



**PERBANDINGAN ALGORITMA *EXTREME GRADIENT BOOSTING*
(XGBOOST) DAN *RANDOM FOREST* DALAM KLASIFIKASI
KUALITAS UDARA DI DAERAH KHUSUS JAKARTA**

SKRIPSI

**Untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
HANI PUSPITA
NIM. 2111017120001**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JANUARI 2025**



**PERBANDINGAN ALGORITMA *EXTREME GRADIENT BOOSTING*
(XGBOOST) DAN *RANDOM FOREST* DALAM KLASIFIKASI
KUALITAS UDARA DI DAERAH KHUSUS JAKARTA**

SKRIPSI

**Untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
HANI PUSPITA
NIM. 2111017120001**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JANUARI 2025**

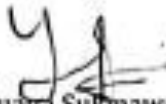
SKRIPSI

**PERBANDINGAN ALGORITMA *EXTREME GRADIENT BOOSTING*
(XGBOOST) DAN *RANDOM FOREST* DALAM KLASIFIKASI KUALITAS
UDARA DI DAERAH KHUSUS JAKARTA**

Oleh:
Hani Puspita
2111017120001

Telah dipertahankan pada hari Jumat, tanggal 17-01-2025 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I



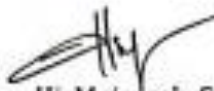
Yuana Sukmawaty, S.Si., M.Si
NIP. 198810152015042002

Penguji I



Dewi Sri Susanti, S.Si., M.Si
NIP. 197305161999032002

Pembimbing II



Hj. Malsarah, S.Pd.I., M.Pd
NIP. 19890713201801213001

Penguji II



Sigit Dwi Prabowo, S.Mat., M.Stat
NIP. 199605282024061003

Banjarbaru, 30 Januari 2025

Mengetahui,
Koordinator Program Studi Statistika
EMIPA ULM



Prof. Dewi Angraini, S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 17 Januari 2025



Hani Puspita

NIM. 2111017120001

ABSTRAK

Perbandingan Algoritma *Extreme Gradient Boosting* (XGBoost) Dan *Random Forest* Dalam Klasifikasi Kualitas Udara Di Daerah Khusus Jakarta (Oleh: Hani Puspita; Pembimbing: Yuana Sukmawaty dan Maisarah, 2025; 33 halaman)

Daerah Khusus Jakarta merupakan salah satu kota metropolitan dengan tingkat aktivitas ekonomi dan kepadatan penduduk yang sangat tinggi. Seiring dengan pesatnya pembangunan, Jakarta menghadapi berbagai tantangan lingkungan, termasuk masalah kualitas udara. Klasifikasi kualitas udara menggunakan *machine learning* menjadi penting untuk memberikan informasi yang lebih akurat dan mendukung pengambilan kebijakan berbasis data dalam mengatasi tantangan ini. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Extreme Gradient Boosting* (XGBoost) dan *Random Forest* dalam klasifikasi kualitas udara di Daerah Khusus Jakarta. Kualitas udara dinilai berdasarkan Indeks Standar Pencemar Udara (ISPU) dengan parameter PM_{10} , $PM_{2.5}$, SO_2 , CO , O_3 , dan NO_2 . Penelitian ini melibatkan analisis deskriptif untuk memahami pola distribusi kualitas udara, penanganan ketidakseimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE), implementasi algoritma XGBoost dan *Random Forest*, serta evaluasi kinerja model berdasarkan akurasi, sensitivitas, dan spesifisitas. Hasil analisis deskriptif menunjukkan bahwa kualitas udara di Jakarta sebagian besar berada pada kategori sedang, yang memiliki proporsi tertinggi dibandingkan kategori baik, tidak sehat, sangat tidak sehat, dan berbahaya. Algoritma XGBoost menunjukkan kinerja yang lebih baik dibandingkan *Random Forest* dalam mengklasifikasikan kualitas udara. Penanganan ketidakseimbangan data menggunakan SMOTE terbukti mampu meningkatkan performa kedua algoritma, terutama pada kategori pada kelas minoritas.

Kata Kunci: XGBoost, *Random Forest*, Kualitas Udara, Daerah Khusus Jakarta, SMOTE

ABSTRACT

Comparison of Extreme Gradient Boosting (XGBoost) and Random Forest Algorithms in Air Quality Classification in Jakarta Special Region (By: Hani Puspita; Advisors: Yuana Sukmawaty and Maisarah, 2025; 36 pages)

The Special Region of Jakarta is one of the metropolitan cities with a very high level of economic activity and population density. Along with rapid development, Jakarta faces various environmental challenges, including air quality issues. Air quality classification using machine learning is important to provide more accurate information and support data-driven policy making in addressing these challenges. This study aims to compare the performance of Extreme Gradient Boosting (XGBoost) and Random Forest algorithms in air quality classification in Jakarta Special Region. Air quality is assessed based on the Air Pollution Standard Index (ISPU) with parameters PM_{10} , $PM_{2.5}$, SO_2 , CO , O_3 , and NO_2 . This research involves descriptive analysis to understand the distribution pattern of air quality, handling data imbalance using Synthetic Minority Oversampling Technique (SMOTE), implementation of XGBoost and Random Forest algorithms, and evaluation of model performance based on accuracy, sensitivity, and specificity. The results of descriptive analysis show that air quality in Jakarta is mostly in the moderate category, which has the highest proportion compared to the good, unhealthy, very unhealthy, and dangerous categories. Handling data imbalance using SMOTE is proven to improve the performance of both algorithms, especially in categories with minority classes.

Keywords: XGBoost, Random Forest, Air Quality, Special Region of Jakarta, SMOTE

PRAKATA

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah dan inayah-Nya sehingga penulis dapat penyusunan tugas akhir dengan judul “Perbandingan Algoritma *eXtreme Gradient Boosting* (XGBoost) Dan *Random Forest* Dalam Klasifikasi Kualitas Udara Di Daerah Khusus Jakarta”. Penyusunan tugas akhir ini bertujuan untuk memenuhi persyaratan dalam rangka menyelesaikan program sarjana S-1 di program studi Statistika FMIPA ULM.

Dalam proses penyelesaian tugas akhir ini, penulis mendapat berbagai bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang mendalam kepada:

1. Kedua orang tua dan keluarga yang selalu mendukung serta doa, nasihat, dan bantuan baik secara moril maupun materil.
2. Ibu Prof. Dewi Anggraini, S.Si., M.App.Sci., Ph.D selaku Koordinator Program Studi Statistika FMIPA ULM atas segala dukungan dan ilmu yang diberikan selama belajar di Program Studi Statistika.
3. Ibu Yuana Sukmawaty, S.Si., M.Si., dan Ibu Maisarah, S.Pd.I., M.Pd. selaku dosen pembimbing atas segala bimbingan, arahan, dan saran yang diberikan kepada penulis sehingga penyusunan tugas akhir ini dapat diselesaikan.
4. Ibu Dewi Sri Susanti, S.Si., M.Si., dan Bapak Sigit Dwi Prabowo, S.Mat., M.Stat selaku dosen penguji atas segala masukan dan saran yang diberikan selama proses penyusunan tugas akhir.
5. Seluruh Dosen dan staff Program Studi Statistika FMIPA ULM yang telah memberikan ilmu, motivasi, dan segala nasihat selama masa perkuliahan.
6. Orang-orang terkasih lainnya yang belum penulis sebutkan secara spesifik karena telah kebersamaian penulis selama masa-masa perkuliahan.
7. Terakhir, Terima kasih kepada Wanita sederhana yang memiliki keinginan tinggi namun terkadang sulit dipahami isi kepalanya, sang penulis karya tulis ini yaitu diri saya sendiri, Hani Puspita. Seorang anak bungsu yang berjalan menuju 22 tahun yang keras kepala namun terkadang sifatnya seperti anak kecil pada umumnya. Terima kasih telah hadir di dunia dan sudah bertahan sampai sejauh ini melewati banyaknya tantangan rintangan yang alam semesta berikan. Terima kasih, kamu sudah sangat hebat, saya bangga dengan atas pencapaian yang telah diraih dalam hidupmu dan selalu merayakan dirimu sampai dititik ini, walau seringkali pengharapan tidak sesuai dengan ekspektasi, namun harus tetap bersyukur, Terima kasih selalu mau berusaha, bekerja sama dan tidak lelah mencoba hal-hal positif. Berbahagialah selalu dimanapun, kapanpun, kamu berada, hani. Rayakan selalu kehadiranmu, jadilah bersinar dimanapun kamu memijakkan kaki. Semoga Langkah kebaikan terus berada padamu dan semoga allah selalu meridhoi setiap perbuatanmu dan selalu dalam lindungannya. Aamiin.

Banjarbaru, 13 Januari 2025

A handwritten signature in black ink, appearing to read 'Hani Puspita', with a horizontal line extending from the bottom right of the signature.

Hani Puspita

NIM. 2111017120001

DAFTAR ISI

PERNYATAAN.....	i
ABSTRAK.....	ii
ABSTRACT	iii
PRAKATA.....	iv
DAFTAR ISI	vi
DAFTAR GAMBAR.....	viii
DAFTAR TABEL	ix
DAFTAR LAMPIRAN.....	x
DAFTAR LAMBANG DAN SINGKATAN	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	4
2.1 Kajian Penelitian Terdahulu	4
2.2 Kajian teori.....	5
2.2.1 Kualitas Udara.....	5
2.2.2 Statistika Deskriptif.....	6
2.2.3 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	7
2.2.4 Klasifikasi.....	7
2.2.5 <i>Decision Tree</i>	7
2.2.6 <i>Extreme Gradient Boosting</i>	8
2.2.7 <i>Random Forest</i>	10
2.2.8 <i>Confusion Matrix</i>	11
BAB III METODE PENELITIAN.....	13
3.1 Sumber data.....	13
3.2 Variabel penelitian.....	13
3.3 Prosedur penelitian.....	14
BAB IV HASIL DAN PEMBAHASAN	17
4.1 Deskripsi Data Kualitas Udara Daerah Khusus Jakarta.....	17
4.1.1 Data <i>pre-processing</i>	17
4.1.2 Statistika Deskriptif	18
4.2 Implementasi algoritma XGBoost dan <i>Random Forest</i>	21
4.2.1 <i>Resampling</i> Data.....	21
4.2.2 Analisis Klasifikasi Menggunakan Algoritma XGBoost.....	22
4.2.3 Analisis Klasifikasi Menggunakan Algoritma <i>Random Forest</i>	24
4.3 Evaluasi Kinerja Model Klasifikasi	26
4.3.1 Kinerja Model Pada Algoritma XGBoost.....	26
4.3.2 Kinerja Model Pada Algoritma <i>Random Forest</i>	28
4.4 Perbandingan Kinerja Model Klasifikasi.....	30
BAB V PENUTUP.....	32
5.1 Kesimpulan	32
5.2 Saran.....	32

DAFTAR PUSTAKA	33
LAMPIRAN	36
Lampiran 1. Data Penelitian	36
Lampiran 2. Data <i>Training</i>	37
Lampiran 3. Data <i>Testing</i>	38
Lampiran 4. Hasil kombinasi <i>tuning parameter</i> Algoritma XGBoost.....	39
Lampiran 5. <i>Syntax</i> python menggunakan <i>Google Colab</i>	42
Lampiran 6. <i>Output</i> dari visualisasi data	51
Lampiran 7. <i>Output</i> pohon pertama XGBoost.....	52
Lampiran 8. <i>Output</i> pohon pertama <i>random forest</i>	53
Lampiran 9. <i>Output</i> dari <i>confusion matrix</i> XGBoost	54
Lampiran 10. <i>Output</i> dari <i>confusion matrix random forest</i>	55
RIWAYAT HIDUP.....	56