Febrianti et al., n.d.))

2.7 Randomized Search Cross Validation

RandomizedSearchCV merupakan salah satu metode optimasi *hyperparameter* yang digunakan untuk memperoleh kombinasi parameter terbaik pada suatu model *machine learning*. Metode ini bekerja dengan memilih sejumlah kombinasi nilai *hyperparameter* secara acak dari ruang pencarian (*search space*) yang telah ditentukan sebelumnya. Melalui proses tersebut, model dilatih dan dievaluasi

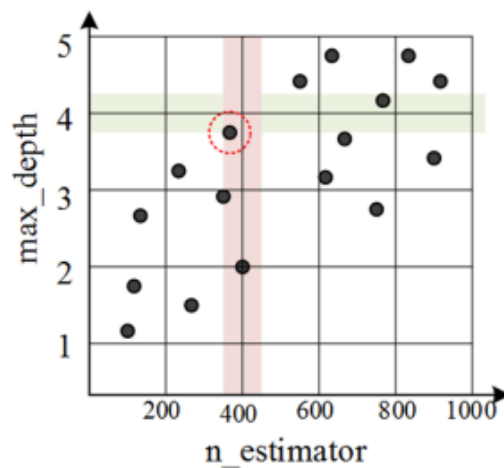


Hak Cipta :

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

menggunakan berbagai kombinasi parameter sehingga dapat ditemukan konfigurasi yang memberikan performa terbaik.

Pendekatan ini memungkinkan proses pencarian parameter dilakukan secara lebih efisien karena tidak perlu mengevaluasi seluruh kombinasi yang tersedia, namun tetap mampu menghasilkan model dengan tingkat performa yang optimal. Pada implementasinya, setiap *hyperparameter* ditentukan terlebih dahulu rentang nilai yang akan diuji, kemudian *RandomizedSearchCV* memilih kombinasi-kombinasi tersebut secara acak sebanyak jumlah iterasi yang telah ditentukan (Arifa et al., n.d.). Cara kerja *RandomizedSearchCV* diilustrasikan pada Gambar 2.3 berikut.



Gambar 2. 3 Cara Kerja RandomizedSearchCV

Sumber : (Oriana et al., 2025)

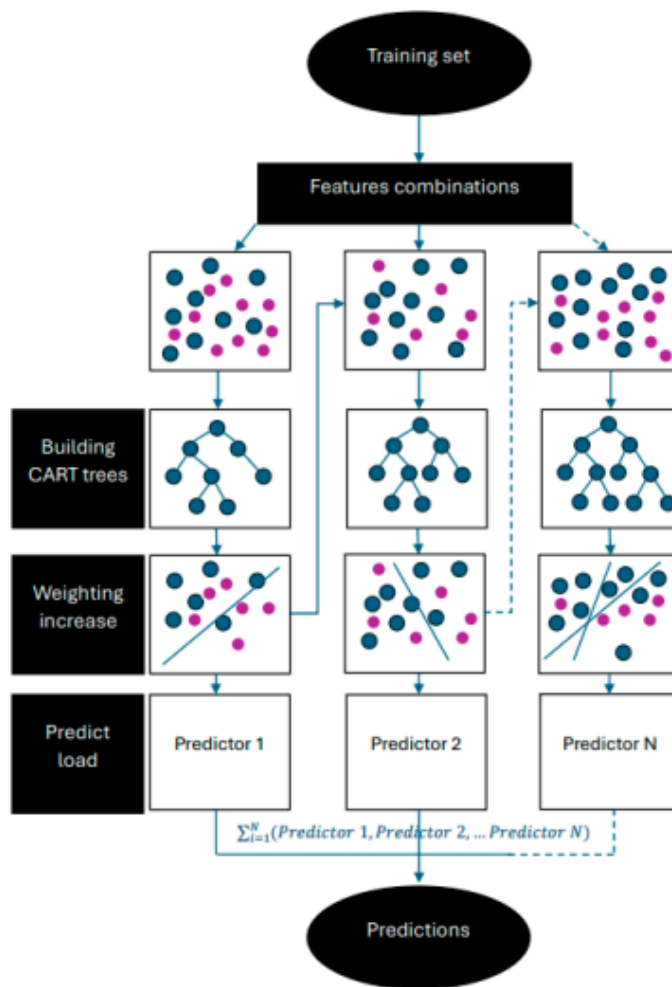
2.8 XGBoost

Extreme Gradient Boosting atau XGBoost merupakan salah satu metode *machine learning* yang digunakan untuk klasifikasi atau regresi. XGBoost adalah algoritma yang menggunakan teknik *tree boosting* yang bisa diperluas untuk menciptakan sistem yang lebih kompleks. XGBoost merupakan bentuk perluasan dari metode *Gradient Boosting*, yang cara kerjanya menggabungkan berbagai model *machine learning* untuk meningkatkan performa klasifikasi.



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 2. 4 Algoritma XGBoost
(Sumber: (Rychlicki et al., 2025))

Di dalam algoritma XGBoost, hasil klasifikasi dari model sebelumnya yang disebut dengan *residuals* dijadikan sebagai bahan evaluasi agar kinerja algoritma meningkat. Algoritma XGBoost dirancang untuk meningkatkan waktu eksekusi terutama saat memproses data yang besar (Nugraha & Irawan, 2023).

Dalam algoritma XGBoost prosesnya dibagi menjadi beberapa tahapan (Firdaus et al., 2025). Tahap pertama yang dilakukan adalah inisialisasi, yaitu pengambilan nilai prediksi awal untuk seluruh data yang berasal dari rata-rata seluruh data.



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

$$L(\phi) = \sum l(y_i, \hat{y}_i^{(t)}) + \sum \Omega(f_k)$$

Keterangan :

$L(\phi)$: Fungsi objektif total dari model.

y_i : Nilai prediksi.

$\Omega(f_k)$: Fungsi regularisasi terhadap pohon ke-k.

$l(y_i, \hat{y}_i)$: Fungsi kerugian prediksi.

Prediksi awal inilah yang menjadi dasar untuk proses perbaikan lebih lanjut yang dilakukan di dalam pohon keputusan dalam iterasi selanjutnya yang bertujuan untuk mengurangi kesalahan prediksi yang dilakukan secara bertahap melalui proses *boosting*. Langkah selanjutnya adalah menghitung nilai gradien dan hessian. Gradien adalah turunan pertama dari fungsi loss terhadap proses prediksi, yang menunjukkan nilai kesalahan prediksi yang saat ini dilakukan dan menentukan langkah perbaikan yang harus dilakukan model.

Hessian merupakan turunan kedua dari keturunan kedua dari nilai fungsi kerugian terhadap proses prediksi. Fungsi dari nilai Hessian adalah mengukur kelengkungan fungsi kerugian yang berguna untuk menentukan besarnya nilai angka perbaikan yang harus dilakukan.

**POLITEKNIK
NEGERI
JAKARTA**



© 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

$$g_i = \partial l / \partial \hat{y}_i, \quad h_i = \partial^2 l / \partial \hat{y}_i^2$$

Keterangan :

g_i : Gradien dari loss terhadap prediksi.

h_i : Hessian dari loss.

$\hat{y}_i$: Probabilitas hasil prediksi untuk data ke-i.

Langkah selanjutnya dalam tahap perhitungan pada algoritma XGBoost adalah memilih pembagian yang terbaik berdasarkan nilai dari hessian dan gradient. Tahap ini bertujuan untuk menghitung nilai peningkatan kualitas setelah proses data yang dibagi menjadi dua bagian pada suatu node pohon.

$$\text{Gain} = \frac{1}{2} \left[\frac{(\sum g_L)^2}{(\sum h_L + \lambda)} + \frac{(\sum g_R)^2}{(\sum h_R + \lambda)} - \frac{(\sum g)^2}{(\sum h + \lambda)} \right] - \gamma$$

Keterangan :

Gain : Keuntungan yang diperoleh dari *split*.

Λ : Parameter regularisasi untuk penalti besar bobot leaf.

γ : Parameter regularisasi untuk penalti jumlah daun.

g_L : Jumlah gradien pada cabang kiri setelah split.

h_L : Jumlah hessian pada cabang kiri.

g_R : Jumlah gradien pada cabang kanan.

h_R : Jumlah hessian pada cabang kanan.



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

Pada tahap pembagian ini, algoritma akan melakukan berbagai kemungkinan cara untuk membagi data berdasarkan fitur-fitur yang ada. Selain itu tujuan dari tahap ini adalah menemukan pemisahan data yang lebih homogen, sehingga nilainya lebih dekat dengan nilai target dan juga meningkatkan nilai akurasi prediksi. Tahap selanjutnya adalah regularisasi. Tujuan dari proses ini adalah untuk mencegah *overfitting* dengan cara menambahkan penalti terhadap kompleksitas dari model yang dilihat berdasarkan jumlah daun pada pohon dan besar nilai pada daun tersebut.

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum w_j^2$$

Keterangan :

$\Omega(f_k)$: Fungsi regularisasi pada pohon ke-k.

λ : Parameter regularisasi untuk penalti besar bobot leaf.

γ : Parameter regularisasi untuk penalti jumlah daun.

T : Jumlah daun dalam satu pohon.

w_j : Bobot pada daun ke-j.

Setelah pembagian yang optimal dan pohon telah dibuat, tahap terakhir yaitu memperbarui prediksi akhir. Tahap ini dilakukan dengan menambahkan nilai hasil prediksi dengan nilai hasil prediksi sebelumnya, dengan hasil perkalian dengan *learning rate*. Proses ini bertujuan untuk meningkatkan nilai akurasi model secara bertahap dengan cara mengintegrasikan setiap kontribusi dari masing-masing pohon baru yang dibuat dalam iterasi.

$$\hat{y}_i = \hat{y}_i + \eta f_k(x_i)$$

Keterangan :

$\hat{y}_i$: Nilai aktual data.

η : Learning rate.

$f_k(x_i)$: Kontribusi pohon keputusan baru untuk input x .



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

- Accuracy : Mengukur persentase prediksi yang benar dibandingkan dengan keseluruhan data

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

- Precision : Mengukur ketepatan prediksi positif yang dihasilkan model.

$$Precision = \frac{TP}{TP+FP}$$

- Recall : Mengukur kemampuan model dalam menemukan seluruh data positif yang benar.

$$Recall = \frac{TP}{TP+FN}$$

- F1-Score : Merupakan *harmonic mean* dari *Precision* dan *Recall*.

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

- Log Loss : Mengukur seberapa baik model mengeluarkan probabilitas prediksi yang mendekati nilai sebenarnya. Nilai yang lebih rendah menunjukkan model lebih baik.

$$LogLoss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \cdot \log(p_{ij})$$

- ROC-AUC (Area Under Curve) : Mengukur kemampuan model dalam membedakan antar kelas. Nilai AUC yang mendekati 1 menunjukkan model memiliki performa yang sangat 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

2.9 Waterfall

Metode waterfall merupakan metode klasik dalam pengembangan perangkat lunak. Proses pengembangan perangkat lunak pada metode ini bersifat linear dan berurutan, setiap fasenya dilakukan secara bertahap dan tidak ada yang boleh terlewat dari awal hingga akhir proses pengembangan perangkat lunak. Metode waterfall cocok digunakan untuk proyek pembuatan perangkat lunak yang kebutuhannya jelas dan stabil, artinya komponen-komponen yang mendukung proses pembuatan perangkat lunak tidak berubah (Subandi et al., 2025).

Metode waterfall memiliki struktur proses yang sangat terorganisir. Setiap fase yang ada dijalankan secara berurutan, sehingga memberikan gambaran mengenai urutan apa yang dilakukan terlebih dahulu dan apa yang harus dilakukan setelahnya. Hal ini membuat pengembang untuk menyusun jadwal dan menetapkan anggaran selama proses pembuatan. Selain itu, metode ini merupakan metode yang sifatnya sederhana dalam pelaksanaannya. Selama proses berjalan pengembang sudah tahu mengenai apa saja yang harus dilakukan di setiap fasenya, artinya pengembang pengembang bisa menyelesaikan satu tahap penuh sebelum berpindah ke tahap berikutnya tanpa mendapatkan gangguan dari tahap-tahap yang lainnya. Hal ini bisa mengurangi risiko kesalahan dalam pengerjaan.

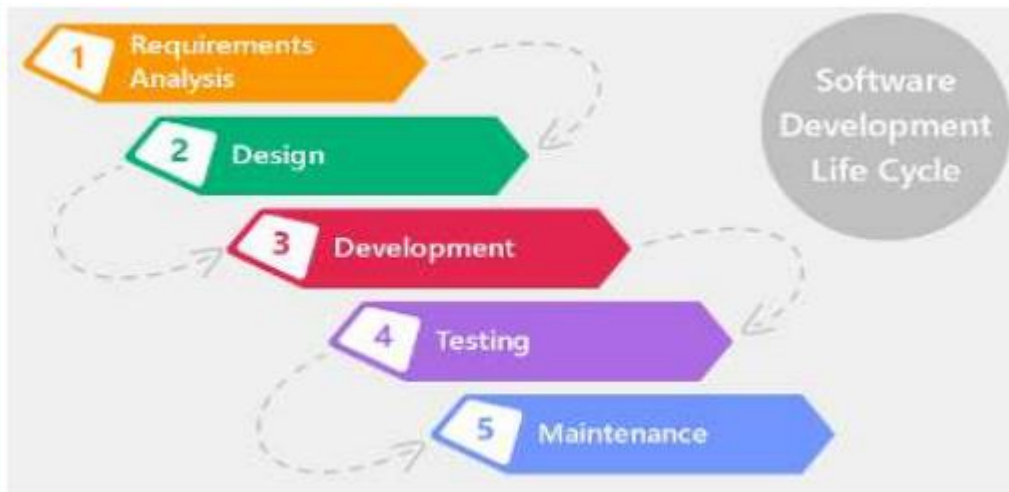
Meskipun waterfall merupakan metode yang terstruktur dan sederhana, metode ini juga memiliki kelemahan yaitu kurangnya fleksibilitas. Karena pada prosesnya setiap tahap harus dikerjakan secara tuntas sebelum berpindah ke fase yang berikutnya, hal ini justru sangat rentan jika ada perubahan di dalam pengerjaan. Tentunya perubahan yang dilakukan di tengah proses pengerjaan dapat mempengaruhi jadwal dan biaya yang telah ditetapkan. Jika ada perubahan besar yang harus dilakukan, hal itu akan membuat proses pengerjaan bisa diulang kembali ke tahapan awal (Muhammad Ibnu Sa'ad et al., 2025).



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

Dalam pengembangannya metode waterfall terdiri dari beberapa tahapan yang berurutan, yaitu :



Gambar 2. 5 Alur Metode Waterfall

(Sumber:(Muhammad Zen et al., 2022))

1. Analisis Kebutuhan

Tahap ini merupakan tahapan awal dari proses metode waterfall. Dimana pengembang melakukan pengumpulan informasi mengenai hal-hal apa saja yang dibutuhkan dalam proses pembuatan sebuah perangkat lunak. Informasi tersebut didapatkan melalui berbagai cara yaitu wawancara, observasi, atau survei langsung. Kemudian informasi yang didapatkan dianalisis untuk selanjutnya ditentukan kebutuhan fungsional maupun non fungsional berdasarkan hasil analisis tersebut.

2. Desain Sistem

Setelah menentukan kebutuhan apa saja yang diperlukan untuk membuat sistem, maka kebutuhan tersebut akan diterjemahkan untuk membuat desain teknis sistem. Desain yang dibuat dalam tahap ini adalah desain antarmuka pengguna, desain arsitektur perangkat lunak, desain struktur sistem, desain database. Desain yang dibuat dibuat dalam bentuk dokumen teknis, gambar, diagram, skema yang menjadi landasan atau fondasi dalam pembuatan sebuah sistem perangkat lunak.



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

3. Implementasi

Pada tahap inilah sistem mulai dibuat, yaitu dengan menulis kode berdasarkan desain sistem yang telah dibuat pada tahap sebelumnya. Struktur dan skema-skema sistem perangkat lunak mulai dikembangkan dengan mengikuti rancangan yang sudah ditetapkan. Setiap bagian kode harus efisien dan bisa diintegrasikan dengan bagian yang lain.

4. Integrasi dan Testing

Setelah sistem selesai dibuat maka tahap selanjutnya adalah pengintegrasian setiap bagiannya menjadi sebuah sistem yang utuh. Setelah diintegrasikan maka sistem diuji kinerjanya. Pengujian dilakukan secara bertahap, mulai dari pengujian unit, pengujian integrasi antar modul, hingga pengujian keseluruhan sistem. Jika ditemukan kesalahan atau *bug*, maka hal itu harus segera diperbaiki agar sistem berjalan sesuai dengan yang diharapkan.

5. Deployment dan Maintenance

Tahap akhir dalam metode waterfall adalah perangkat lunak yang sudah jadi secara utuh dan telah diuji kinerjanya mulai digunakan ke pengguna akhir. Tahap ini mencakup instalasi sistem serta penyusunan dokumentasi seperti panduan penggunaan dan dokumentasi teknis. Selain itu, perlu dilakukannya maintenance, dimana performa atau kinerja sistem akan terus dipantau dan memperbaiki bug yang muncul setelah perangkat lunak digunakan secara nyata oleh pengguna akhir (M Zen & Rizki, 2024).



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

2.10 Python

Python adalah bahasa pemrograman model skrip yang berorientasi objek. Python digunakan untuk berbagai kebutuhan pengembangan perangkat lunak dan berjalan pada *platform* sistem operasi. Python adalah bahasa pemrograman *freeware* yang tidak memiliki Batasan dalam hal penyalinan atau penyebaran lengkap dengan berbagai arsitekturnya (Tarigan & Kurniawan, 2022).

Python merupakan bahasa pemrograman yang memiliki beberapa keunggulan yaitu Python dapat digunakan untuk banyak keperluan yaitu pembuatan aplikasi berbasis *desktop*, pembuatan aplikasi berbasis *web*, Python juga memiliki Pustaka pihak ketiga yang mendukung proses komputasi kompleks yang berbasis kecerdasan buatan seperti *machine learning* dan *deep learning*, Python juga dapat terhubung ke basis data dan dapat membaca dan memodifikasi file, python juga dapat digunakan untuk menangani data dengan jumlah besar dan dapat melakukan perhitungan matematika yang kompleks (Dr. Abdiansah, 2022).

2.11 Laravel

Laravel merupakan *framework* PHP yang bersifat *open source* yang dibangun dengan menggunakan konsep *model view controller* dan menyediakan berbagai macam sintaks. Laravel diluncurkan pada tahun 2011, sejak saat itu Laravel digemari oleh berbagai komunitas programmer dan akhirnya banyak dipakai oleh *programmer* di seluruh dunia. Laravel terus mengalami pengembangan dengan menyediakan berbagai fitur baru, memperbaiki bug, meningkatkan kinerja, dan arsitekturnya selalu diperbarui secara teratur. Laravel menyediakan struktur yang jelas dan konsisten dalam proses pengembangan aplikasi web, karena Laravel didasarkan pada beberapa konsep seperti pola desain MVC (*Model-View-Controller*), ORM (*Object-Relational Mapping*), serta *routing* yang kuat (H. Kurniawan et al., 2023).



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

Selain karena kepopuleran dari Laravel karena menjadi *framework* .yang disukai oleh banyak *programmer*, Laravel memiliki beberapa keunggulan lainnya, yaitu :

- Laravel memiliki beberapa fitur yang tidak dimiliki oleh *framework* lainnya, fitur-fitur tersebut seperti Artisan, migrations, blade, dan masih banyak lagi.
- Laravel memiliki sintaks yang mudah dimengerti oleh *programmer*.
- Dokumentasi dari Laravel sangat lengkap, bahkan di setiap versi Laravel memiliki dokumensinya masing-masing.
- Karena Laravel merupakan *framework* yang disukai oleh banyak *programmer*, maka banyak library pendukung yang dibuat.
- Laravel didukung oleh Composer, sehingga banyak library Laravel yang bisa didapatkan dengan menggunakan Composer. (Abdulloh, 2022)

2.12 UML

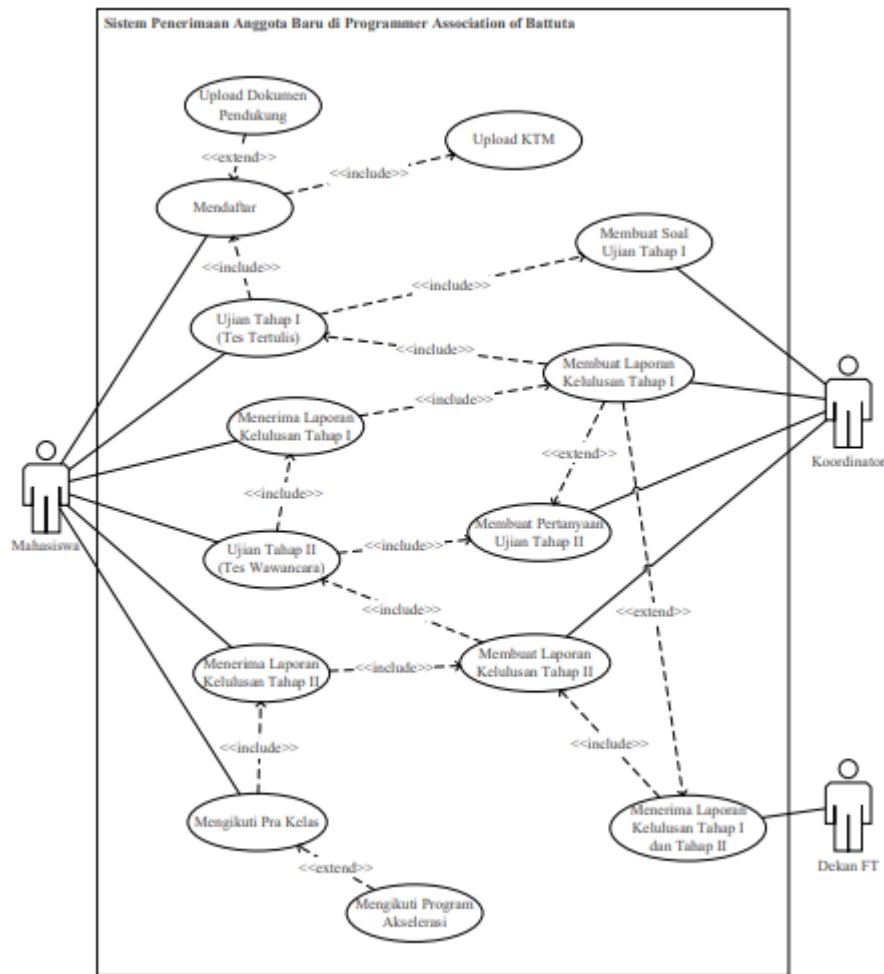
UML (*Unified Modeling Language*) merupakan kelompok notasi grafis yang didukung oleh model-model tunggal, yang bertujuan untuk membantu penggambaran, pendeskripsian, dan desain sistem perangkat lunak khususnya sistem yang berorientasi objek. Selain itu UML juga menjabarkan sebuah sistem secara rinci yang digunakan untuk menganalisis dan mencari apa yang diperlukan oleh sistem, UML juga digunakan untuk mendokumentasikan sebuah sistem beserta proses-proses dan organisasinya. Dalam pengertian lain UML merupakan sebuah bahasa visual yang digunakan untuk pemodelan dan penggambaran komunikasi sebuah sistem dengan dilengkapi dengan teks-teks pendukung (Anggrawan et al., 2023).



2.13 Use Case Diagram

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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



Gambar 2. 6 Contoh Diagram Use Case

(Sumber:(Wayahdi & Ruziq, 2023))



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

Use Case Diagram merupakan salah satu dari berbagai macam jenis UML diagram yang menggambarkan hubungan antara aktor dengan sebuah sistem. Use Case Diagram menjadi sebuah model fungsional yang menggambarkan sebuah perilaku (*Behavior System*) dari sebuah sistem yang akan dibangun. Sebuah Use Case Diagram dapat dicirikan dengan sebuah atau beberapa aktor yang terikat dengan sistem. Aktor yang terikat pada sistem memiliki status serta aksesnya sesuai dengan fungsi di dalam sebuah sistem aplikasi. Pada intinya, *Use Case Diagram* digunakan untuk mengetahui dan menggambarkan sebuah fungsi di dalam sebuah sistem dan siapa saja aktor yang terikat pada sistem yang dibuat (Ahmad et al., 2022).

2.14 ISPU

Indeks Standar Pencemaran Udara (ISPU) merupakan angka yang tidak memiliki satuan yang menjadi dasar dalam menggambarkan kondisi atau kualitas udara di sebuah wilayah dalam waktu tertentu yang didasarkan pada dampak kesehatan yang ditimbulkan terhadap masyarakat dan lingkungan sekitar.

Nilai ISPU merupakan hasil dari perhitungan rata-rata partikel yang terkandung di dalam udara yang menjadi parameter dalam pengukuran ISPU. Parameter tersebut diantaranya yaitu partikulat matter (PM10), ozon (O3), karbon monoksida (CO), sulfur dioksida (SO2), dan nitrogen dioksida (NO2) (Giantara et al., 2021).



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 III

METODE PENELITIAN

3.1 Rancangan Penelitian

Penelitian ini berfokus pada pengembangan aplikasi berbasis web yang bertujuan untuk melakukan klasifikasi pencemaran udara di wilayah Jakarta dengan menggunakan data ISPU kota Jakarta periode 2010 – 2024 dan model *machine learning* yang digunakan adalah algoritma XGBoost. Pendekatan kuantitatif digunakan dalam penelitian ini, di mana analisis dilakukan terhadap data numerik yang dihasilkan dari model XGBoost serta hasil pengujian aplikasi web untuk mengevaluasi dan memvalidasi eksperimen yang dilakukan. Oleh karena itu, penelitian ini termasuk ke dalam kategori riset eksperimental, yang berfokus pada pengujian dan evaluasi model klasifikasi serta implementasinya dalam sebuah sistem berbasis web.

Dalam pengembangan model klasifikasi pencemaran udara, penelitian ini menerapkan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang mencakup tahapan *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*. Sedangkan untuk pengembangan aplikasi web, digunakan metode *Waterfall* yang mencakup analisis kebutuhan, desain sistem, implementasi, pengujian, dan pemeliharaan.

Pengumpulan data yang digunakan untuk membangun model dilakukan dengan memanfaatkan dataset Indeks Standar Pencemar Udara (ISPU) DKI Jakarta periode 2010 -2024. Dataset tersebut bersumber dari *Satu Data Jakarta* sebagai portal data resmi Pemerintah Provinsi DKI Jakarta. Teknik pengumpulan data yang digunakan adalah pemanfaatan data sekunder sebagai bahan pengembangan model *machine learning* serta uji coba langsung terhadap aplikasi web yang dikembangkan.



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

Tahap pengujian pada pengembangan aplikasi dilakukan untuk memastikan setiap fungsi berjalan sesuai dengan yang diharapkan. Pengujian ini menggunakan metode *Black Box Testing* yaitu pengujian yang berfokus pada hasil keluaran berdasarkan masukan yang diberikan tanpa memeriksa proses internal sistem. Pengujian dilakukan pada seluruh fitur utama aplikasi, seperti proses unggah file CSV berisi data ISPU, klasifikasi data dengan model XGBoost, dan menampilkan hasil klasifikasi pada antarmuka pengguna.

Tujuan utama dari penelitian ini adalah menghasilkan sebuah sistem yang dapat melakukan klasifikasi data Indeks Standar Pencemar Udara (ISPU) secara akurat, stabil, dan efisien. Selain itu, penelitian ini juga bertujuan untuk menganalisis kelebihan dan kekurangan yang dimiliki baik dari sisi model XGBoost maupun dari sisi implementasi web, sehingga dapat memberikan rekomendasi untuk peningkatan dan pengembangan lebih lanjut.

3.2 Tahap Penelitian

Tahap penelitian dibagi menjadi dua bagian, yaitu tahap pembuatan model *machine learning* dan tahapan pembuatan aplikasi web. Kedua proses tersebut saling terintegrasi untuk menghasilkan sistem klasifikasi pencemaran udara berbasis web.

3.3 Tahapan Pembangunan Model Machine Learning

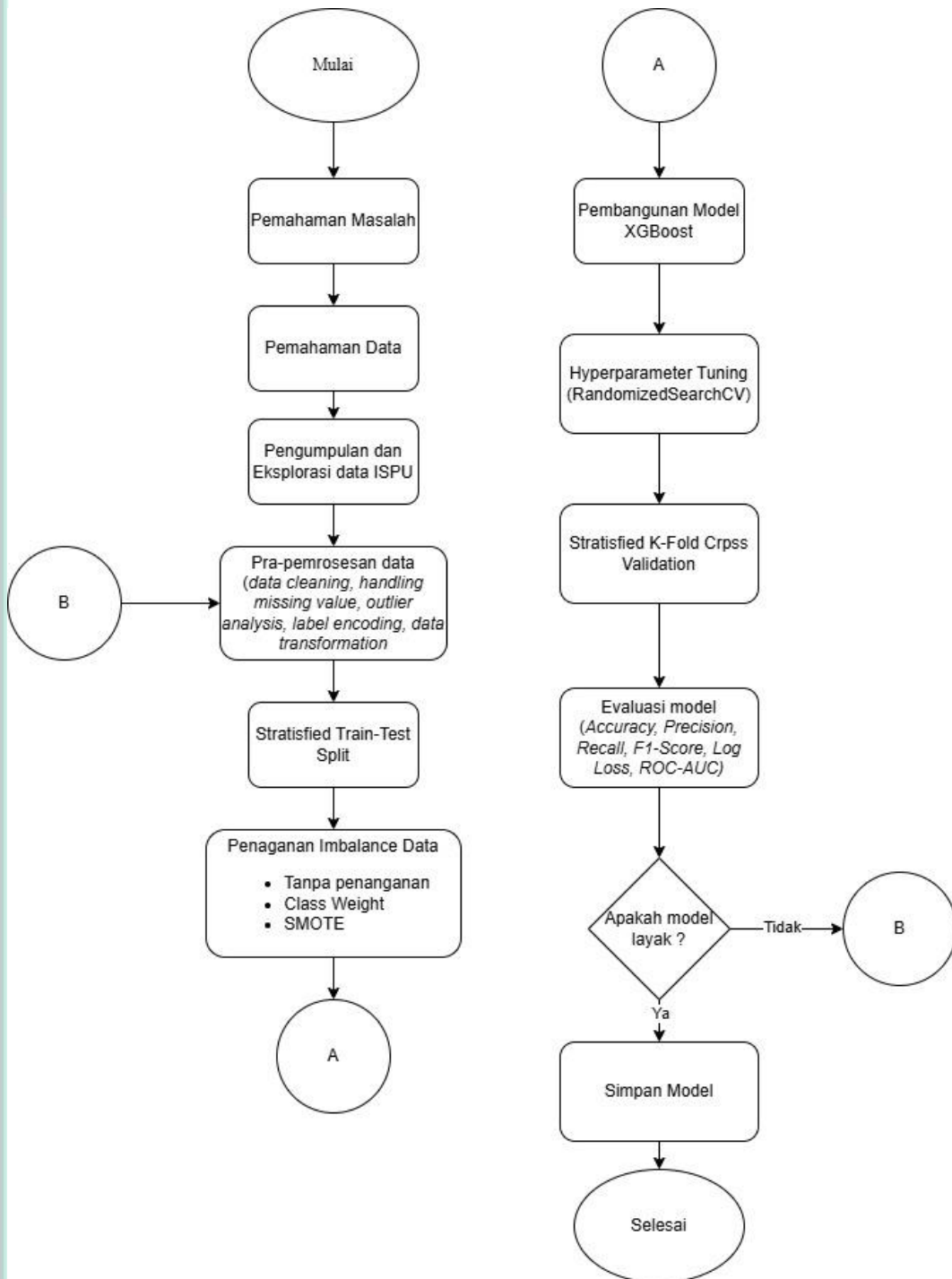
Proses pembuatan model *machine learning* pada penelitian ini menerapkan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Sehingga setiap tahapan dalam pembuatannya disesuaikan dengan alur tahapan pada metode CRISP-DM.



© 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 3. 1 Flowchart Diagram Model Machine Learning



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. *Business Understanding*

Pada tahap *Business Understanding*, ditentukan tujuan utama penelitian, yaitu mengembangkan sistem klasifikasi kualitas udara berbasis web yang mampu mengklasifikasikan kategori kualitas udara berdasarkan data Indeks Standar Pencemar Udara (ISPU). Selain itu, penelitian ini juga bertujuan memperoleh model klasifikasi yang memiliki performa yang baik melalui penerapan beberapa teknik, seperti penanganan ketidakseimbangan kelas (*imbalanced data handling*), *hyperparameter tuning*, serta evaluasi model menggunakan berbagai metrik klasifikasi sehingga model yang dihasilkan memiliki kemampuan generalisasi yang baik terhadap data baru.

2. *Data Understanding*

Tahap *Data Understanding* dilakukan dengan mengumpulkan dan memahami karakteristik dataset yang akan digunakan dalam proses pembangunan model. Dataset yang digunakan merupakan data sekunder berupa data Indeks Standar Pencemar Udara (ISPU) DKI Jakarta periode 2010–2024 sebanyak 4.625 data. Dataset tersebut bersumber dari Satu Data Jakarta sebagai portal data resmi Pemerintah Provinsi DKI Jakarta. Pada tahap ini dilakukan identifikasi terhadap atribut yang digunakan, tipe data, distribusi kelas, serta eksplorasi awal untuk mengetahui keberadaan missing value, distribusi nilai pada setiap parameter polutan, dan karakteristik data yang akan mempengaruhi proses pembangunan model.

3. *Data Preparation*

Tahap *Data Preparation* bertujuan menyiapkan data agar siap digunakan pada proses pembentukan model. Tahapan yang dilakukan meliputi normalisasi label kategori, normalisasi nama atribut, konversi nilai numerik, penanganan nilai negatif dan nilai nol menjadi *missing value*, serta imputasi nilai kosong menggunakan median.



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

Selanjutnya dilakukan penghapusan atribut yang tidak digunakan sebagai prediktor, analisis *outlier* tanpa menghapus data, transformasi label menggunakan *Label Encoding*, dan pemisahan data menggunakan *Stratified Train-Test Split* untuk mempertahankan proporsi setiap kelas pada data latih dan data uji. Pada tahap ini juga dilakukan penerapan dan perbandingan tiga skenario penanganan ketidakseimbangan data, yaitu tanpa penanganan, *Class Weight*, dan *Synthetic Minority Oversampling Technique* (SMOTE).

4. Modeling

Pada tahap *Modeling*, dilakukan pembangunan model klasifikasi menggunakan algoritma *Extreme Gradient Boosting* (XGBoost). Sebelum proses pelatihan model, dilakukan optimasi parameter menggunakan *RandomizedSearchCV* untuk memperoleh kombinasi parameter yang paling sesuai. Proses pencarian parameter terbaik divalidasi menggunakan *Stratified K-Fold Cross Validation* sebanyak lima lipatan (5-fold) sehingga setiap kombinasi parameter dapat dievaluasi secara lebih objektif. Setelah diperoleh parameter terbaik, model kemudian dilatih menggunakan data latih sesuai dengan skenario penanganan imbalanced data yang sedang diuji.

5. Evaluation

Tahap *Evaluation* dilakukan untuk menilai performa model klasifikasi yang telah dibangun. Evaluasi dilakukan menggunakan beberapa metrik, yaitu *Accuracy*, *Precision*, *Recall*, *F1-Score*, *Log Loss*, dan *ROC-AUC*, serta divisualisasikan menggunakan *Confusion Matrix* untuk mengetahui distribusi hasil prediksi pada setiap kelas. Selain itu, dilakukan pula evaluasi menggunakan *Stratified K-Fold Cross Validation* untuk mengukur konsistensi performa model pada beberapa pembagian data, serta analisis *Feature Importance* untuk mengetahui tingkat kontribusi masing-masing parameter polutan terhadap hasil klasifikasi. Hasil evaluasi dari setiap skenario penanganan imbalanced data kemudian dibandingkan untuk menentukan model terbaik yang akan diimplementasikan pada aplikasi web.



© 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

6. Deployment

Tahap *Deployment* merupakan tahap akhir dalam proses pembangunan model *machine learning*. Model terbaik hasil proses evaluasi disimpan dalam format *Pickle* (.pkl) beserta *Label Encoder* sehingga dapat diintegrasikan ke dalam aplikasi web berbasis *Laravel*. Model yang telah diintegrasikan memungkinkan sistem melakukan klasifikasi secara otomatis terhadap data ISPU yang diunggah pengguna melalui berkas *CSV*, kemudian menampilkan hasil klasifikasi beserta informasi pendukung yang tersedia pada aplikasi.

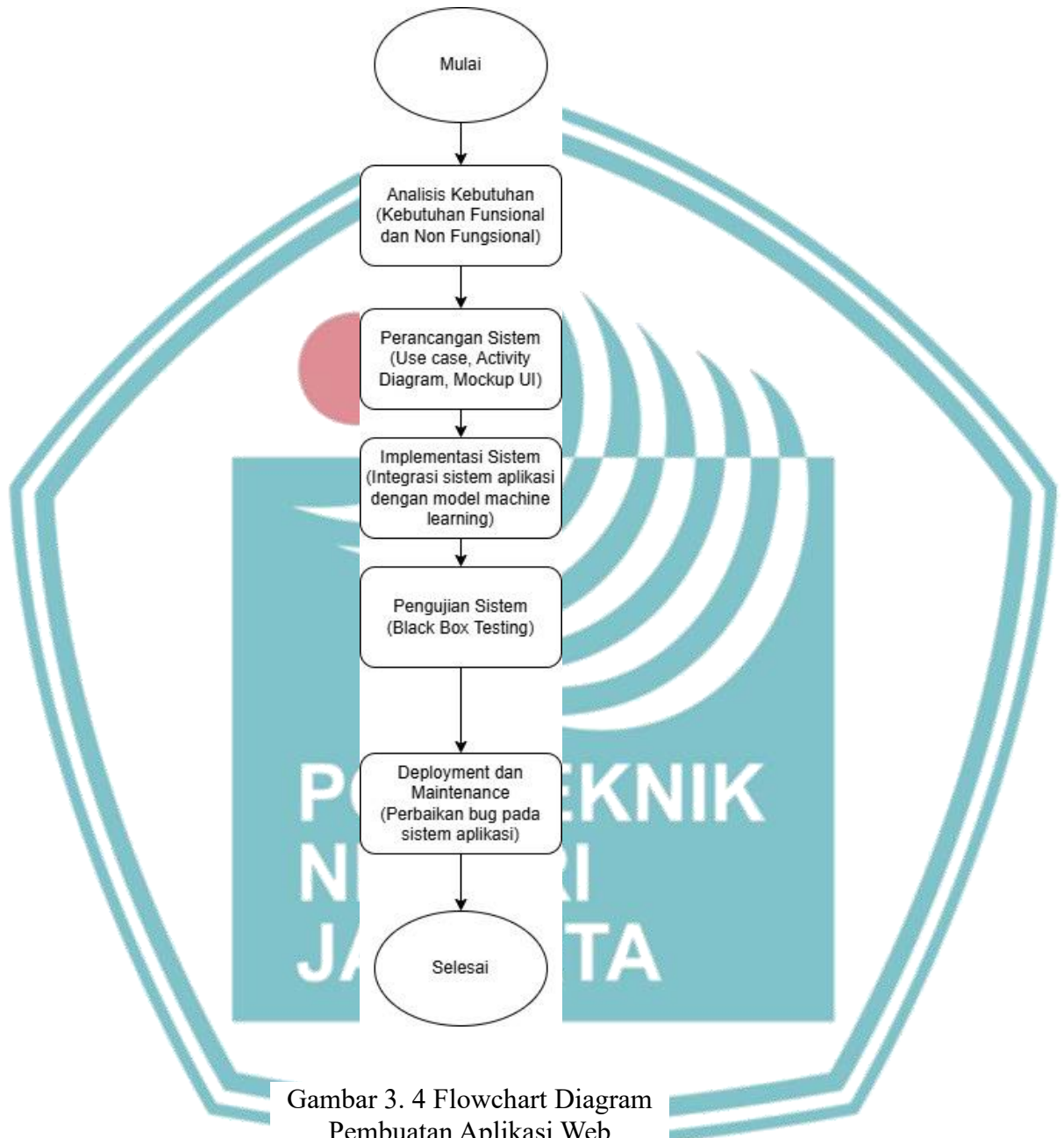




Hak Cipta :

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

3.4 Tahap Pembuatan Aplikasi Web



Gambar 3. 4 Flowchart Diagram Pembuatan Aplikasi Web



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

Proses pembuatan aplikasi web pada penelitian ini menerapkan metode pengembangan Waterfall, di mana tahapannya meliputi :

1. Analisis Kebutuhan

Tahapan awal dalam melakukan pembuatan sistem aplikasi web adalah mengidentifikasi kebutuhan fungsional dan kebutuhan non fungsional. Kebutuhan fungsional meliputi proses pengunggahan file CSV, klasifikasi data menggunakan model XGBoost, dan menampilkan hasil klasifikasi. Sedangkan kebutuhan non fungsional meliputi tampilan antarmuka yang responsif, kecepatan proses klasifikasi.

2. Desain Sistem

Pada tahap ini, dilakukan perancangan arsitektur sistem dengan menggunakan diagram UML untuk menggambarkan bagaimana sistem aplikasi bekerja.

3. Implementasi

Sistem aplikasi mulai dibangun pada tahap ini. Rancangan sistem yang telah dibuat mulai diterjemahkan ke dalam bentuk kode untuk menghasilkan sistem aplikasi yang sesuai dengan tujuan. Aplikasi ini dibangun menggunakan *framework* Laravel dan untuk membuat tampilan atau antarmuka menggunakan *framework* Tailwind CSS. Model XGBoost diintegrasikan dengan sistem aplikasi melalui perantara file python yang juga dihubungkan dengan *controller* aplikasi web.

4. Pengujian

Aplikasi yang telah selesai dibangun dilakukan pengujian terhadap kinerjanya dengan menggunakan metode *Black Box Testing*. Pengujian dilakukan pada fitur upload file CSV, proses klasifikasi, tampilan hasil klasifikasi, serta mengunduh data hasil klasifikasi.



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

5. Pemeliharaan (*Maintenance*)

Aplikasi yang telah selesai dibangun dilakukan pemeliharaan terhadap kinerja nya. Pemeliharaan tersebut meliputi perbaikan *bug* yang terjadi selama penggunaan, pembaruan model dalam menangani data terbaru, serta penambahan fitur sesuai dengan kebutuhan pengguna di masa yang akan datang.

3.5 Objek Penelitian

Objek pada penelitian ini adalah perancangan sistem klasifikasi pencemaran udara menggunakan data Indeks Standar Pencemar Udara (ISPU). Dataset yang digunakan pada penelitian ini diambil dari portal data *Satu Data 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

BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisa Kebutuhan

Berdasarkan studi literatur, beberapa kebutuhan untuk penelitian ini berhasil teridentifikasi. Kebutuhan yang terkait penelitian ini meliputi kebutuhan pengembangan model, pengembangan web, serta kebutuhan perangkat keras dan kebutuhan perangkat lunak. Semua kebutuhan ini harus dipastikan semua pemenuhannya untuk kelancaran dan keberhasilan jalannya proses penelitian.

4.1.1 Kebutuhan Pengembangan Model

- Kebutuhan Dataset

Dalam membangun sebuah model yang efektif dan akurat, diperlukan dataset yang sesuai dengan kebutuhan penelitian. Dataset harus mencakup variabel-variabel yang relevan dengan penelitian, seperti parameter PM_{10} , $PM_{2.5}$, SO_2 , CO , O_3 , dan NO_2 , beserta kategori tingkat pencemaran udara. Jumlah data yang digunakan harus cukup besar untuk memastikan model memiliki kemampuan generalisasi yang baik dan tidak *overfitting*. Dataset harus didapatkan dari sumber yang terpercaya, dalam penelitian ini dataset bersumber dari portal data *Satu Data Jakarta*, memiliki kualitas data yang terjaga minim kesalahan, serta minim jumlah data yang hilang. Sebelum digunakan dalam proses pelatihan dan pengujian model, dataset perlu melalui tahap *preprocessing data* yang mencakup pembersihan data, penanganan nilai N/A, penyesuaian format, konversi label kategori, serta pemilihan fitur agar dapat diproses oleh algoritma yang digunakan dalam penelitian, dalam hal ini XGBoost. Kriteria *dataset* yang dibutuhkan untuk mengembangkan model klasifikasi pencemaran udara menggunakan algoritma XGBoost disajikan pada Tabel 4.1

Tabel 4. 1 Tabel Kebutuhan Dataset

Kode	Nama	Deskripsi
KD-1	Sumber Data	Sumber data harus terpercaya
KD-2	Periode Data	Data ISPU Jakarta periode 2010 – 2024
KD-3	Format Data	File data ISPU memiliki format CSV
KD-4	Variabel Utama	Variabel yang dibutuhkan penelitian yaitu PM ₁₀ , PM _{2.5} , SO ₂ , CO, O ₃ , NO ₂ , kategori pencemaran udara
KD-5	Kualitas Data	Kualitas data yang dibutuhkan harus minim <i>missing value</i> , bebas duplikasi, dan format konsisten
KD-6	Jumlah Data	Jumlah data keseluruhan pada <i>dataset</i> yaitu 4626.

Dataset ini digunakan dalam proses training dan testing pada proses pengembangan model. Data parameter pencemaran udara digunakan sebagai data input pada model, sementara data kategori pencemaran udara digunakan sebagai target *output* model. Semakin besar jumlah dan bervariasi data yang digunakan pada model, maka semakin baik kinerja model dalam mengenali pola data dan menghasilkan hasil prediksi yang akurat.

**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

• **Kebutuhan Model**

Fokus utama pada penelitian ini adalah pembangunan model klasifikasi pencemaran udara. Model ini akan diimplementasikan pada sistem aplikasi web sehingga dapat digunakan oleh *user*. Oleh karena itu, pengembangan model yang memiliki performa kinerja yang akurat sangat penting. Kebutuhan model dalam penelitian ini ditunjukkan pada Tabel 4.2

Tabel 4. 2 Tabel Kebutuhan Model

Kode	Nama	Deskripsi
KM-1	Kemampuan Prediksi	Model dapat melakukan klasifikasi pada kategori pencemaran udara berdasarkan input parameter ISPU
KM-2	Performa	Model memiliki akurasi, presisi, recall, dan F1-score yang optimal.
KM-3	Efisiensi	Model dapat melakukan proses klasifikasi dengan baik.

Apabila kebutuhan-kebutuhan ini dapat terpenuhi, maka diharapkan model dapat digunakan oleh pengguna akhir untuk membantu mempermudah proses pengolahan data kualitas udara atau data ISPU



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

4.1.2 Kebutuhan Pengembangan Aplikasi Web

Aplikasi berbasis web yang dibangun pada penelitian ini berfungsi sebagai media implementasi dari model XGBoost untuk klasifikasi pencemaran udara yang telah dibuat. Selain itu, aplikasi ini juga diharapkan bisa digunakan sebagai *platform* untuk melakukan pengolahan data ISPU secara ilmiah bagi pihak-pihak yang membutuhkan.

- Kebutuhan Fungsional

Untuk memastikan aplikasi web yang telah dibuat dapat digunakan secara optimal, maka harus dipastikan kebutuhan fungsional dari sistem harus terpenuhi. Fitur utama yang ada pada aplikasi ini meliputi kemampuan untuk mengunggah file data, melakukan proses klasifikasi otomatis, dan menampilkan hasil klasifikasi. Kebutuhan fungsional sistem tercantum pada Tabel 4.3

Tabel 4. 3 Tabel Kebutuhan Fungsional

Kode	Nama	Deskripsi
KF-1	Unggah Data	Pengguna dapat mengunggah file data ISPU yang berformat CSV pada halaman upload
KF-2	Klasifikasi Data	Sistem dapat melakukan klasifikasi otomatis menggunakan model XGBoost
KF-3	Visualisasi Data	Hasil dari klasifikasi ditampilkan dalam bentuk tabel dan grafik.



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

Kebutuhan fungsional ini diharapkan dapat membantu pengguna dalam melakukan pengolahan dan analisis data tingkat pencemaran udara.

- Kebutuhan Non Fungsional

Selain kebutuhan fungsional, ada juga kebutuhan non fungsional. Kebutuhan non fungsional berkaitan dengan aspek teknis dan perilaku sistem. Kebutuhan non fungsional sistem dijabarkan dalam Tabel 4.4

Tabel 4. 4 Tabel Kebutuhan Non Fungsional

Kode	Nama	Deskripsi
KNF-1	Kompatibilitas	Sistem dapat diakses melalui berbagai <i>web browser</i>
KNF-2	Responsivitas	Tampilan web reponsif

4.1.3 Kebutuhan Perangkat Keras & Perangkat Lunak

Agar dapat mencapai kinerja optimal dan sesuai dengan tujuan penelitian, dalam pembangunan sistem aplikasi web dan model XGBoost ini, direkomendasikan spesifikasi dari perangkat keras dan perangkat lunak tertentu yang sesuai. Namun, jika spesifikasi yang sesuai dengan kebutuhan sistem tidak tersedia maka dapat menyesuaikan dengan sumber daya yang dimiliki. Kebutuhan perangkat keras dan perangkat lunak dijelaskan pada Tabel 4.5 dan Tabel 4.6

Tabel 4. 5 Tabel Kebutuhan Perangkat Keras

Kode	Nama	Deskripsi
KPK-1	CPU	Minimal Intel Core i5 atau setara
KPK-2	RAM	Minimal 8 GB.
KPK-3	Penyimpanan	SSD minimal 256 GB.
KPK-4	GPU	GPU NVIDIA untuk mempercepat proses komputasi jika diperlukan.

Tabel 4. 6 Tabel Kebutuhan Perangkat Lunak

Kode	Nama	Deskripsi
KPL-1	Visual Studio Code	<i>Code editor</i> utama untuk pengembangan aplikasi
KPL-2	Python	Bahasa pemrograman untuk pembuatan model XGBoost
KPL-3	Tailwind CSS	<i>Framework</i> CSS untuk desain antarmuka yang responsif.
KPL-4	Laravel	<i>Framework</i> PHP untuk pengembangan <i>backend</i> aplikasi web
KPL-5	Jupyter Notebook	Lingkungan interaktif untuk melakukan pembuatan dan pengujian model.

**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

KPL-6	Python Virtual Environment	Lingkungan virtual untuk mengelola dependensi dan pustaka
-------	----------------------------	---

4.2 Perancangan Sistem

Setelah kebutuhan penelitian berhasil diidentifikasi, langkah berikutnya adalah merancang sistem berdasarkan kebutuhan tersebut. Perancangan ini bertujuan untuk menjadi panduan dalam proses pengembangan dan implementasi sistem. Dalam penelitian ini, perancangan sistem dibagi menjadi dua bagian utama, yaitu perancangan model XGBoost untuk klasifikasi pencemaran udara dan perancangan aplikasi web.

4.2.1 Perancangan Model

Proses perancangan model pada penelitian ini dimulai dengan pengumpulan dan pengolahan *dataset* Indeks Standar Pencemaran Udara (ISPU) Jakarta yang akan digunakan untuk melatih model XGBoost. Setelah data terkumpul, dan telah melalui tahap pembersihan data, maka langkah selanjutnya adalah tahap melatih model yang telah dibuat menggunakan algoritma XGBoost dengan tujuan memastikan bahwa model dapat berfungsi dengan baik sesuai dengan kebutuhan penelitian.

Proses perancangan model pada penelitian ini menerapkan metode CRISP-DM yang meliputi tahapan *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Pendekatan metode CRISP-DM ini bersifat fleksibel dimana setiap tahapan tidak selalu dijalankan secara berurutan namun bisa kembali ke tahap sebelumnya jika diperlukan, hal ini bertujuan untuk memastikan bahwa hasil yang diperoleh akurat dan sesuai dengan tujuan dari penelitian.



© 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

- Pengumpulan dan Pembuatan Dataset

Dataset yang digunakan dalam penelitian ini adalah data ISPU Jakarta periode 2010 – 2024 dengan jumlah keseluruhan data yaitu 4626 data yang diperoleh dari portal data *Satu Data Jakarta*. *Dataset* ini mencakup parameter pencemar udara yaitu PM₁₀, PM₂₅, SO₂, CO, O₃, NO₂, serta kolom kategori pencemaran udara yang menunjukkan tingkat kualitas udara, dalam konteks penelitian ini kategori pencemaran udara yang ada yaitu baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya.

Dataset akan dibagi menjadi dua bagian, yaitu *training set* dan *testing set*. *Training set* atau data latih digunakan untuk melatih model XGBoost dengan tujuan agar model pola nilai parameter pencemar udara yang ada pada dataset. Sedangkan data uji digunakan untuk mengevaluasi performa model.

Sebelum digunakan, dataset melalui proses pembersihan data (*data cleaning*) untuk menghilangkan nilai kosong (*missing values*), mengkonversi nilai N/A menjadi angka yang sesuai, merubah format kolom non numerik menjadi numerik, serta memilih data yang dijadikan fitur. Proses ini sangat penting tujuannya adalah agar model dapat memproses data dengan baik dan menghasilkan hasil klasifikasi yang akurat.

- Pelatihan Model

Pada tahap pelatihan model, algoritma yang digunakan adalah *Xtreme Gradient Boosting* (XGBoost), faktor yang menjadi dasar digunakannya algoritma XGBoost pada penelitian ini adalah keunggulan dalam akurasi, kemampuan dalam menangani data tabular, dan terdapat fitur *regularization* yang berfungsi untuk mengatasi *overfitting*.

Tahap pelatihan dilakukan dengan menggunakan bahasa pemrograman Python dan menggunakan pustaka pendukung seperti *xgboost*, *pandas*, *numpy*, dan *scikit-learn*. Selain itu untuk menangani persebaran kelas data pada fitur yang tidak merata atau *imbalanced data*.



Hak Cipta :

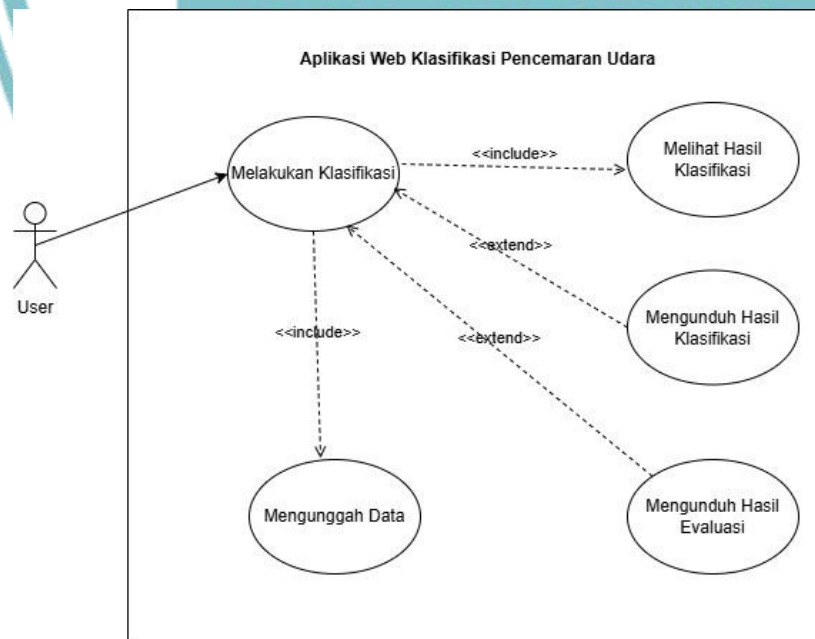
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

Selain itu tahap pengembangan model pada penelitian ini menggunakan metode *class weight* untuk menangani *imbalanced data*. Metode *class weight* bekerja dengan memberikan bobot yang lebih besar pada kelas dengan jumlah data yang sedikit, serta bobot yang lebih kecil pada kelas data yang banyak. Dengan cara ini, kesalahan prediksi pada kelas minoritas akan diberikan penalti yang lebih tinggi sehingga model terdorong untuk lebih memperhatikan kelas tersebut selama proses pelatihan.

Evaluasi performa model dilakukan dengan menghitung akurasi, presisi, recall, dan F1-score berdasarkan *confusion matrix*. Model dengan evaluasi terbaik kemudian akan digunakan untuk diintegrasikan dengan sistem aplikasi web untuk klasifikasi pencemaran udara. Perancangan Web

- Use Case Diagram

Use case diagram dalam penelitian ini dibangun berdasarkan kebutuhan fungsional web yang telah dijelaskan sebelumnya. Diagram ini menggambarkan interaksi antara aktor dengan sistem aplikasi web mengenai berbagai aksi yang dapat dilakukan oleh user di dalam sistem aplikasi web. Gambar 4.1 menampilkan *use case diagram* untuk web klasifikasi pencemaran udara.



Gambar 4. Use Case Diagram



© 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

Gambar 4.1 menunjukkan *use case diagram* untuk aplikasi web yang dibuat dalam penelitian ini. Pada *use case diagram* tersebut dapat dilihat bahwa terdapat satu aktor pada sistem, konteks *user* pada penelitian ini adalah semua pihak yang membutuhkan aplikasi web ini untuk melakukan pengolahan data ISPU secara ilmiah seperti mahasiswa, peneliti dan praktisi. *User* dapat mengunggah data ISPU, melihat hasil klasifikasi, serta mengunduh hasil klasifikasi.

- Activity Diagram

Secara sederhana *activity diagram* menggambarkan alur fungsional dari sebuah sistem. Namun, secara lengkap *activity diagram* mendefinisikan alur kerja fungsional sebuah sistem, urutan dari semua aktivitas fungsional, dimana titik dimulainya *workflow*, proses apa saja yang terjadi, sampai titik berakhir dari *workflow*.

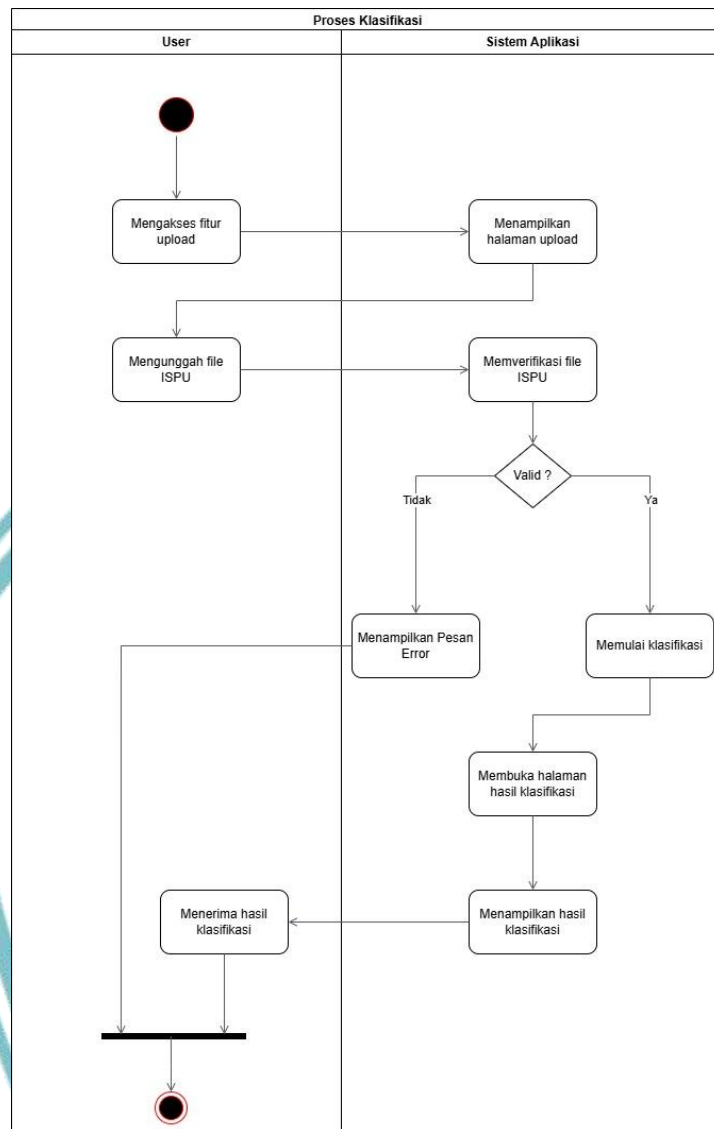




Hak Cipta :

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

• Proses Klasifikasi



Gambar 4. 1 Activity Diagram Proses Klasifikasi

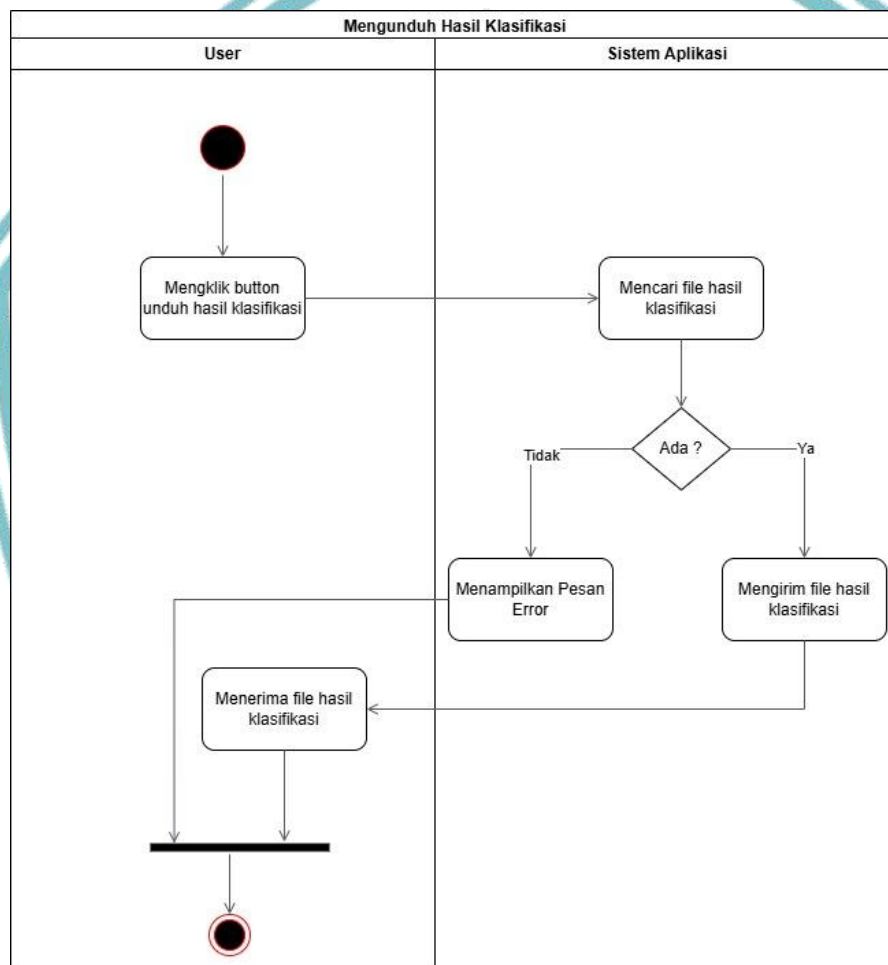


Hak Cipta :

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 4.2 menunjukkan proses klasifikasi pada aplikasi web klasifikasi pencemaran udara yang dimulai ketika *user* membuka halaman upload, lalu *user* memilih file input berupa file ISPU, setelah itu *user* menekan *button* untuk memulai proses klasifikasi. Setelah proses klasifikasi selesai dilakukan maka sistem akan *redirect* ke halaman hasil klasifikasi untuk menampilkan hasil dari proses klasifikasi yang telah dilakukan.

- Mengunduh Hasil Klasifikasi



Gambar 4. 2 *Activity Diagram* Mengunduh Hasil Klasifikasi

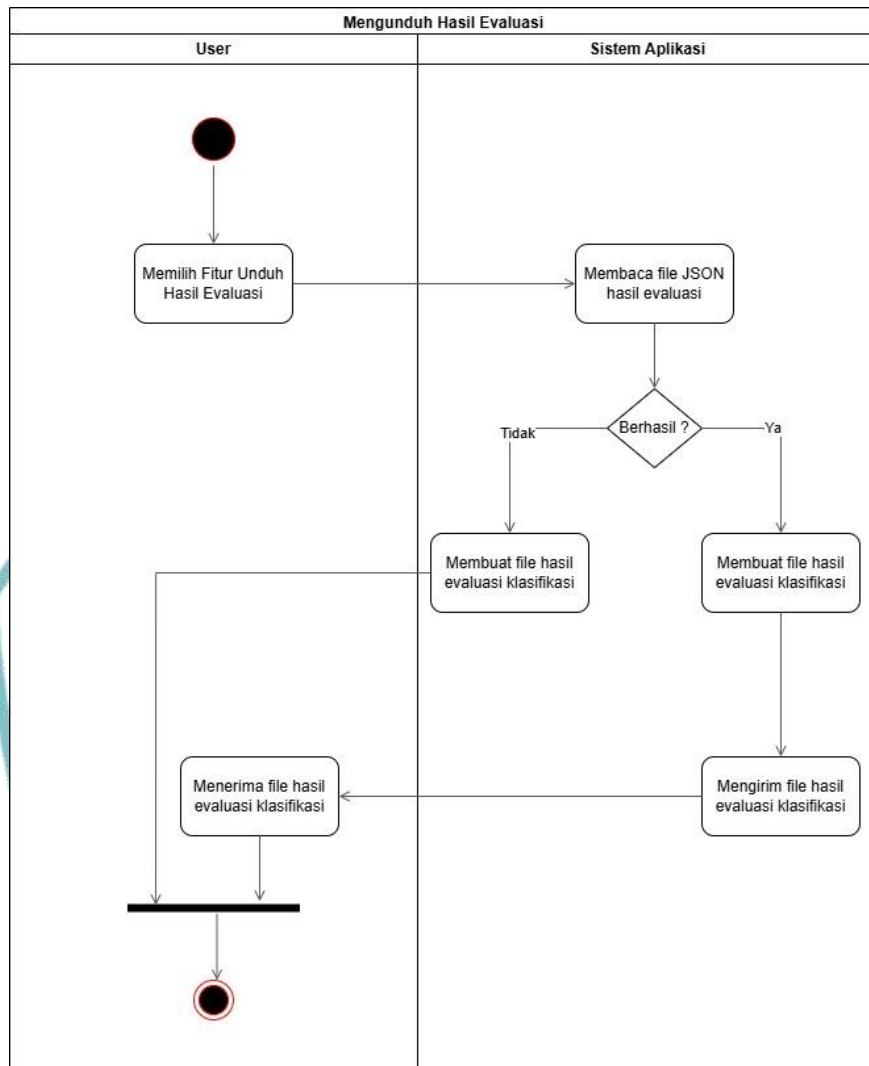
Gambar 4.3 menunjukkan proses mengunduh hasil klasifikasi dari aplikasi web klasifikasi pencemaran udara yang dimulai ketika *user* memilih fitur unduh hasil klasifikasi, lalu sistem akan mencari file hasil dari proses klasifikasi, selanjutnya sistem akan mengirimkan file hasil dari proses klasifikasi tersebut kepada user.



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

- Mengunduh Hasil Evaluasi



Gambar 4. 3 Activity Diagram Unduh Hasil Evaluasi

Gambar 4.4 menunjukkan mengenai gambaran alur dari proses unduh hasil evaluasi. Langkah tersebut dimulai dari *user* memilih fitur unduh hasil evaluasi, lalu sistem akan membuat file hasil evaluasi dari proses klasifikasi dan selanjutnya file tersebut dikirimkan ke oleh sistem kepada *user*.



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

• Wireframe Antarmuka

Wireframe merupakan kerangka awal antarmuka pengguna yang berfungsi untuk memberikan gambaran mengenai struktur halaman, seperti tata letak dari komponen-komponen utama seperti navigasi, button, serta isi atau konten dari halaman tanpa menggunakan desain atau gambaran yang lengkap dari halaman tersebut. Berikut ini *wireframe* antarmuka yang terdiri dari halaman upload dan halaman hasil klasifikasi.

• Halaman Upload

Gambar 4. 4 *Wireframe* Halaman Upload

Wireframe halaman upload menampilkan rancangan tampilan halaman upload, dimana terdapat elemen berupa form untuk mengupload file data ISPU dan terdapat elemen *button* untuk mengupload file data ISPU dan memulai proses klasifikasi.



© 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

- Halaman Hasil Klasifikasi

Halaman Upload

Upload

Hasil Klasifikasi

ISPU Jakarta

Hasil Klasifikasi

Penjelasan Kategori & Parameter ISPU

Download Hasil Klasifikasi (CSV)

Download Hasil Evaluasi (PDF)

Tanggal | PM10 | PM25 | SO2 | NO2 | CO | O3 | Kategori Aktual | Kategori Hasil

2010-01-01 | 60.0 | 94.0 | 4.0 | 14.0 | 73.0 | 27.0 | Sedang | Sedang

2010-01-02 | 32.0 | 94.0 | 2.0 | 15.0 | 20.0 | 33.0 | Baik | Baik

2010-01-03 | 22.0 | 94.0 | 3.0 | 16.0 | 15.0 | 20.0 | Sedang | Sedang

2010-01-04 | 25.0 | 94.0 | 4.0 | 19.0 | 17.0 | 15.0 | Baik | Baik

2010-01-05 | 24.0 | 94.0 | 3.0 | 13.0 | 24.0 | 17.0 | Sedang | Baik

2010-01-06 | 55.0 | 94.0 | 5.0 | 16.0 | 32.0 | 23.0 | Baik | Sedang

2010-01-07 | 60.0 | 94.0 | 4.0 | 13.0 | 28.0 | 25.0 | Sedang | Sedang

Legenda Kategori ISPU

☐ Baik

☐ Sedang

☐ Tidak Sehat

☐ Sangat Tidak Sehat

☐ Berbahaya

Visualisasi Jumlah Kategori

Tips Kesehatan Berdasarkan Kategori Dominan

Gambar 4. 5 *Wireframe* Halaman Hasil Klasifikasi

Wireframe halaman hasil klasifikasi menampilkan rancangan halaman hasil klasifikasi, yang terdiri dari berbagai elemen, terutama yaitu tabel yang menampilkan hasil dari proses klasifikasi, keterangan mengenai warna dari label kategori, diagram yang menampilkan jumlah data label kategori, *button* untuk menampilkan informasi tentang ISPU serta *button* untuk mengunduh hasil klasifikasi.



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

4.3 Implementasi Sistem

Berbagai rancangan yang telah dibuat pada tahap sebelumnya, selanjutnya akan direalisasikan pada tahap implementasi sistem. Pada penelitian ini, implementasi sistem terbagi menjadi dua, yaitu implementasi model *machine learning* dan implementasi sistem aplikasi km,web. Pembagian pada tahap implementasi ini dilakukan dengan tujuan karena proses pengembangan web hanya dapat dilakukan apabila model *machine learning* telah selesai dibuat. Tujuan lainnya juga, untuk memastikan bahwa model *machine learning* yang dibuat merupakan model yang terbaik ketika diimplementasikan pada sistem web.

4.3.1 Implementasi Model *Machine Learning*

Seluruh proses implementasi model *machine learning* pada penelitian ini dijelaskan pada langkah-langkah berikut sesuai dengan konsep metode CRISP-DM.

- *Business Understanding*

Tujuan dari implementasi model klasifikasi ini adalah untuk mengembangkan sistem klasifikasi yang mampu memetakan kategori kualitas udara berdasarkan data ISPU (Indeks Standar Pencemar Udara). Data ini mencakup parameter pencemar utama, seperti PM₁₀, PM_{2.5}, SO₂, CO, O₃, dan NO₂. Dengan memanfaatkan algoritma XGBoost, diharapkan sistem mampu menghasilkan klasifikasi yang lebih akurat. Secara praktis, dengan adanya sistem klasifikasi ini dapat digunakan untuk mendukung pengambilan keputusan terkait kondisi udara di Jakarta.



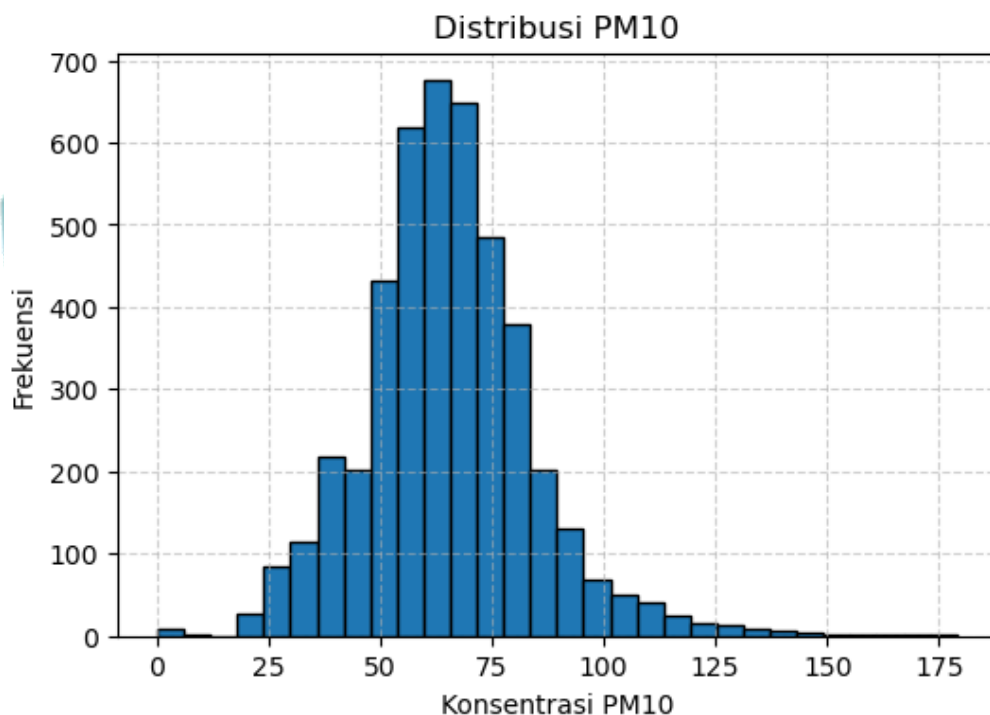
© Hak Cipta milik Politeknik Negeri Jakarta

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

• Data Understanding

Dataset yang digunakan dalam penelitian ini adalah data ISPU yang diperoleh dari *platform* portal data *Satu Data Jakarta*, yang merupakan salah satu repository publik yang banyak digunakan untuk kebutuhan penelitian dan pengembangan model *machine learning*. *Dataset* ini berisi data Indeks Standar Pencemar Udara (ISPU) yang dikumpulkan dari beberapa stasiun pemantauan kualitas udara di Jakarta. Variabel yang tersedia dalam *dataset* meliputi parameter pencemar udara seperti PM_{10} , $PM_{2.5}$, SO_2 , CO , O_3 , dan NO_2 , serta atribut tambahan berupa periode pengukuran, tanggal, dan lokasi stasiun.



Gambar 4. 6 Distribusi Data Pencemar

Dataset tersebut dipilih karena struktur datanya telah disusun secara rapi, memiliki label dengan kategori kualitas udara (baik, sedang, tidak sehat, sangat tidak sehat, berbahaya), serta dapat digunakan secara langsung untuk proses pelatihan dan pengujian model klasifikasi. *Dataset* ini mencakup 4626 data status kualitas udara per harinya dari tahun 2010 sampai tahun 2024.



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

Data diperoleh melalui pemindaian menggunakan alat *monitoring* kualitas udara, yang menghasilkan data harian status pencemaran udara. Data harian tersebut berisi informasi mengenai tanggal, lokasi stasiun, nilai pencemar udara, nilai pencemar terbesar, serta kategori status kondisi udara di wilayah tersebut.

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

```
print("=== Informasi Dataset ===")
print(df.info())
```

```
=== Informasi Dataset ===
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4626 entries, 0 to 4625
Data columns (total 11 columns):
#   Column      Non-Null Count  Dtype
---  -
0   tanggal     4626 non-null   object
1   stasiun     4626 non-null   object
2   pm10        4466 non-null   float64
3   pm25        723 non-null    float64
4   so2         4607 non-null   float64
5   co          4618 non-null   float64
6   o3          4621 non-null   float64
7   no2         4618 non-null   float64
8   max         4626 non-null   float64
9   critical    4625 non-null   object
10  kategori    4626 non-null   object
dtypes: float64(7), object(4)
memory usage: 397.7+ KB
None
```

```
df.shape
```

```
(4626, 11)
```

Gambar 4. 7 Informasi Umum *Dataset*

Struktur *dataset* yang digunakan pada penelitian ini terdiri dari 11 kolom. Kolom-kolom yang ada pada *dataset* terbagi menjadi dua jenis berdasarkan tipe datanya, yaitu kolom yang bertipe *object*, seperti : tanggal yang menunjukkan data pencatatan data kualitas udara, stasiun menunjukkan lokasi atau titik pemantauan kualitas udara, *critical* menunjukkan jenis polutan yang menjadi parameter dominan, *kategori* menunjukkan status dari kualitas udara berdasarkan ISPU. Ada juga kolom yang bertipe numerik seperti kolom parameter pencemar udara yaitu PM₁₀, PM_{2.5}, SO₂, CO, O₃, NO₂, dan kolom *max* yang menunjukkan nilai maksimum jenis polutan pada periode tertentu.



© 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

```
print("\n=== Jumlah Data Non Null pada Setiap Kolom ===")
print(df.count())
```

```
=== Jumlah Data Non Null pada Setiap Kolom ===
tanggal      4626
stasiun      4626
pm10         4466
pm25         723
so2          4607
co           4618
o3           4621
no2          4618
max          4626
critical     4625
kategori     4626
dtype: int64
```

Gambar 4. 8 Informasi Data *Non Null*

Berikut merupakan informasi jumlah data pada setiap kolom yang ada pada *dataset*. *Dataset* yang dipakai juga memiliki banyak nilai yang kosong yang paling banyak jumlahnya ada pada kolom polutan PM_{2.5}.

```
print("\n=== Total Missing Values ===")
print(df.isnull().sum().sum())
```

```
=== Total Missing Values ===
4104
```

```
print(df['pm25'].isnull().sum())
```

```
3903
```

Gambar 4. 9 Total *Missing Value*

Jumlah keseluruhan nilai yang kosong pada dataset sebanyak 4104 data, dan seperti yang sudah disebutkan sebelumnya data polutan PM_{2.5} menjadi kolom yang memiliki nilai yang hilang terbanyak yaitu 3903 data yang hilang.



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 mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

```
print("\n=== Statistik Deskriptif ===")
print(df.describe())
```

```
=== Statistik Deskriptif ===
```

	pm10	pm25	so2	co	o3
count	4466.000000	723.000000	4607.000000	4618.000000	4621.000000
mean	65.086879	94.374827	28.183851	30.263534	86.586670
std	19.243150	27.427553	13.822107	13.640344	50.295879
min	0.000000	23.000000	0.000000	0.000000	0.000000
25%	54.000000	76.000000	18.000000	21.000000	51.000000
50%	64.000000	94.000000	27.000000	29.000000	78.000000
75%	75.000000	111.000000	34.000000	37.000000	110.000000
max	179.000000	287.000000	126.000000	134.000000	314.000000

	no2	max
count	4618.000000	4626.000000
mean	18.120832	99.432123
std	8.451605	43.284775
min	0.000000	0.000000
25%	13.000000	70.000000
50%	17.000000	91.000000
75%	21.000000	118.000000
max	134.000000	314.000000

```
print("=== Distribusi Kelas ===")
distribution = df['kategori'].value_counts()
print(distribution)
```

```
=== Distribusi Kelas ===
```

kategori	
SEDANG	2574
TIDAK SEHAT	1568
BAIK	280
SANGAT TIDAK SEHAT	202
BERBAHAYA	1
TIDAK ADA DATA	1

Name: count, dtype: int64

Gambar 4. 10 Informasi Distribusi Kelas

Distribusi kelas pada *dataset* ISPU menunjukkan adanya ketidakseimbangan (*imbalanced class*) di mana terdapat beberapa kelas kategori memiliki selisih jumlah data yang sangat signifikan. Persentase jumlah data setiap kelas pada *dataset* ditunjukkan pada Tabel 4.7 berikut.



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

Tabel 4. 7 Distribusi Kelas

Kelas Kategori	Jumlah Persentase
Baik	6%
Sedang	555,6%
Tidak Sehat	33,9%
Sangat Tidak Sehat	4%
Berbahaya	0,0216%

Berdasarkan tabel tersebut bisa dilihat bahwa jumlah data kategori “sedang” memiliki jumlah data yang paling banyak, sedangkan kategori “berbahaya” merupakan kategori dengan jumlah data yang paling sedikit, yaitu hanya terdapat 1 data pada dataset. Tentunya perlu dilakukan penanganan untuk mengatasi distribusi kelas yang tidak seimbang agar proses klasifikasi dapat berjalan dengan baik.

Salah satu bagian penting dalam tahap ini adalah mendeteksi adanya *outlier*, yaitu nilai ekstrim yang menyimpang jauh dari sebaran data mayoritas.

```
def detect_outliers_iqr(df, column):
    """Fungsi untuk mendeteksi outlier menggunakan metode IQR."""
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = df[(df[column] < lower_bound) | (df[column] > upper_bound)]
    return outliers

# Visualisasi boxplot sebelum penanganan
plt.figure(figsize=(12, 6))
df[numerical_columns].boxplot()
plt.title("Boxplot Sebelum Penanganan Outlier")
plt.show()
```

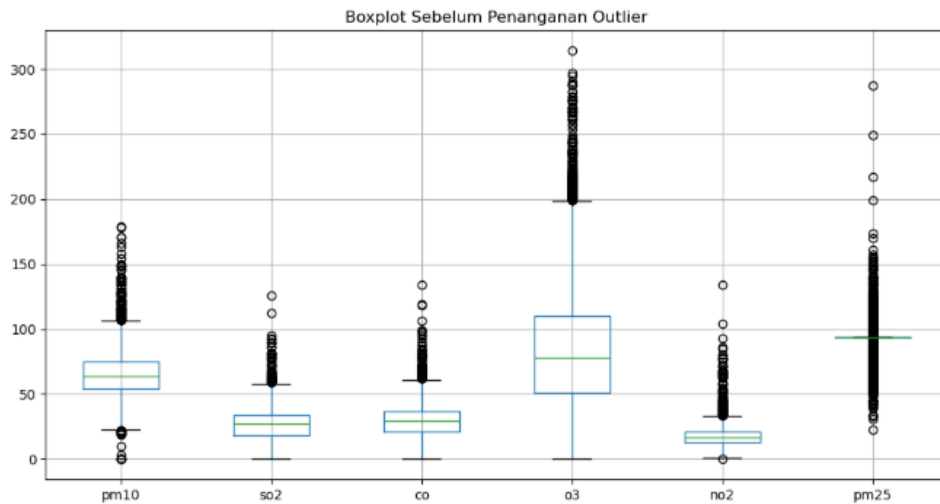
Gambar 4. 11 Deteksi Outlier

Untuk mendeteksi *outlier* digunakan metode *Interquartile Range* (IQR). Metode ini bekerja dengan menghitung nilai IQR yaitu dengan menghitung selisih kuartil pertama (Q1) dan kuartil ketiga (Q3). Nilai batas bawah ditentukan dengan rumus $Q1 - 1.5 \times IQR$, sedangkan untuk nilai batas atas didapatkan dengan rumus $Q3 - 1.5 \times IQR$.



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. 12 *Boxplot Outlier*

Hasil deteksi menunjukkan adanya sejumlah data ekstrem pada parameter tertentu, misalnya PM10 dan PM2.5, yang memiliki nilai konsentrasi jauh lebih tinggi dibandingkan rentang normal. Keberadaan outlier yang cukup signifikan ini berpotensi mempengaruhi hasil analisis dan kinerja model klasifikasi, sehingga diperlukan langkah penanganan *outlier* agar distribusi data lebih stabil dan model dapat menghasilkan prediksi yang lebih akurat.

- Data Preparation

Berdasarkan pemaparan informasi tentang dataset pada tahap sebelumnya, terdapat data anomali yang ada pada kolom kategori, yaitu “tidak ada data”. Pada tahap ini dilakukan pembersihan data, khususnya pada data kategori. Selain itu data kelas kategori “berbahaya” hanya ada 1 entri, tentunya ini diperlukan penanganan untuk menjaga performa dari model.

```
def normalize_labels(y):
    y = y.astype(str).str.strip().str.upper()
    y = y.replace({
        'TIDAK ADA DATA': np.nan,
        'BERBAHAYA': 'SANGAT TIDAK SEHAT'
    })
    return y

df['kategori'] = normalize_labels(df['kategori'])
df = df.dropna(subset=['kategori'])
```

Gambar 4. 13 Tahap Pembersihan Data



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

Pada tahap ini dilakukan proses normalisasi label target dan pembersihan nilai numerik yang tidak logis untuk memastikan kualitas serta konsistensi data sebelum digunakan dalam pelatihan model klasifikasi. Tahapan ini bertujuan agar model yang dihasilkan tidak terpengaruh oleh kesalahan penulisan, ketidakteraturan label, maupun nilai yang tidak masuk akal secara ilmiah.

Langkah pertama, adalah melakukan normalisasi terhadap label target pada kolom *kategori*. Proses ini meliputi perubahan seluruh nilai menjadi menjadi huruf besar dan penghapusan spasi yang tidak diperlukan agar data lebih seragam, Selain itu, label yang memiliki makna serupa seperti “berbahaya” disamakan menjadi “sangat tidak sehat” untuk mengurangi ketidakseimbangan kelas akibat jumlah data yang sangat sedikit pada kategori tertentu. Label “tidak ada data ” diubah menjadi nilai kosong (*NaN*) karena tidak merepresentasikan kondisi kualitas udara yang valid. Setelah dilakukan normalisasi, seluruh baris data yang tidak memiliki label kategori dihapus agar hanya tersisa data dengan target yang benar dan dapat digunakan untuk proses pelatihan model.

```
numerical_columns = ['pm10', 'so2', 'co', 'o3', 'no2', 'pm25']
for col in numerical_columns:
    df[col] = pd.to_numeric(df[col], errors='coerce')
    df = df[df[col].isna() | (df[col] >= 0)]
```

Gambar 4. 14 Tahap Pembersihan Nilai Negatif

Tahap berikutnya adalah pembersihan nilai negatif pada kolom numerik yang berisi parameter konsentrasi polutan, yaitu PM₁₀, SO₂, CO, O₃, NO₂, dan PM₂₅. Seluruh nilai pada kolom tersebut dikonservasi menjadi tipe data numerik agar data yang masih berbentuk teks dapat dikenali sebagai angka. Jika terdapat nilai yang tidak dapat dikonversi, maka nilai tersebut diubah menjadi *NaN* untuk ditangani pada tahap imputasi berikutnya, Kemudian, semua baris data yang memiliki nilai negatif dihapus karena secara fisik, konsentrasi polutan udara tidak mungkin bernilai negatif. Nilai negatif biasanya disebabkan oleh kesalahan alat ukur atau kesalahan pencatatan data.



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

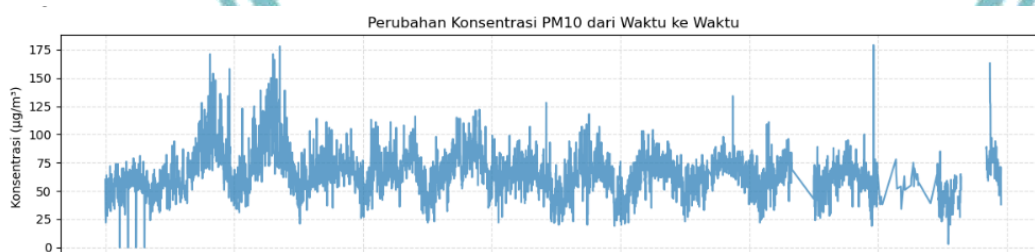
```
le_target = LabelEncoder()
df['kategori'] = le_target.fit_transform(df['kategori'])
```

Gambar 4. 15 Tahap *Encoding Label*

Tahap berikutnya adalah melakukan encoding pada label target menggunakan metode *Label Encoding*. Langkah ini bertujuan untuk mengubah data kategorikal pada kolom *kategori* menjadi bentuk numerik agar dapat diproses oleh algoritma *machine learning* XGBoost, yang secara *default* hanya dapat menerima masukan dalam bentuk numerik. Setiap label unik pada kolom *kategori* diberikan kode numerik secara berurutan, label “baik” menjadi 0, “sedang” menjadi 1, “tidak sehat” menjadi 2, “sangat tidak sehat” menjadi 3. Proses ini mengonversi format dari setiap kategori agar dapat dikenali oleh model *machine learning*.

Proses pembuatan model *machine learning* untuk melakukan klasifikasi terhadap data ISPU pada penelitian ini tidak menerapkan penanganan *outlier* yang ditemukan di dalam *dataset*. Hal itu dilakukan karena *outlier* yang ditemukan merupakan hal yang terjadi secara alami. Maksudnya adalah nilai-nilai ekstrem yang muncul bisa saja disebabkan oleh peningkatan kondisi pencemar udara yang disebabkan karena jumlah kendaraan, aktivitas industri, maupun faktor cuaca.

Untuk mendukung pernyataan tersebut, dilakukan analisis terhadap nilai-nilai ekstrem tersebut, yaitu dengan menerapkan teknik analisis temporal (*time series consistency*) dan teknik analisis korelasi antar parameter (*cross-check parameter*). Hal ini bertujuan untuk membuktikan apakah *outlier* yang ada merupakan terjadi secara alami atau merupakan *error* dari alat pemantau pencemaran udara, yang mana hal ini akan menentukan cara dalam menangani *outlier* yang ada.



Gambar 4. 16 Analisis Temporal PM₁₀



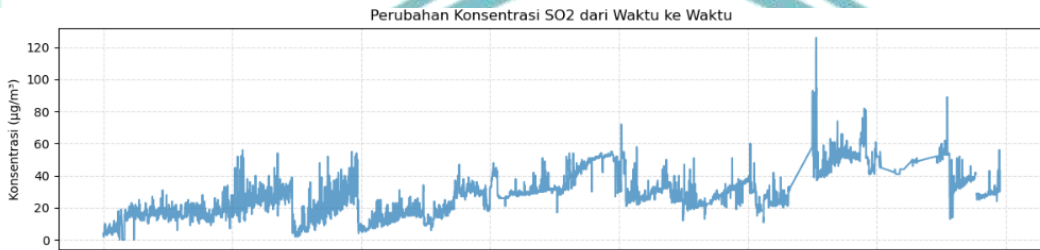
© Hak Cipta milik Politeknik Negeri Jakarta

Hak Cipta :

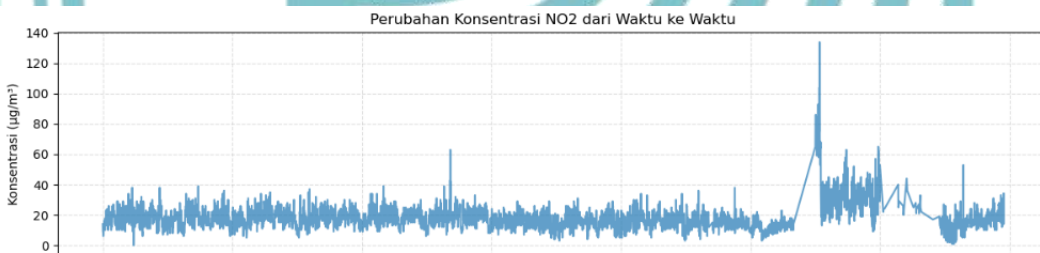
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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



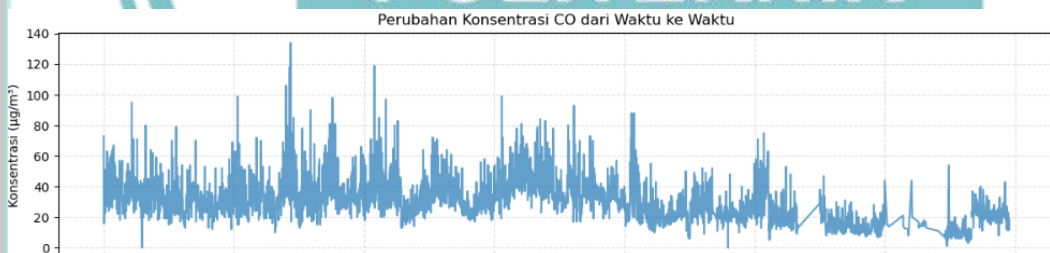
Gambar 4. 17 Analisis Temporal PM₂₅



Gambar 4. 18 Analisis Temporal SO₂



Gambar 4. 19 Analisis Temporal NO₂

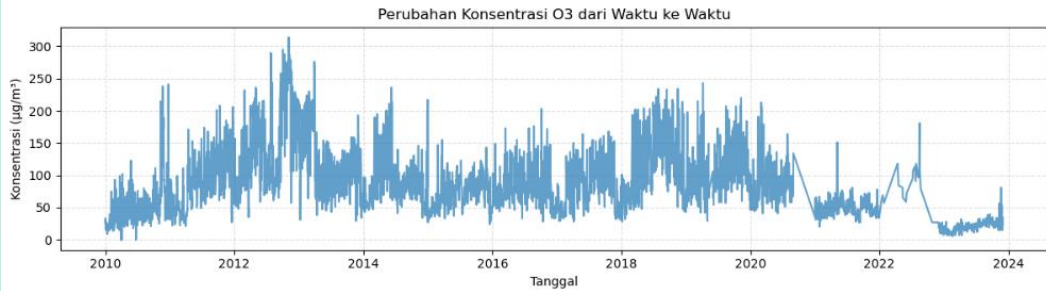


Gambar 4. 20 Analisis Temporal SO₂



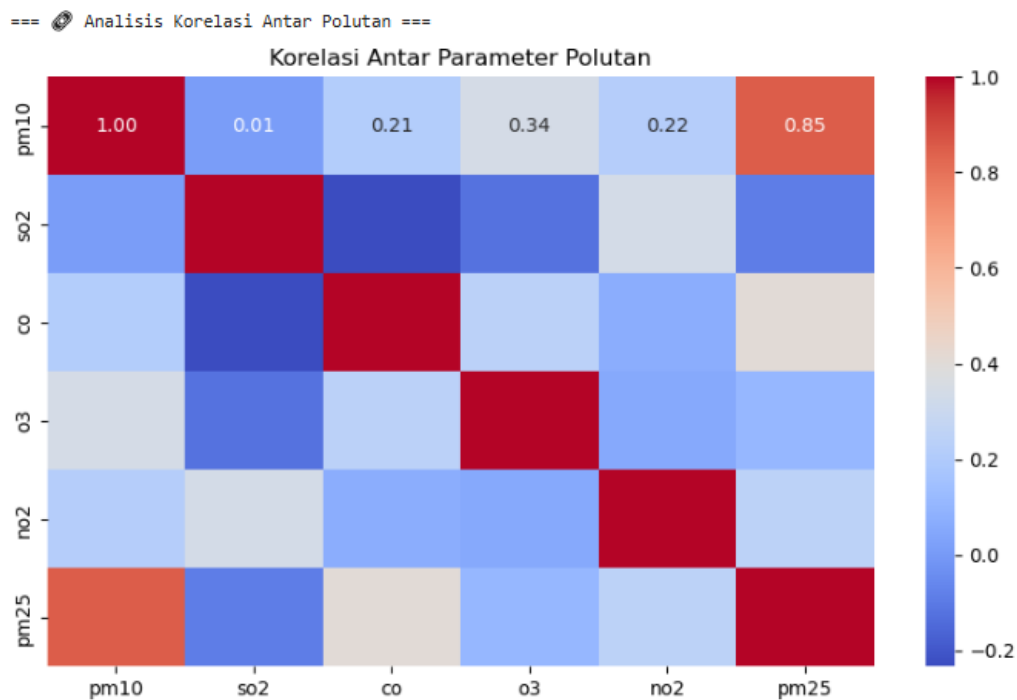
Hak Cipta :

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 O₃

Analisis temporal dilakukan untuk meninjau pola perubahan konsentrasi parameter pencemar udara dari waktu ke waktu. Tujuan dilakukan analisis ini adalah untuk mengidentifikasi tren jangka panjang, variasi musiman, serta potensi nilai ekstrim (*outlier*) yang dapat muncul akibat *error* dari alat pemantau kualitas udara atau fenomena alami tertentu. Secara keseluruhan, pola temporal pada masing-masing parameter pencemar udara yang dihasilkan menunjukkan bahwa sebagian besar fluktuasi konsentrasi bersifat musiman dan alami, bukan merupakan *error* dari alat pemantau kualitas udara.



Gambar 4. 22 Analisis Korelasi Antar Polutan



© 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

Analisis korelasi antar parameter polutan dilakukan untuk mengetahui sejauh mana hubungan antar parameter saling mempengaruhi satu sama lain dalam membentuk kondisi kualitas udara. Berdasarkan hasil dari analisis tersebut dapat disimpulkan bahwa polutan PM10 dan PM2.5 memiliki keterkaitan paling kuat, sehingga keduanya dapat dijadikan indikator utama untuk menilai tingkat pencemaran partikulat di udara. Kedua polutan tersebut memiliki nilai korelasi sebesar 0.85, keterkaitan yang tinggi ini memperkuat dugaan bahwa peningkatan konsentrasi salah satu parameter polutan diikuti oleh peningkatan parameter lainnya secara bersamaan.

Fenomena ini menggambarkan kondisi lingkungan yang saling berpengaruh secara alami, bukan akibat kesalahan dari alat. Sedangkan untuk polutan gas seperti CO, NO₂, O₃, dan SO₂, memiliki hubungan yang lebih lemah satu sama lain, hal itu menandakan bahwa setiap jenis polutan memiliki sumber dan dinamika yang berbeda.

Berdasarkan pertimbangan dari hasil dua analisis yang telah dilakukan, disimpulkan bahwa outlier yang ada pada *dataset* bukan merupakan anomali data yang harus dihapus, melainkan merupakan fenomena yang terjadi secara alami yang bisa disebabkan oleh berbagai kejadian di lingkungan di wilayah pengamatan. Oleh karena itu, dalam penelitian ini tidak menerapkan proses penanganan *outlier*, hal itu bertujuan agar model yang dibuat mampu menampilkan hasil yang selaras dengan keadaan yang sebenarnya di lapangan dan tidak kehilangan informasi penting mengenai fenomena ekstrem yang relevan terhadap proses klasifikasi tingkat pencemaran udara.



- Modeling

Pada tahap ini, data yang telah melalui proses pembersihan (*data cleaning*) dan persiapan (*data preparation*) selanjutnya digunakan untuk membangun model klasifikasi pencemaran udara. Sebelum proses pelatihan model dilakukan, dataset dibagikan menjadi dua bagian utama, yaitu data latih (*training set*) dan data uji (*testing set*). Penelitian ini menggunakan metode *Stratified Train-Test Split* dengan perbandingan 80% data latih dan 20% data uji. Metode ini dipilih karena mampu menjaga proporsi distribusi setiap kelas pada data latih dan data uji agar tetap konsisten dengan distribusi data asli. Pendekatan ini menjadi penting dalam penelitian yang menggunakan dataset dengan distribusi kelas yang tidak seimbang (*imbalanced dataset*), karena memastikan bahwa kelas minoritas tetap memiliki representasi yang proporsional pada setiap partisi data.

```
X = df[numerical_columns].copy()
y = df["kategori"].copy()

X_train, X_test, y_train, y_test = train_test_split(
    X,
    y,
    test_size=0.20,
    stratify=y,
    random_state=42
)
```

Gambar 4. 23 Tahap *Data Splitting*

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



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

Tahap selanjutnya adalah penanganan *imbalanced dataset*, karena berdasarkan informasi umum dari *dataset* terlihat bahwa terdapat ketidakseimbangan pada distribusi kelas. Oleh karena itu, perlu dilakukan penerapan teknik *resampling* yang tepat sesuai dengan karakteristik data pada *dataset*. Dalam penelitian ini dilakukan tiga pendekatan berbeda untuk menangani ketidakseimbangan data pada model klasifikasi, yaitu tanpa penanganan *imbalanced data*, dengan penerapan *class weight*, dan menggunakan metode SMOTE (*Synthetic Minority Over-Sampling Technique*). Ketiga pendekatan ini bertujuan untuk meningkatkan performa model, khususnya pada kelas minoritas yang jumlah datanya relatif sangat sedikit dibandingkan dengan kelas lainnya.

```
random_search.fit(
    X_train,
    y_train,
)

pipeline = random_search.best_estimator_
```

Gambar 4. 24 Tanpa Penerapan Metode *Imbalance Data Handling*

Pada skenario pertama, model XGBoost dilatih menggunakan data hasil *Stratified Train-Test Split* tanpa menerapkan teknik penanganan ketidakseimbangan data. Seluruh data latih digunakan secara langsung untuk proses pelatihan model.

```
classes = np.unique(y_train)

weights = compute_class_weight(
    class_weight="balanced",
    classes=classes,
    y=y_train
)

class_weights = dict(zip(classes, weights))

sample_weight = np.array([
    class_weights[y]
    for y in y_train
])
```

Gambar 4. 25 Penerapan Teknik Class Weight



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

Pada skenario kedua diterapkan metode *Class Weight* menggunakan fungsi `compute_class_weight()` yang disediakan disediakan oleh pustaka *Scikit-Learn* dengan parameter `class_weight = "balanced"`. Cara kerja dari metode *Class Weight* pada penelitian ini adalah memberikan bobot pada seluruh label kelas yang terdapat pada data latih (`y_train`). Bobot setiap kelas dihitung secara otomatis oleh fungsi `compute_class_weight()` berdasarkan distribusi jumlah data pada masing-masing kelas. Ketika parameter `class_weight = "balanced"` digunakan, maka *library Scikit-Learn* menerapkan rumus dari *Class Weight* yaitu :

$$\omega_i = \frac{N}{K \times n_i}$$

Tabel 4. 8 Tabel Distribusi Kelas Data

Kategori	Jumlah Data
BAIK	224
SANGAT TIDAK SEHAT	161
SEDANG	2059
TIDAK SEHAT	1256

Berdasarkan tabel tersebut, jumlah seluruh data *train* yaitu 3700, jadi $N = 3700$, jumlah data pada kelas BAIK yaitu 224, $n_i = 224$, jumlah data pada kelas SANGAT TIDAK SEHAT 161, maka $n_i = 161$, jumlah data pada kelas SEDANG 2059, maka $n_i = 2059$, dan jumlah data pada kelas TIDAK SEHAT 1256, maka $n_i = 1256$, sedangkan jumlah kelas ada 4, maka $K = 4$. Berikut perhitungan bobot setiap kelas pada penelitian ini :

- Bobot Kelas BAIK

$$\omega_i = \frac{3700}{4 \times 224}$$

$$\omega_i = 4.13$$

- Bobot Kelas SANGAT TIDAK SEHAT

$$\omega_i = \frac{3700}{4 \times 161}$$

$$\omega_i = 5.45$$



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

- Bobot Kelas SEDANG

$$\omega_i = \frac{3700}{4 \times 2059}$$

$$\omega_i = 0.45$$

- Bobot Kelas TIDAK SEHAT

$$\omega_i = \frac{3700}{4 \times 1256}$$

$$\omega_i = 0.47$$

Berdasarkan perhitungan bobot kelas pada metode *Class Weight* bisa dilihat bahwa semakin sedikit jumlah data suatu kelas, maka semakin besar bobot yang dihasilkan. Hal tersebut menyebabkan kesalahan prediksi pada kelas minoritas akan memberikan kontribusi yang lebih besar terhadap fungsi kerugian (*loss function*) dibandingkan kesalahan pada kelas mayoritas. Penggunaan `compute_class_weight()` pada penelitian ini dipilih karena fungsi tersebut merupakan implementasi resmi dari *library Scikit-Learn* untuk menghitung bobot kelas secara otomatis berdasarkan distribusi data latih menggunakan skema *balanced*. Pendekatan ini banyak digunakan dalam penelitian klasifikasi dengan data tidak seimbang karena mampu meningkatkan perhatian model terhadap kelas minoritas tanpa mengubah distribusi data asli atau menghasilkan data sintetis seperti pada metode *SMOTE*.



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

```
imputer = SimpleImputer(strategy="median")

X_train = imputer.fit_transform(X_train)

X_test = imputer.transform(X_test)

smote = SMOTE(
    random_state=42,
    k_neighbors=5
)

X_train_smote, y_train_smote = smote.fit_resample(
    X_train,
    y_train
)
```

Gambar 4. 26 Penerapan Teknik *SMOTE*

Pada skenario ketiga diterapkan metode *SMOTE* (*Synthetic Minority Oversampling Technique*) untuk menghasilkan sampel sintesis pada kelas minoritas sehingga distribusi data *train* menjadi lebih seimbang sebelum dilakukan proses pelatihan model *XGBoost*.

```
param_dist = {

    "classifier__n_estimators": [100,200,300,400],

    "classifier__learning_rate": [0.01,0.03,0.05,0.1],

    "classifier__max_depth": [3,4,5,6,7],

    "classifier__subsample": [0.7,0.8,0.9,1.0],

    "classifier__colsample_bytree": [0.7,0.8,0.9,1.0],

    "classifier__gamma": [0,0.1,0.3,0.5],

    "classifier__min_child_weight": [1,3,5],

    "classifier__reg_alpha": [0,0.1,0.5,1],

    "classifier__reg_lambda": [1,1.5,2]

}
```

Gambar 4. 27 Tahap *Hyperparameter Tuning*



© 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

Tahap *hyperparameter tuning* dilakukan untuk memperoleh kombinasi nilai *hyperparameter* terbaik pada algoritma *XGBoost* sehingga model mampu menghasilkan performa klasifikasi yang optimal. Nilai *hyperparameter* merupakan parameter yang harus ditentukan sebelum proses pelatihan dimulai. Nilai *hyperparameter* akan memengaruhi kompleksitas model, kemampuan generalisasi, serta risiko terjadinya *overfitting*.

```
cv = StratifiedKFold(
    n_splits=5,
    shuffle=True,
    random_state=42
)
```

Gambar 4. 28 Stratified K-Fold Cross Validation

Metode ini membagi data *train* menjadi lima bagian (*fold*) dengan mempertahankan proporsi setiap kelas pada masing-masing *fold*. Pada setiap iterasi, empat *fold* digunakan sebagai data *training* sedangkan satu *fold* digunakan sebagai data validasi. Proses tersebut dilakukan sebanyak lima kali sehingga setiap *fold* pernah menjadi data validasi. Penggunaan metode ini bertujuan menghasilkan evaluasi yang lebih stabil pada dataset yang memiliki distribusi kelas yang tidak seimbang.

**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

```
random_search = RandomizedSearchCV(

    estimator=pipeline,

    param_distributions=param_dist,

    n_iter=30,

    scoring="f1_macro",

    cv=cv,

    verbose=2,

    n_jobs=-1,

    random_state=42

)
```

Gambar 4. 29 RandomizedSearchCV

Setelah pencarian *hyperparameter* ditentukan, proses pencarian kombinasi terbaik dilakukan menggunakan *RandomizedSearchCV*. Pada penelitian ini digunakan 30 kombinasi *hyperparameter* yang dipilih secara acak dari seluruh pencarian. Setiap kombinasi dievaluasi menggunakan 5 *fold Stratified Cross Validation* dan kinerja setiap kombinasi dinilai menggunakan metrik *Macro f1-Score*. Kombinasi yang menghasilkan nilai *Macro f1-Score* tertinggi akan dipilih sebagai model terbaik.

- Evaluation

Evaluasi model dilakukan untuk menilai seberapa baik model klasifikasi yang dibangun mampu melakukan prediksi sesuai dengan label aktual pada data uji.



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

```

y_pred = pipeline.predict(X_test)
y_pred_proba = pipeline.predict_proba(X_test)

accuracy = accuracy_score(y_test, y_pred)
print(f"\n♦ Accuracy: {accuracy:.4f}")

report = classification_report(
    y_test, y_pred, target_names=le_target.classes_, output_dict=True
)
report_df = pd.DataFrame(report).transpose()

print("\n=== 🔍 Classification Report ===")
print(report_df)

cm = confusion_matrix(y_test, y_pred)
cm_df = pd.DataFrame(
    cm, index=le_target.classes_, columns=le_target.classes_
)

```

Gambar 4. 30 Tahap Evaluasi

Setelah model XGBoost selesai dilatih menggunakan parameter terbaik hasil proses *hyperparameter tuning*, tahap selanjutnya adalah melakukan evaluasi terhadap performa model menggunakan data uji. Model terlebih dahulu menghasilkan prediksi kelas untuk setiap data uji, kemudian menghasilkan probabilitas prediksi yang menunjukkan tingkat keyakinan model terhadap setiap kelas.

Selanjutnya nilai *accuracy* dihitung untuk mengukur proporsi keseluruhan prediksi yang benar terhadap seluruh data uji. Selain itu, dilakukan pembentukan *classification report* yang menyajikan metrik *precision*, *recall*, dan *F1-Score* untuk masing-masing kelas, serta nilai rata-rata *macro average*, dan *weighted average* sebagai gambaran performa model secara keseluruhan, terutama dalam menangani distribusi kelas yang tidak seimbang. Tahap evaluasi juga mencakup pembentukan *confusion matrix*, yaitu metrik yang memperlihatkan jumlah prediksi benar dan salah pada setiap kelas sehingga dapat dianalisis kelas mana yang berhasil diklasifikasikan dengan baik maupun kelas yang masih mengalami kesalahan klasifikasi.



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

Melalui kombinasi metrik evaluasi tersebut, performa model dapat dianalisis secara lebih komprehensif, tidak hanya berdasarkan tingkat akurasi, tetapi juga berdasarkan kemampuan model dalam mengenali setiap kategori kualitas udara secara individual. Berikut hasil dari metrik evaluasi model klasifikasi pada penelitian ini bisa dilihat pada tabel 4.9 berikut.

Tabel 4. 9 Ringkasan Metrik Model

Metrik	Nilai
Akurasi	99.89%
Precision (Macro Avg)	99.56%
Recall (Macro Avg)	99.95%
F1-Score (Macro Avg)	99.75%
Precision (Weighted Avg)	99.89%
Recall (Weighted Avg)	99.89%
F1-Score (Weighted Avg)	99.89%

Selain itu, *Classification Report* menunjukkan nilai precision, recall, dan F1-score untuk setiap kelas kategori pencemaran udara. Nilai-nilai dari *Classification Report* ditunjukkan pada tabel 4.10 berikut :

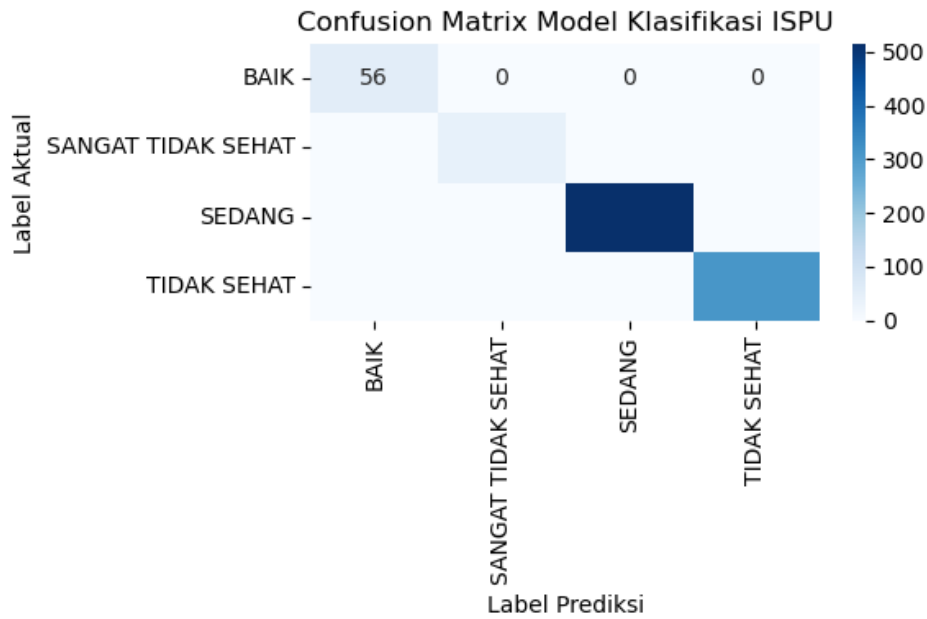
Tabel 4. 10 Hasil Classification Report

Kategori	Precision	Recall	F1-score	Support
BAIK	98.25%	100.00%	99.12%	56
SANGAT TIDAK SEHAT	100.00%	100.00%	100.00%	41
SEDANG	100.00%	99.81%	99.90%	515
TIDAK SEHAT	100.00%	100.00%	100.00%	313
Macro Avg	99.56%	99.95%	99.75%	925
Weighted Avg	99.89%	99.89%	99.89%	925



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. 31 Diagram *Confusion Matrix*

Confusion matrix menunjukkan model mampu mengklasifikasikan data kualitas udara dengan tingkat ketepatan yang sangat tinggi, Sebigain besar data pada setiap kategori berhasil diprediksi dengan benar. Kelas BAIK, SANGAT TIDAK SEHAT, dan TIDAK SEHAT berhasil diklasifikasikan seluruhnya tanpa kesalahan, sedangkan kelas SEDANG hanya terdapat satu data yang salah diklasifikasikan ke kelas TIDAK SEHAT. Hasil ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membendakan setiap kateogori kualitas udara dengan tingkat kesalahan klasifikasi yang sangat rendah, sehinnga mendukung nilai *accuracy*, *precision*, *recall*, dan *F1-Score* yang tinggi pada hasil evaluasi model.



© 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

- Deployment

```
joblib.dump(pipeline, "xgboost_model.pkl")
```

Gambar 4. 32 Tahap Evaluasi

Setelah model *XGBoost* berhasil dilatih dan dievaluasi, langkah selanjutnya adalah melakukan *deployment* agar model dapat digunakan secara praktis untuk melakukan proses klasifikasi pada data baru. *Deployment* ini memungkinkan model tidak hanya berhenti pada tahap penelitian, tetapi juga dapat diintegrasikan ke dalam aplikasi berbasis web yang siap digunakan oleh pengguna.

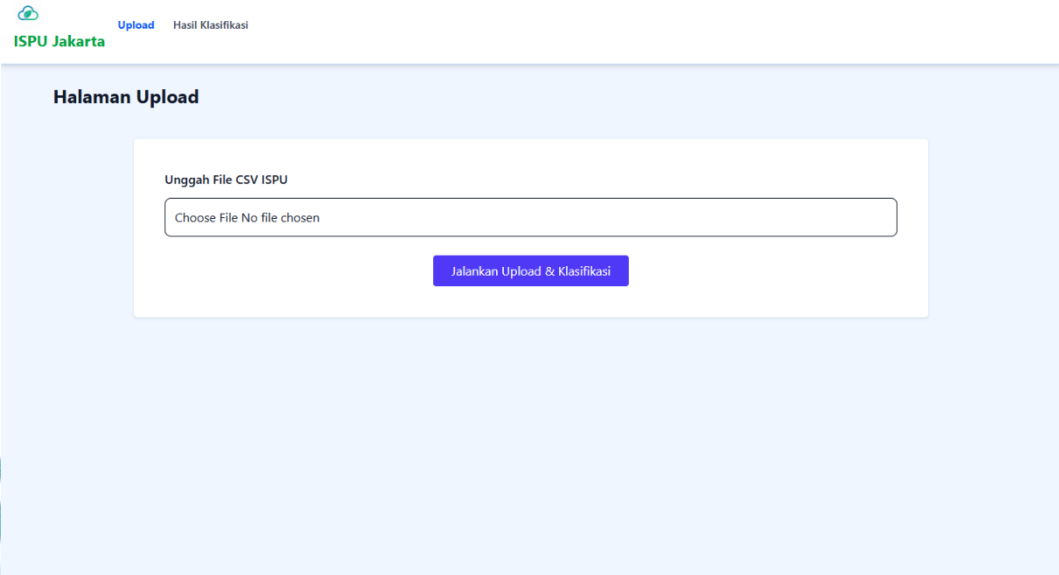
Dalam implementasinya, model *XGBoost* yang telah dilatih disimpan dalam bentuk file terkompresi (.pkl) sehingga dapat dimuat Kembali tanpa harus melakukan proses pelatihan dari awal. File model ini kemudian dihubungkan dengan aplikasi web berbasis Laravel yang telah dikembangkan.

Dengan pendekatan ini, pengguna dapat secara langsung memperoleh hasil klasifikasi pencemaran udara berdasarkan data terbaru. Model yang telah di-*deploy* juga dapat diperbarui secara berkala dengan data training tambahan agar performanya tetap optimal dan mampu beradaptasi dengan kondisi lingkungan yang dinamis.



4.3.2 Implementasi Web

- Halaman Upload



Gambar 4. 33 Tampilan Halaman Upload

Halaman upload merupakan halaman dimana proses klasifikasi pada aplikasi web klasifikasi pencemaran udara dimulai. Untuk memulai proses klasifikasi, pengguna dapat mengunggah file data ISPU yang berformat CSV. Terdapat *button* yang berfungsi untuk menjalankan fitur unggah data ISPU ke sistem sekaligus menjalankan fitur klasifikasi.



• Halaman Hasil Klasifikasi

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

Upload Hasil Klasifikasi

ISPU Jakarta

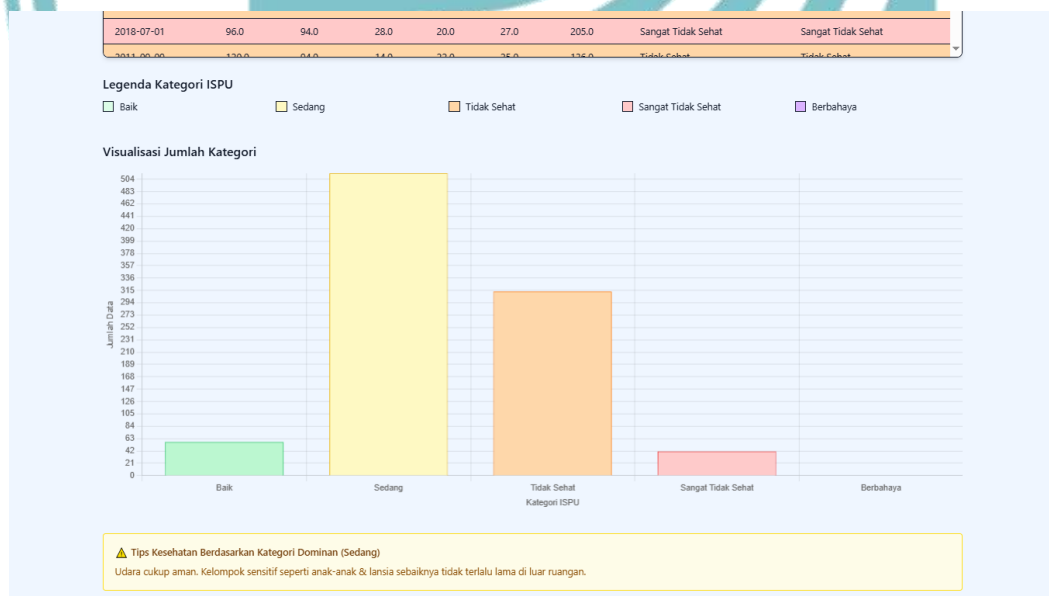
Hasil Klasifikasi

Penjelasan Kategori & Parameter ISPU

Download Hasil Klasifikasi (CSV) Unduh Laporan Evaluasi (PDF)

Tanggal	Pm10	Pm25	So2	No2	Co	O3	Kategori Aktual	Kategori Hasil
2015-06-26	114.0	94.0	34.0	18.0	31.0	67.0	Tidak Sehat	Tidak Sehat
2015-05-13	75.0	94.0	24.0	43.0	33.0	63.0	Sedang	Sedang
2017-09-22	68.0	94.0	52.0	12.0	26.0	161.0	Tidak Sehat	Tidak Sehat
2017-04-30	63.0	94.0	43.0	14.0	18.0	139.0	Tidak Sehat	Tidak Sehat
2017-09-23	81.0	94.0	53.0	19.0	31.0	137.0	Tidak Sehat	Tidak Sehat
2018-02-01	44.0	94.0	43.0	13.0	37.0	73.0	Sedang	Sedang
2010-07-04	56.0	94.0	13.0	23.0	46.0	51.0	Sedang	Sedang
2019-12-16	76.0	94.0	36.0	18.0	22.0	88.0	Sedang	Sedang
2015-03-28	68.0	94.0	17.0	17.0	33.0	77.0	Sedang	Sedang
2022-12-17	65.0	98.0	52.0	15.0	14.0	13.0	Sedang	Sedang
2010-05-19	69.0	94.0	14.0	18.0	39.0	79.0	Sedang	Sedang
2016-05-06	81.0	94.0	29.0	16.0	78.0	128.0	Tidak Sehat	Tidak Sehat

Gambar 4. 34 Tampilan Halaman Hasil Klasifikasi



Gambar 4. 35 Tampilan Diagram Hasil Klasifikasi



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

Hasil dari proses klasifikasi ditampilkan dalam bentuk tabel pada halaman ini. Selain itu terdapat informasi tambahan bagi pengguna seperti informasi mengenai penjelasan tentang kategori sesuai dengan aturan ISPU, visualisasi banyaknya data kategori yang dihasilkan dari proses klasifikasi, serta informasi mengenai tindakan kesehatan berdasarkan kategori yang dominan. Pada halaman ini juga terdapat *button* yang berfungsi bagi pengguna untuk mengunduh hasil klasifikasi dalam file berformat CSV serta untuk mengunduh hasil evaluasi model.

4.4 Pengujian

Tujuan utama dari penelitian ini adalah merancang, mengimplementasikan, serta menguji sistem klasifikasi pencemaran udara berbasis web menggunakan algoritma *XGBoost*. Pengujian dilakukan untuk memastikan bahwa sistem dapat berjalan sesuai dengan rancangan serta seluruh fungsionalitas yang ada dapat digunakan dengan baik oleh pengguna. Selain itu, pengujian juga dimaksudkan untuk memastikan keandalan model dalam menghasilkan hasil klasifikasi yang akurat.

4.4.1 Deskripsi Pengujian

Proses pengujian pada penelitian ini mencakup dua bagian utama. Pertama, dilakukan pengujian *Black Box Testing* terhadap aplikasi web yang dikembangkan. Pengujian ini bertujuan untuk memverifikasi apakah semua fitur seperti upload data ISPU, proses klasifikasi, serta tampilan hasil klasifikasi dapat berjalan sesuai dengan kebutuhan pengguna tanpa melihat kode program secara langsung. Setiap skenario pengujian disusun untuk mempresentasikan interaksi nyata antara pengguna dengan sistem.



© 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

Kedua, dilakukan pengujian terhadap performa model *XGBoost* yang telah dibangun. Pengujian ini menggunakan data uji yang berbeda dari data latih untuk menilai tingkat akurasi dan kemampuan generalisasi model. Ada beberapa metode pengujian yang diterapkan pada tahap pengujian performa model ini, yaitu pengujian performa model dengan membandingkan pengaruh penerapan teknik resampling terhadap performa model dan penggunaan matrik evaluasi. Beberapa matrik evaluasi yang digunakan dalam tahap ini meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Dengan evaluasi tersebut, diperoleh gambaran mengenai sejauh mana model mampu mengklasifikasikan kualitas udara dengan benar pada data baru.

- Black Box Testing

Tujuan diterapkannya metode *Black Box Testing* adalah untuk memastikan kinerja dari fungsionalitas dari sistem web dapat berjalan dengan sesuai tanpa terkendala *bug* atau *error*. Selain itu, diharapkan sistem web juga dapat memberikan hasil klasifikasi yang dapat diandalkan bagi pengguna. Rancangan pengujian melalui metode *Black Box Testing* ditampilkan pada Tabel 4. 11 dan Tabel 4. 12

Tabel 4. 11 Rancangan *Black Box Testing* 1

Kode	Kasus	Skenario	Harapan
B-B 1.1	Unggah Data	Pengguna berhasil mengunggah data yang valid	Sistem <i>redirect</i> ke halaman Hasil Klasifikasi
B-B 1.2			Sistem menampilkan pesan error mengenai data yang tidak valid

Tabel 4. 12 Rancangan *Black Box Testing* 2

Kode	Kasus	Skenario	Harapan
BB-2	Unduh hasil klasifikasi	Pengguna berhasil mengunduh data hasil klasifikasi	File data hasil klasifikasi
BB-3	Visualisasi data	Pengguna berhasil mengunggah data yang valid	Sistem menampilkan visualisasi data yang jelas

- Model Performance Testing

Pengujian kinerja model klasifikasi *XGBoost* dilakukan berdasarkan tiga metode yang berbeda, yaitu pengujian terhadap pengaruh dari penerapan teknik *resampling* terhadap performa model, pengujian terhadap pengaruh variasi persentase *split data* antara data uji dan data latih menggunakan metode *Stratified Train-Test Split*, dan pengujian berdasarkan hasil matrik evaluasi yang relevan untuk data multi-kelas, yaitu *accuracy*, *precision*, *recall*, *F1-Score*, *log-loss*, dan *ROC-AUC*.

Pada tahap pertama, pengujian dilakukan dengan membandingkan performa model pada tiga skenario yang berbeda, yaitu tanpa penerapan *resampling*, penerapan teknik *class weight*, serta penerapan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Perbandingan ini bertujuan untuk menilai pengaruh dari masing-masing teknik dalam mengatasi ketidakseimbangan kelas (*imbalanced dataset*) pada data ISPU. Dengan adanya perbedaan distribusi antar kelas, penggunaan *resampling* menjadi penting untuk memastikan model tidak hanya mempelajari pola dari kelas mayoritas, tetapi juga mampu mengenali kelas minoritas dengan 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



© 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

Pada tahap kedua, dilakukan pengujian terhadap beberapa variasi persentasi pembagian data latih dan data uji menggunakan metode *Stratified Train-Test Split*, yaitu 70:30, 80:20, dan 90:10. Metode ini dipilih karena mampu mempertahankan proporsi distribusi setiap kelas pada data latih maupun data uji sehingga

karakteristik dataset tetap terjaga. Pengujian ini bertujuan untuk mengetahui pengaruh perbedaan proporsi data latih dan data uji terhadap kemampuan generalisasi model serta menentukan komposisi pembagian data yang menghasilkan performa klasifikasi paling optimal.

Selanjutnya, tahap ketiga dilakukan dengan melakukan evaluasi performa model berdasarkan metrik-metrik utama. Setiap metrik memberikan gambaran yang berbeda terhadap kualitas hasil klasifikasi model. Pengujian kinerja model klasifikasi *XGBoost* dilakukan dengan menggunakan beberapa metrik evaluasi yang relevan untuk data multi-kelas, yaitu *Accuracy*, *Precision*, *Recall*, *F1-Score*, *Log Loss*, dan *ROC-AUC*. Pemilihan metrik-metrik ini didasarkan pada tujuan penelitian, yakni menilai seberapa baik model dalam mengklasifikasikan kategori kualitas udara berdasarkan data ISPU.

Dengan melakukan ketiga jenis pengujian tersebut, diharapkan dapat diperoleh gambaran yang komprehensif mengenai kinerja model dalam mengklasifikasikan tingkat pencemaran udara. Hasil pengujian ini menjadi dasar dalam menentukan model terbaik yang akan digunakan, serta memberikan pemahaman yang lebih mendalam terkait pengaruh penerapan penanganan ketidakseimbangan data, komposisi pembagia data latih dan data uji, dan efektivitas model *XGBoost* dalam menangani data dengan distribusi kelas yang tidak seimbang. Gambaran dari *model performance testing* dapat dilihat pada Tabel 4. 13, Tabel 4. 14, dan Tabel 4.15.



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

Tabel 4. 13 Tabel *Model Performance Testing 1*

Kode	Skenario Uji Coba	Deskripsi Pengujian
MPT-1.1	Model tanpa penanganan <i>imbalanced data</i>	Menguji performa dasar model <i>XGBoost</i> pada data asli tanpa penyeimbangan kelas untuk melihat <i>baseline performance</i> .
MPT-1.2	Model dengan penerapan <i>class weight</i>	Menguji pengaruh pemberian bobot kelas secara proporsional terhadap performa model <i>XGBoost</i> pada data <i>imbalanced</i> .
MPT-1.3	Model dengan penerapan SMOTE	Menguji pengaruh penambahan sampel sintetis pada kelas minoritas terhadap peningkatan performa model <i>XGBoost</i> .

Tabel 4. 14 Tabel Uji Model Performance Testing 2

Kode	Skenario Uji Coba	Deskripsi Pengujian
MPT-2.1	<i>Stratified Train-Test Split 70:30</i>	Membagi data latih dan data uji sebesar 70% dan 30% dengan tetap mempertahankan proporsi setiap kelas.



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

MPT-2.2	<i>Stratified Train-Test Split 80:20</i>	Membagi data latih dan data uji sebesar 80% dan 20% dengan tetap mempertahankan proporsi setiap kelas.
MPT-2.3	<i>Stratified Train-Test Split 90:10</i>	Membagi data latih dan data uji sebesar 90% dan 10% dengan tetap mempertahankan proporsi setiap kelas.

Tabel 4. 15 Rancangan *Model Performance Testing 3*

Kode	Skenario Uji Coba	Jumlah Data
MPT-3.1	Evaluasi dengan metrik <i>Accuracy, Precision, Recall, dan F1-Score</i>	925
MPT-3.2	Evaluasi dengan metrik <i>Log Loss</i>	925
MPT-3.3	Evaluasi dengan metrik <i>ROC-AUC (Macro)</i>	925

4.4.2 Prosedur Pengujian

Terdapat dua jenis pengujian yang digunakan pada penelitian ini, yaitu pengujian mandiri untuk memverifikasi fungsi sistem web dan pengujian kinerja model untuk mengukur tingkat akurasi dan efektivitas dari model klasifikasi yang telah dibuat.

1. Pengujian Mandiri

Pengujian mandiri dilakukan untuk mengevaluasi fungsionalitas sistem secara menyeluruh. Tahapan pengujian meliputi pengujian dengan menerapkan *Black Box Testing* untuk memverifikasi apakah setiap fitur dalam aplikasi web berjalan sesuai rancangan. Proses ini dilakukan dengan meniru interaksi pengguna terhadap sistem, misalnya proses unggah file, klasifikasi data, hingga menampilkan hasil klasifikasi.



© 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

Hasil yang didapat dari pengujian ini digunakan untuk memverifikasi bahwa sistem dapat melakukan *input*, menjalankan algoritma, dan menampilkan hasil klasifikasi sesuai dengan alur yang telah ditentukan dalam tahap perancangan sistem.

2. Pengujian Kinerja Model

Pengujian kinerja model dilakukan untuk mengevaluasi performa model klasifikasi *XGBoost* dalam mengklasifikasikan kualitas udara berdasarkan data ISPU. Pengujian kinerja model pada penelitian ini dilakukan dengan dua pendekatan :

a. Pengujian pengaruh penerapan teknik penanganan ketidakseimbangan data

Pengujian ini bertujuan untuk melihat sejauh mana penerapan teknik *resampling* (tanpa *resampling*, *class weight*, dan SMOTE) mempengaruhi performa model terhadap data yang bersifat *imbalanced*.

Setiap model diuji menggunakan data latih dan data uji yang sama agar hasil perbandingan bersifat objektif. Hasil pengujian kemudian dibandingkan menggunakan matrik evaluasi yang sama, sehingga dapat diketahui teknik mana yang memberikan hasil yang paling optimal.

b. Pengujian pengaruh variasi persentase pembagian data menggunakan *Stratified Train-Test Split*

Pengujian ini bertujuan untuk menganalisis pengaruh variasi proporsi data latih dan data uji terhadap performa model klasifikasi *XGBoost*. Metode ini digunakan karena mampu mempertahankan proporsi distribusi setiap kelas pada data latih maupun data uji sehingga karakteristik dataset tetap terjaga. Pengujian dilakukan menggunakan tiga variasi persentase pembagian data, yaitu 70:30, 80:20, dan 90:10. Hasil dari setiap skenario kemudian dibandingkan menggunakan metrik evaluasi yang sama untuk menentukan proporsi pembagian data yang menghasilkan kemampuan generalisasi dan performa klasifikasi terbaik.



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

c. Pengujian berdasarkan matrik evaluasi performa model

Setelah model terbaik diperoleh, dilakukan pengujian lanjutan menggunakan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, *F1-Score*, *log-loss*, dan ROC-AUC. Pemilihan metrik tersebut didasarkan pada karakteristik data multi-kelas serta kebutuhan penelitian untuk menilai akurasi, keseimbangan prediksi antar kelas, dan kemampuan model dalam membedakan kelas satu dengan lainnya.

Dengan menggunakan metrik-metrik ini, performa model dapat dievaluasi secara komprehensif baik dari sisi akurasi klasifikasi keseimbangan prediksi antar kelas, maupun probabilitas prediksi yang dihasilkan.

4.4.3 Data Hasil Pengujian

1. Black Box Testing

Pengujian Black Box dilakukan dengan membandingkan hasil aktual dengan hasil yang diharapkan. Jika keduanya sesuai, fitur dianggap berhasil. Hasil pengujian aplikasi web klasifikasi pencemaran udara ditunjukkan pada Tabel 4. 16 dan Tabel 4. 17.

Tabel 4. 16 Hasil *Black Box Testing 1*

Kode	Harapan	Aktual	Hasil
BB-1.1	Sistem redirect ke halaman Hasil Klasifikasi	Proses upload dan klasifikasi berhasil dan sistem redirect ke halaman Hasil Klasifikasi	Sesuai
BB-1.2	Proses upload dan klasifikasi gagal karena format file tidak valid	Sistem menampilkan pesan error dan tetap di halaman upload	Sesuai

Tabel 4. 17 Hasil *Black Box Testing* 2

BB-2	File data hasil klasifikasi berhasil terunduh	File data hasil klasifikasi terunduh ke penyimpanan lokal	Sesuai
BB-3	Sistem menampilkan visualisasi data yang jelas dan informatif	Sistem menampilkan visualisasi data yang jelas dan informatif	Sesuai

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

© Hak Cipta milik Politeknik Negeri Jakarta

- Model Performance Testing

Pengujian performa model dilakukan untuk mengetahui pengaruh variasi persentase pembagian data terhadap kemampuan model XGBoost dalam mengklasifikasikan kualitas udara berdasarkan data ISPU. Pada penelitian ini, pembagian data dilakukan menggunakan metode *Stratified Train-Test Split* dengan tiga variasi persentase, yaitu 70:30, 80:20, dan 90:10. Penggunaan metode *Stratified Train-Test Split* bertujuan untuk mempertahankan proporsi distribusi setiap kelas pada data latih dan data uji sehingga karakteristik dataset tetap terjaga selama proses pelatihan dan pengujian model.

Setiap scenario pembagian data menggunakan konfigurasi model yang sama sehingga perbedaan hasil yang diperoleh hanya dipengaruhi oleh proporsi data latih dan data uji. Selanjutnya, performa model dievaluasi menggunakan metrik *Accuracy*, *Precision (Macro Average)*, *Recall (Macro Average)*, *F1-Score (Macro Average)*, *Log Loss*, dan *ROC-AUC*. Penggunaan metrik tersebut bertujuan untuk memberikan evaluasi yang komprehensif terhadap kemampuan model dalam melakukan klasifikasi, baik dari sisi ketepatan prediksi secara keseluruhan maupun kemampuan model dalam mengenali seluruh kategori kualitas udara secara seimbang. Hasil dari *model performance testing* dapat dilihat pada Tabel 4. 18 dan Tabel 4.19



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

Tabel 4. 18 Hasil Pengujian Variasi Persentase Split Data

Persentase Split Data	Akurasi	Precision	Recall	F1-Score	Log Loss	ROC AUC
70:30	99.64%	99.29%	98.93%	99.11%	0.0262	0.9976
80:20	99.89%	99.56%	99.95%	99.75%	0.0049	1.0000
90:10	99.78%	99.14%	99.90%	99.51%	0.0057	1.0000

Hasil pengujian menunjukkan bahwa pembagian data 80:20 menghasilkan performa terbaik dibandingkan variasi lainnya, Skenario ini memperoleh nilai *Accuracy* sebesar 99.89%, *F1-Score Macro* sebesar 99.75%, serta *Log Loss* terendah sebesar 0.0049 sehingga dipilih sebagai konfigurasi pembagian data terbaik pada penelitian ini.

Tabel 4. 19 Hasil Performance Testing

Metode	Akurasi	Precision	Recall	F1-Score	Log Loss	ROC AUC
Tanpa Penanganan Imbalance	99.78%	99.31%	99.31%	99.31%	0.0057	1.0000
Class Weight	99.89%	99.56%	99.95%	99.75%	0.0049	1.0000
SMOTE	99.78%	98.97%	99.87%	99.41%	0.0066	1.0000

Berdasarkan hasil pengujian, metode *Class Weight* memberikan performa terbaik dibandingkan metode lainnya. Model memperoleh nilai *Accuracy* sebesar 99.89%, *F1-Score Macro* sebesar 99.75%, serta *Log Loss* terendah sebesar 0.0049. Hal ini menunjukkan bahwa pemberian bobot pada kelas minoritas lebih efektif dibandingkan tanpa penanganan ketidakseimbangan data maupun metode *SMOTE* dalam meningkatkan kemampuan model mengklasifikasikan seluruh kategori kualitas udara.



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

4.4.4 Evaluasi Pengujian

Dalam penelitian ini, salah satu tantangan utama yang dihadapi adalah kualitas dan karakteristik data ISPU yang diperoleh dari sumber resmi. *Dataset* masih mengandung beberapa nilai yang tidak valid, seperti *missing value*, nilai nol yang tidak merepresentasikan hasil pengukuran, serta nilai negatif pada beberapa parameter pencemar. Oleh karena itu, sebelum dilakukan proses pelatihan model, diterapkan tahapan *data preprocessing* yang meliputi normalisasi label target, penyesuaian nama atribut, penanganan nilai kosong menggunakan metode *median imputation*, konversi nilai negatif menjadi *missing value*, serta analisis terhadap keberadaan *outlier* berdasarkan batas logis setiap parameter pencemar. Tahapan tersebut bertujuan untuk meningkatkan kualitas data sehingga dapat digunakan secara optimal pada proses klasifikasi.

Selain itu, distribusi kelas pada *dataset* ISPU menunjukkan kondisi *imbalanced dataset*, di mana kategori SEDANG dan TIDAK SEHAT memiliki jumlah data yang jauh lebih besar dibandingkan kategori BAIK dan SANGAT TIDAK SEHAT. Ketidakseimbangan tersebut berpotensi menyebabkan model lebih cenderung mempelajari pola dari kelas mayoritas sehingga kemampuan klasifikasi terhadap kelas minoritas menjadi menurun. Oleh karena itu, penelitian ini melakukan pengujian terhadap tiga pendekatan penanganan ketidakseimbangan data, yaitu tanpa penerapan metode *imbalance data handling*, *Class Weight*, dan *Synthetic Minority Oversampling Technique* (SMOTE). Pengujian dilakukan untuk mengetahui metode yang paling efektif dalam meningkatkan performa model pada seluruh kategori kualitas udara.

Selain membandingkan metode penanganan ketidakseimbangan data, penelitian ini juga melakukan pengujian terhadap variasi pembagian data menggunakan metode *Stratified Train-Test Split* dengan tiga skenario, yaitu 70:30, 80:20, dan 90:10. Metode *Stratified Train-Test Split* dipilih karena mampu mempertahankan proporsi distribusi setiap kelas pada data latih maupun data uji sehingga karakteristik dataset tetap terjaga selama proses pelatihan dan pengujian model.



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

Berdasarkan hasil pengujian, pembagian data 80:20 menghasilkan performa terbaik dibandingkan dengan persentase *split data* yang lainnya, dengan 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. Oleh karena itu, konfigurasi pembagian data tersebut digunakan pada tahap pengujian metode penanganan ketidakseimbangan data.

Model klasifikasi yang digunakan dalam penelitian ini adalah *XGBoost Classifier* karena berbagai penelitian terdahulu algoritma ini menunjukkan kinerja yang baik dalam menangani data tabular, mendukung klasifikasi multi-kelas, serta menerapkan mekanisme *gradient boosting* yang efektif dalam mempelajari hubungan antar fitur. Untuk memperoleh konfigurasi model yang lebih optimal, dilakukan proses *Hyperparameter Tuning* menggunakan metode *RandomizedSearchCV* yang dikombinasikan dengan *Stratified K-Fold Cross Validation* sebanyak lima lipatan (5-fold). Pendekatan ini bertujuan untuk memperoleh kombinasi parameter terbaik sekaligus memastikan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya.

Hasil pengujian menunjukkan bahwa metode *Class Weight* memberikan performa terbaik dibandingkan dua pendekatan lainnya yaitu tanpa penanganan *imbalanced dataset* dan SMOTE. Model yang menerapkan *Class Weight* memperoleh *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, dan ROC-AUC sebesar 1,0000. Dibandingkan dengan model tanpa penanganan ketidakseimbangan data maupun model yang menerapkan SMOTE, penggunaan *Class Weight* menghasilkan nilai *F1-Score Macro* yang lebih tinggi serta *Log Loss* yang lebih rendah. Hal tersebut menunjukkan bahwa pemberian bobot pada kelas minoritas mampu meningkatkan keseimbangan performa model tanpa mengubah distribusi data asli, sehingga model menjadi lebih sensitif dalam mengenali kelas minoritas sekaligus tetap mempertahankan performa yang tinggi pada kelas mayoritas.



© Hak Cipta milik Politeknik Negeri Jakarta

Meskipun menghasilkan performa yang sangat baik, penelitian ini masih memiliki beberapa keterbatasan. Model sangat bergantung pada kualitas data masukan sehingga kesalahan format, nilai yang hilang, maupun data yang tidak sesuai standar dapat memengaruhi hasil klasifikasi. Selain itu, penelitian ini hanya memanfaatkan enam parameter pencemar udara sebagai variabel prediktor tanpa mempertimbangkan faktor eksternal seperti kondisi meteorologi, curah hujan, kecepatan angin, kelembapan udara, maupun musim yang diketahui dapat memengaruhi konsentrasi polutan di atmosfer. Oleh karena itu, penelitian selanjutnya diharapkan dapat mengintegrasikan faktor-faktor tersebut agar model mampu menghasilkan performa klasifikasi yang lebih komprehensif dan memiliki kemampuan generalisasi yang lebih 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 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

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.

