



**PREDIKSI KUALITAS UDARA DI SUMATERA SELATAN
BERDASARKAN INDEKS STANDAR PENCEMARAN UDARA
MENGUNAKAN ALGORITMA EXTREME GRADIENT
BOOSTING**

SKRIPSI

MUHAMMAD JODI NOVIANDRI

21141045P

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS TEKNOLOGI

UNIVERSITAS BINA DARMA

2024



**PREDIKSI KUALITAS UDARA DI SUMATERA SELATAN
BERDASARKAN INDEKS STANDAR PENCEMARAN UDARA
MENGUNAKAN ALGORITMA EXTREME GRADIENT
BOOSTING**

**MUHAMMAD JODI NOVIANDRI
21141045P**

**Skripsi ini diajukan sebagai syarat memperoleh gelar
Sarjana Komputer di Universitas Bina Darma**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS TEKNOLOGI
UNIVERSITAS BINA DARMA
2024**

HALAMAN PENGESAHAN

**PREDIKSI KUALITAS UDARA DI SUMATERA SELATAN
BERDASARKAN INDEKS STANDAR PENCEMARAN UDARA
MENGUNAKAN ALGORITMA EXTREME GRADIENT
BOOSTING**

MUHAMMAD JODI NOVIANDRI
21141045P

Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi

Palembang, 15 Juli 2024
Fakultas Sains Teknologi
Universitas Bina Darma
Dekan,

Pembimbing



Ferdiansyah, M.Kom., Ph.D.


Universitas Bina Darma
Fakultas Sains Teknologi

Dr. Tata Sutabri, S.Kom., MMSI., MKM

HALAMAN PERSETUJUAN

Skripsi Berjudul "PREDIKSI KUALITAS UDARA DI SUMATERA SELATAN BERDASARKAN INDEKS STANDAR PENCEMARAN UDARA MENGGUNAKAN ALGORITMA EXTREME GRADIENT BOOSTING" Oleh "Muhammad Jodi Novlandri", telah dipertahankan di depan komisi penguji pada hari Senin tanggal 15 Juli 2024.

Komisi Penguji

1. Ketua : Ferdiansyah, M.Kom., Ph.D.
2. Anggota : Andri, S.Kom., M.CS
3. Anggota : Novri Hadinata, M.Kom.

()

()

()

Mengetahui,
Program Studi Sistem Informasi
Fakultas Sains Teknologi
Universitas Bina Darma
Ketua,

Universitas Bina Darma
Fakultas Sains Teknologi

()

Nita Rosa Damayanti, M.Kom., Ph.D.

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Jodi Noviandri

NIM : 21141045P

Dengan ini menyatakan bahwa:

1. Karta tulis saya (Skripsi) adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan dari tim pembimbing ;
3. Di dalam karya tulis ini terdapat karya atau pendapat yang telah ditulis untuk dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan ;
4. Saya bersedia tugas skripsi, di cek keasliannya menggunakan plagiarism checker serta diunggah ke internet, sehingga dapat di akses secara daring ;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan perundang-undangan yang berlaku ;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Jakarta, 10 September 2024

Yang membuat pernyataan



Muhammad Jodi Noviandri

21141045P

MOTTO DAN PERSEMBAHAN

Motto

"Karena sesungguhnya sesudah kesulitan itu ada kemudahan, sesungguhnya sesudah kesulitan itu ada kemudahan."

(Q.S Al Insyirah: 5-6)

PERSEMBAHAN

Dengan penuh rasa syukur, skripsi ini saya persembahkan kepada:

1. Kedua Orang Tua Tercinta, yang selalu memberikan doa, dukungan, dan cinta tanpa batas. Terima kasih atas segala pengorbanan, kasih sayang, dan motivasi yang tiada henti.
2. Saudara-Saudaraku, yang selalu memberikan semangat dan dukungan dalam setiap langkah perjalanan ini.
3. Diri sendiri, yang selalu berusaha menyelesaikan hal yang telah dimulai dengan baik.

ABSTRACT

Forest and land fires are events that cause forest or land areas on a small or large scale to burn. In Indonesia, especially in the area of South Sumatra Province, forest and land fires every year have become like an annual "tradition" but with different levels of pollution every year. At this time on October 2, 2023 based on data from the IQAir website accessed at 10:00 PM that Palembang City, South Sumatra Province, occupies the first position of the 10 cities in Indonesia with the worst air pollution levels with an AQI value of US or average unit 259 based on AQI parameters. From January to August 2023, an area of 4,082.8 hectares of land in South Sumatra burned, this happened because it was on mineral and peat land.

This reason makes the discussion of air quality prediction one of the things that is widely researched. This research uses a supervised learning machine learning algorithm, namely eXtreme Gradient Boosting to get prediction results with the best level of accuracy. This research will also create a website-based application that can be used by the general public to find out and predict air quality based on AQI directly.

Keywords: Prediction, Air Quality, Air Pollutant Standard Index, XGBoost

ABSTRAK

Kebakaran hutan dan lahan merupakan suatu kejadian yang menyebabkan Kawasan hutan ataupun lahan dalam skala kecil maupun luas terbakar. Di Indonesia, khususnya di daerah Provinsi Sumatera Selatan, kebakaran hutan dan lahan setiap tahunnya sudah menjadi seperti “tradisi” tahunan namun dengan tingkat polusi yang berbeda setiap tahunnya. Pada saat ini tanggal 2 Oktober 2023 berdasarkan data dari website IQAir yang di akses pada jam 10.00 PM bahwa Kota Palembang Provinsi Sumatera Selatan menempati posisi pertama dari 10 kota di Indonesia yang paling buruk tingkat polusi udaranya dengan nilai AQI US atau rata rata unit 259 berdasarkan parameter ISPU. Sejak Januari hingga Agustus 2023, seluas 4.082,8 hektare lahan di Sumatera Selatan terbakar, hal itu terjadi dikarenakan berada di lahan mineral dan gambut.

Alasan tersebut membuat pembahasan mengenai prediksi kualitas udara menjadi salah satu hal yang banyak diteliti. Dalam penelitian ini menggunakan algoritma *machine learning supervised learning* yaitu *eXtreme Gradient Boosting* untuk mendapatkan hasil prediksi dengan tingkat akurasi terbaik. Dalam penelitian ini juga akan membuat sebuah aplikasi berbasis website yang dapat digunakan oleh khalayak umum untuk mengetahui dan memprediksi kualitas udara berdasarkan ISPU secara langsung

Kata Kunci: Prediksi, Kualitas Udara, Indeks Standar Pencemar Udara, XGBoost

KATA PENGANTAR

Alhamdulillah puji dan syukur penulis ucapkan ke hadirat Allah Yang Maha Kuasa, karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul “PREDIKSI KUALITAS UDARA DI SUMATERA SELATAN BERDASARKAN INDEKS STANDAR PENCEMARAN UDARA MENGGUNAKAN ALGORITMA EXTREME GRADIENT BOOSTING” dengan baik dan diselesaikan dengan tepat waktu.

Skripsi ini disusun guna memenuhi salah satu syarat untuk memperoleh gelar S.Kom di Program Studi Sistem Informasi, Fakultas Sains Teknologi, Universitas Bina Darma Palembang. Penulisan Skripsi ini dapat diselesaikan dengan baik atas bantuan dari berbagai pihak.

Dalam kesempatan ini penulis mengucapkan terima kasih kepada :

1. Allah SWT yang Maha Kuasa atas Karunia dan Nikmat-Nya sehingga penulis dapat menyelesaikan Skripsi ini dengan baik.
2. Kedua orang tua saya yang menjadi orang tua yang luar biasa dalam memberikan dukungan Doa, moral serta kasih sayang sehingga penulis dapat mengerjakan dan menyelesaikan Skripsi ini dengan baik.
3. Dr. Sunda Ariana, M.Pd., M.M. Rektor Universitas Bina Darma Palembang
4. Dr. Tata Sutabri, S.Kom., MMSI., MKM Dekan Fakultas Sains Teknologi, Universitas Bina Darma Palembang.
5. Nita Rosa Damayanti., M.Kom., Ph.D. Ketua Program Studi Sistem Informasi Universitas Bina Darma Palembang.
6. Ferdiansyah, M.Kom., Ph.D. Sebagai Pembimbing yang telah membimbing penulis dengan baik hati dan sabar serta memberikan arahan yang baik sehingga penulisan Skripsi ini dapat diselesaikan.
7. Dosen Program Studi Sistem Informasi, Sains Teknologi, Universitas Bina Darma yang telah mendidik serta memberikan ilmu pengetahuan.
8. Kementerian Lingkungan Hidup dan Kehutanan yang telah membantu dalam pemenuhan data yang dibutuhkan.

Semoga Allah yang Maha Kuasa memberikan balasan dengan segala kebaikan dunia dan akhirat atas keikhlasan dan kebaikan kepada semua pihak yang telah diberikan kepada penulis.

Penulis menyadari bahwa dalam penulisan skripsi ini masih terdapat kekurangan. Oleh karena itu, penulis sangat terbuka terhadap kritik dan saran yang membangun demi perbaikan di masa yang akan datang.

Akhir kata, penulis berharap agar skripsi ini dapat bermanfaat bagi pembaca dan dapat memberikan kontribusi positif dalam bidang [Bidang Ilmu Skripsi].

Jakarta, 16 Juni 2024

Penulis,

Muhammad Jodi Noviandri

21141045P

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN	iv
SURAT PERNYATAAN	v
MOTTO DAN PERSEMBAHAN	vi
ABSTRACT	vii
ABSTRAK	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan dan Manfaat Penelitian	5
1.4.1 Tujuan.Penelitian.....	5
1.4.2 Manfaat Penelitian.....	5
1.5 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	7
2.1 Pencemaran Udara	7
2.2 Indeks Standar Pencemar Udara.....	8
2.3.Artificial Intelligence	11
2.4 Machine Learning.....	15
2.4.1 Supervised Learning	16
2.4.2 <i>Unsupervised Learning</i>	17
2.4.3 Semi Supervised Learning dan Reinforcement Learning.....	18
2.5 <i>Deep Learning</i>	18
2.6 Ensemble Learning.....	20
2.6.1 Bagging Classifier (Bootstrap Aggregating)	21

2.6.2 Boosting	22
2.6.3 Stacking	23
2.7 Extreme Gradient Boosting	24
2.7.1 Tahapan algoritma Extreme Gradient Boosting.....	28
BAB III METODOLOGI PENELITIAN	35
3.1 Objek Penelitian	35
3.2 Jenis Data	35
3.3 Metode Penelitian.....	35
3.4 Waktu, tempat dan jadwal penelitian	35
3.5 Kerangka Penelitian	36
3.6 Data Penelitian	40
3.7 Dataset	40
3.8 Pengukuran Kinerja Algoritma Klasifikasi	56
3.8.1 Confusion Matrix	56
3.8.2 Area Under the Curve (AUC)	57
BAB IV PEMBAHASAN	44
4.1 Persiapan Data.....	44
4.2 <i>Preprocessing Data</i>	60
4.2.1 <i>Identification & Handling Missing Values</i>	60
4.2.2 <i>Removing Duplicates</i>	62
4.2.3 Mengkoreksi Data Type pada tiap kolom	62
4.2.4 Menghapus kolom tabel yang tidak digunakan	63
4.2.5 <i>Handling Outliers</i>	64
4.2.6 Mengubah Kategorikal Data ke Data Binary.....	65
4.3 <i>Splitting Train Data & Test Data</i>	66
4.4 Analisis Evaluasi Model XGBoost	67
4.4.1 Perbandingan data train 70% dan data test 30%	67
4.4.1.1 <i>Classification Report Data Train</i>	67
4.4.1.2 <i>Classification Report Data Test</i>	69

4.4.1.3	<i>Confusion Matrix</i>	72
4.4.1.4	<i>AUC-ROC Curve</i>	73
4.4.1.5	<i>XGBoost Tree</i>	74
4.4.2	Perbandingan Data Train 80% dan Data Test 20%	76
4.4.2.1	<i>Classification Report Data Train</i>	76
4.4.2.2	<i>Classification Report Data Test</i>	79
4.4.2.3	<i>Confusion Matrix</i>	81
4.4.2.4	<i>AUC-ROC Curve</i>	82
4.4.2.5	<i>XGboost Tree</i>	83
4.4.3	Perbandingan Data Train 90% dan Data Test 10%.....	85
4.4.3.1	<i>Classification Report Data Train</i>	85
4.4.3.2	<i>Classification Report Data Test</i>	88
4.4.3.3	<i>Confusion Matrix</i>	90
4.4.3.4	<i>AUC-ROC Curve</i>	91
4.4.3.5	<i>XGBoost Tree</i>	92
4.5	Visualisasi Hasil Prediksi ISPU.....	99
4.5.1	Hasil Prediksi dari Parameter PM10	100
4.5.2	Hasil Prediksi dari Parameter PM2.5	101
4.5.3	Hasil Prediksi dari Parameter SO ₂	103
4.5.4	Hasil Prediksi dari Parameter CO.....	105
4.5.5	Hasil Prediksi dari Parameter O ₃	106
4.5.6	Hasil Prediksi dari Parameter NO ₂	108
4.5.7	Hasil Prediksi dari Parameter HC	109
BAB V	KESIMPULAN DAN SARAN	96
5.1	Kesimpulan	96
5.2	Saran.....	97
DAFTAR PUSTAKA	98

DAFTAR GAMBAR

	Halaman
BAB II TINJAUAN PUSTAKA	12
Gambar 2.1 Diagram Jenis AI	16
Gambar 2.2 Flowchart <i>Supervised Learning AI</i>	22
Gambar 2.3 Ilustrasi Proses <i>Ensemble Learning</i>	25
Gambar 2.4 Konsep Kerja <i>Bagging</i>	26
Gambar 2.5 Konsep <i>Boosting</i>	27
Gambar 2.6 Konsep kerja dari <i>Stacking</i>	28
Gambar 2.7 Alur Skematik Dari Xgboost	29
Gambar 2.8 <i>Building Tree</i>	35
Gambar 2.9 Kalkulasi Pada Nilai Similarity Serta Nilai Gain	36
Gambar 2.10 Split Pada <i>Tree XGBoost</i>	36
Gambar 2.11 Turunan Percabangan Yang Di Split	37
Gambar 2.12 Perhitungan Similarity Dan Gain Pada Split Lanjutan	37
Gambar 2.13 Pohon Yang Dilakukan Proses Pruning.	38
Gambar 2.14 Sisa Daun Yang Dilakukan Perhitungan Nilai Keluaran.....	38
BAB IV HASIL DAN PEMBAHASAN	51
Gambar 4.1 Data yang masih ada outliers	56
Gambar 4.2 Data yang sudah dibersihkan dari outliers	57
Gambar 4.3 Hasil Confusion Matrix proporsi data 70:30	64
Gambar 4.4 Kurva ROC XGBoost	65
Gambar 4.5 XGBoost Tree.....	66
Gambar 4.6 Hasil Confusion Matrix proporsi data 80:20	73
Gambar 4.7 Kurva ROC XGBoost	74
Gambar 4.8 XGBoost Tree	75
Gambar 4.9 Hasil Confusion Matrix proporsi data 90:10	82
Gambar 4.10 Kurva ROC XGBoost	83

Gambar 4.11 XGBoost Tree	84
Gambar 4.12 Hasil <i>Learning Curves Logloss</i> di perbandingan 1	95
Gambar 4.13 Hasil <i>Learning Curves Logloss</i> di perbandingan 2	96
Gambar 4.14 Hasil <i>Learning Curves Logloss</i> di perbandingan 3	97
Gambar 4.15 Hasil <i>Feature Importance</i>	99
Gambar 4.16. Hasil Prediksi PM10	100
Gambar 4.17 Forecast pada datatest set	100
Gambar 4.18 Merge dataset PM10 dan hasil forecast	101
Gambar 4.19 Hasil Prediksi PM2.5	102
Gambar 4.20 Forecast pada datatest set	102
Gambar 4.21 Merge dataset PM2.5 dan hasil forecast	103
Gambar 4.22 Hasil Prediksi SO2	103
Gambar 4.23. Forecast pada datatest set	104
Gambar 4.24 Merge dataset SO2 dan hasil forecast	104
Gambar 4.25 Hasil Prediksi CO	105
Gambar 4.26 Forecast pada datatest set	105
Gambar 4.27 Merge dataset CO dan hasil forecast	106
Gambar 4.28 Hasil Prediksi O3	106
Gambar 4.29 Forecast pada datatest set	107
Gambar 4.30 Merge dataset O3 dan hasil forecast	107
Gambar 4.31 Hasil Prediksi NO2	108
Gambar 4.32 Forecast pada datatest set	108
Gambar 4.33 Merge dataset NO2 dan hasil forecast	109
Gambar 4.34 Hasil Prediksi HC	109
Gambar 4.35 Forecast pada datatest set	110
Gambar 4.36 Merge dataset HC dan hasil forecast	110

DAFTAR TABEL

	Halaman
BAB II TINJAUAN PUSTAKA	12
Tabel 2.1 Kategori Indeks Standar Pencemar Udara (ISPU)	23
Tabel 2.2 Koncensi Nilai Konsentrasi Parameter ISPU	24
Tabel 2.3. Contoh Dataset	42
Tabel 2.4 Pehitungan Nilai <i>Error</i> atau <i>Residuals</i>	43
Tabel 2.5 Hasil Perhitungan dan <i>residual</i>	48
BAB III METODOLOGI PENELITIAN	49
Tabel 3.1 Jadwal Penelitian	50
Tabel 3.2 Kerangka Pemikiran	50
Tabel 3.3 Confusion Matrix	55
Tabel 3.4 Nilai dan Kategori AUC	57
BAB IV HASIL DAN PEMBAHASAN	49
Tabel 4.1 Contoh dataset ISPU	50
Tabel 4.2 Tipe Data pada Dataset	59
Tabel 4.3 Jumlah data yang kosong	59
Tabel 4.4 Contoh data yang kosong pada dataset	60
Tabel 4.5 Contoh data yang telah diisi	60
Tabel 4.6 Dataset yang bersih dari duplikasi data	61
Tabel 4.7 Tipe data pada dataset	61
Tabel 4.8 Mengubah tipe data	62
Tabel 4.9 Contoh data yang telah dihapus dua kolom tabel	62
Tabel 4.10 One Hot Encoding	64
Tabel 4.11 Jumlah dari split data	65
Tabel 4.12 Hasil Performa Data Train Proporsi 70%	66
Tabel 4.13 Hasil Performa data test dengan proporsi 30%	68
Tabel 4.14 Hasil Performa data train dengan proporsi 80%	75

Tabel 4.15 Hasil Performa data test dengan proporsi 20%	78
Tabel 4.16 Hasil Performa data train dengan proporsi 90%	84
Tabel 4.17 Hasil Performa data test dengan proporsi 10%	87
Tabel 4.18 Hasil perbandingan performa pada setiap proporsi data	93



DAFTAR LAMPIRAN

Formulir Perbaikan Komprehensif
Surat Keterangan Lulus Komprehensif
SK Pembimbing
Permohonan Acc Pengajuan Judul Skripsi
Lembar Konsultasi Bimbingan Skripsi

