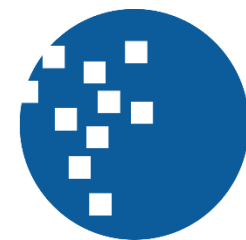


**ANALISIS KLASIFIKASI KUALITAS UDARA DI KOTA
BEKASI BERDASARKAN AQI DENGAN ALGORITMA
SUPPORT VECTOR MACHINE, XGBOOST, DAN RANDOM
FOREST**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Mary Elizabeth Tjang

00000057284

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

**ANALISIS KLASIFIKASI KUALITAS UDARA DI KOTA
BEKASI BERDASARKAN AQI DENGAN ALGORITMA
SUPPORT VECTOR MACHINE, XGBOOST, DAN RANDOM
FOREST**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

Mary Elizabeth Tjang

00000057284

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERSETUJUAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Mary Elizabeth Tjang

Nomor Induk Mahasiswa : 00000057284

Program Studi : Sistem Informasi

Skripsi dengan judul:

**ANALISIS KLASIFIKASI KUALITAS UDARA DI KOTA BEKASI
BERDASARKAN AQI DENGAN ALGORITMA SUPPORT VECTOR
MACHINE, XGBOOST, DAN RANDOM FOREST**

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 4 Juni 2025



Mary Elizabeth Tjang

HALAMAN PERSETUJUAN

SKRIPSI dengan judul

**ANALISIS KLASIFIKASI KUALITAS UDARA DI KOTA BEKASI
BERDASARKAN AQI DENGAN ALGORITMA SUPPORT VECTOR
MACHINE, XGBOOST, DAN RANDOM FOREST**

Oleh

Nama : Mary Elizabeth Tjang
NIM : 00000057284
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah disetujui untuk diajukan pada
Sidang Ujian Skripsi Universitas Multimedia Nusantara
Tangerang, 4 Juni 2025

Pembimbing



Johan Setiawan, S.Kom., MM., M.B.A
327106402

Ketua Program Studi Sistem Informasi



Ririn Ikana Desanti, S.Kom., M.Kom

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS KLASIFIKASI KUALITAS UDARA DI KOTA BEKASI
BERDASARKAN AQI DENGAN ALGORITMA SUPPORT VECTOR
MACHINE, XGBOOST, DAN RANDOM FOREST

Oleh

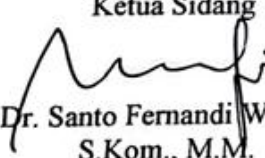
Nama : Mary Elizabeth Tjang
NIM : 00000057284
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Rabu, 11 Juni 2025

Pukul 15.00 s.d 16.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

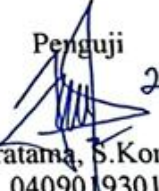
Ketua Sidang

Dr. Santo Fernandi Wijaya,
S.Kom., M.M.
0310016902

Pembimbing


Johan Setiawan, S.Kom., M.M., M.B.A
0327106402

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.

Penguji

Jansen Wiratama, S.Kom., M.Kom.
0409019301

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Mary Elizabeth Tjang

NIM : 00000057284

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah :

ANALISIS KLASIFIKASI KUALITAS UDARA DI KOTA BEKASI BERDASARKAN AQI DENGAN ALGORITMA SUPPORT VECTOR MACHINE, XGBOOST, DAN RANDOM FOREST

Menyatakan dengan sesungguhnya bahwa saya bersedia* (pilih salah satu):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 4 Juni 2025



(Mary Elizabeth Tjang)

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa atas rahmat-Nya sehingga skripsi dengan judul: “Analisis Klasifikasi Kualitas Udara di Kota Bekasi Berdasarkan AQI dengan Algoritma Support Vector Machine, XGBOOST dan Random Forest” dapat terselesaikan. Skripsi ini dilakukan untuk memenuhi salah satu syarat mencapai gelar Strata 1 Jurusan Sistem Informasi pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.

Proses penelitian dan penyusunan skripsi ini tidak terlepas dari bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, terima kasih yang sebesar-besarnya kepada:

Mengucapkan terima kasih

1. Bapak Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Bapak Johan Setiawan, S.Kom., M.M., M.B.A., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Teman-teman saya yang senantiasa memberikan dukungan moral.

Semoga karya ilmiah ini dapat bermanfaat, baik sebagai sumber informasi maupun referensi kualitas udara di Bekasi bagi para pembaca.

Tangerang, 4 Juni 2025



Mary Elizabeth Tjang

ANALISIS KLASIFIKASI KUALITAS UDARA DI KOTA BEKASI BERDASARKAN AQI DENGAN ALGORITMA SUPPORT VECTOR MACHINE, XGBOOST, DAN RANDOM FOREST

Mary Elizabeth Tjang

ABSTRAK

Kualitas udara merupakan faktor penting yang berdampak langsung terhadap kesehatan masyarakat. Di daerah perkotaan seperti Bekasi yang mengalami pertumbuhan industri dan kepadatan lalu lintas yang tinggi, risiko pencemaran udara semakin meningkat. Oleh karena itu, dibutuhkan suatu sistem klasifikasi kualitas udara yang akurat dan dapat diakses oleh masyarakat untuk mendukung pengambilan keputusan dan pencegahan risiko kesehatan lingkungan.

Penelitian ini menggunakan algoritma *machine learning* untuk membangun model klasifikasi kualitas udara berdasarkan data historis. Tiga algoritma yang digunakan adalah Support Vector Machine, XGBoost, dan Random Forest. Dataset yang digunakan terdiri dari parameter pencemar udara seperti PM10, SO₂, CO, O₃, dan NO₂. Proses pelatihan dilakukan dengan teknik *preprocessing*, normalisasi data, pembagian data latih dan uji, serta evaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score*. Model terbaik kemudian *deploy* menggunakan Streamlit untuk membuat antarmuka prediksi berbasis web.

Hasil dari penelitian ini menunjukkan bahwa algoritma XGBoost memberikan performa terbaik dengan akurasi mencapai 99,8%, disusul oleh Random Forest dengan akurasi 98,77%, dan SVM sebesar 97,5%. Sistem klasifikasi kualitas udara yang dibangun dapat digunakan sebagai alat bantu untuk memantau dan mengklasifikasikan kondisi udara di Bekasi secara efektif, sehingga dapat mendukung upaya pengawasan lingkungan secara lebih terstruktur dan akurat. Implementasi menggunakan Streamlit memungkinkan sistem ini dapat diakses oleh pengguna non-teknis serta membuka peluang untuk pengembangan yang lebih luas di masa mendatang.

Kata kunci: Kualitas Udara, Machine Learning, dan Klasifikasi

ANALYSIS OF AIR QUALITY CLASSIFICATION IN BEKASI CITY BASED ON AQI USING SUPPORT VECTOR MACHINE, XGBOOST, AND RANDOM FOREST ALGORITHMS

Mary Elizabeth Tjang

ABSTRACT (English)

Air quality is a crucial factor that directly impacts public health. In urban areas like Bekasi, which are experiencing rapid industrial growth and high traffic density, the risk of air pollution is increasing significantly. Therefore, an accurate and accessible air quality classification system is essential to support decision-making and prevent environmental health risks.

This study employs a machine learning approach to build an air quality classification model based on historical data. Three algorithms were used: Support Vector Machine, XGBOOST, and Random Forest. The dataset includes air pollution parameters such as PM10, SO2, CO, O3, and NO2. The training process involved data preprocessing, normalization, splitting into training and test sets, and evaluation using accuracy, precision, recall, and f1-score metrics. The best-performing model was deployed using Streamlit to create a web-based prediction interface.

The results indicate The findings reveal that XGBoost achieved the highest accuracy of 99.8%, followed by Random Forest at 98.77%, and SVM at 97.5% Therefore, the developed air quality prediction system can serve as an effective tool for monitoring and classifying air conditions in Bekasi. The implementation using Streamlit makes the system accessible to non-technical users and opens up opportunities for broader development in the future.

Keywords: Air Quality, Machine Learning, Classification

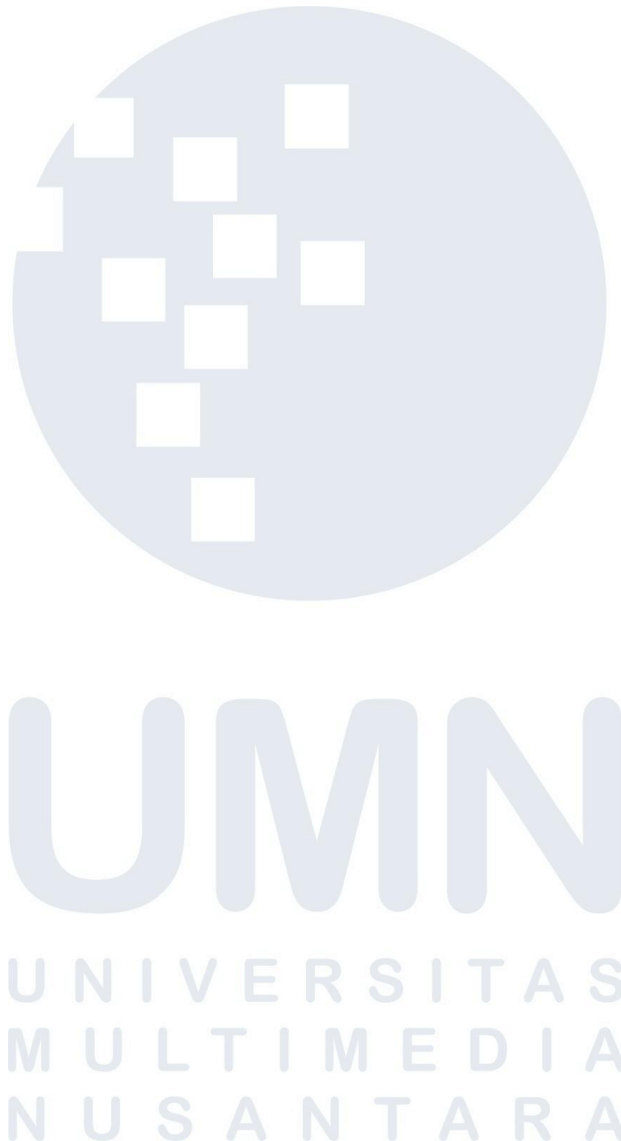
DAFTAR ISI

HALAMAN PERSETUJUAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR.....	vi
ABSTRAK	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.3 Batasan Masalah	6
1.4 Tujuan dan Manfaat Penelitian.....	7
1.4.1 Tujuan Penelitian.....	7
1.4.2 Manfaat Penelitian.....	7
1.5 Sistematika Penulisan	8
BAB II LANDASAN TEORI	10
2.1 Penelitian Terdahulu	10
2.2 Teori Penelitian	17
2.2.1 Udara.....	17
2.2.2 AQI	25
2.3 Framework dan Algoritma Penelitian.....	26
2.3.1 CRISP-DM	26
2.3.2 Support Vector Machine.....	28
2.3.3 XGBoost (Xtreme Gradient Boosting).....	29
2.3.4 Random Forest.....	30
2.4 Tools dan Software Penelitian.....	31

2.4.1	Jupyter Notebook.....	31
2.4.2	Python.....	32
2.4.3	Anaconda	33
BAB III METODOLOGI PENELITIAN		35
3.1	Gambaran Umum Objek Penelitian.....	35
3.4	Metode Penelitian.....	36
3.1.1	Alur Penelitian	36
3.1.2	Metode Data Mining.....	38
3.4	Teknik Pengumpulan Data	42
3.4	Teknik Analisis Data.....	43
3.5	Teknik Pengujian.....	44
BAB IV ANALISIS DAN HASIL PENELITIAN		47
4.1	Business Understanding	47
4.2	Data Understanding	48
4.3	Data Preparation.....	50
4.4	Modelling	57
4.5	Evaluation	65
4.6	Deployment	69
4.7	Hasil Penelitian	73
4.8	Pembahasan Hasil Penelitian	74
BAB V SIMPULAN DAN SARAN.....		75
5.1	Simpulan.....	75
5.2	Saran.....	77
DAFTAR PUSTAKA.....		78
LAMPIRAN.....		83

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	10
Tabel 2. 2 Indeks Air Quality Index (AQI).....	25
Tabel 3. 1 Flowchart Tahap Alur Penelitian	38
Tabel 3. 2 Sample Data Kualitas Udara Kota Bekasi Tahun 2021 – 2025	40
Tabel 4. 1 Hasil evaluasi dari 3 algoritma.....	73

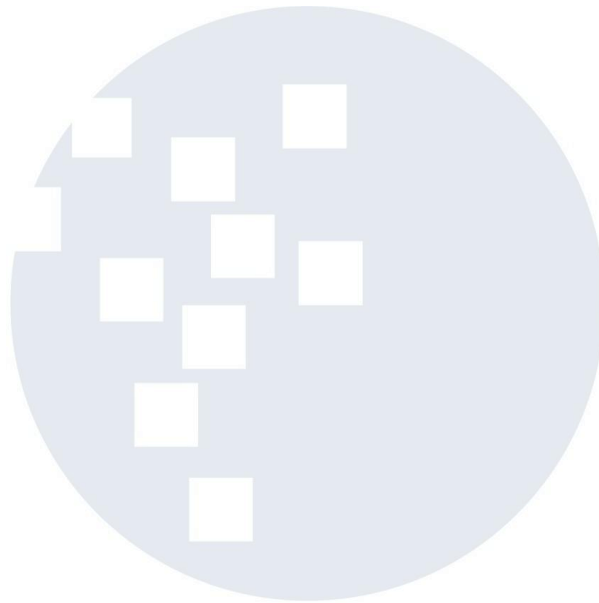


DAFTAR GAMBAR

Gambar 1. 1 Berita tentang kualitas udara di Bekasi	2
Gambar 1. 2 Grafik Rangking kota paling berpolusi di Indonesia	3
Gambar 2. 1 Diagram CRISP-DM	26
Gambar 2. 3 Hyperplane SVM.....	28
Gambar 2. 4 Cara kerja Random Forest.....	31
Gambar 2. 5 Logo Jupyter.....	32
Gambar 2. 6 Logo Python	33
Gambar 2. 7 Anaconda Navigator.....	34
Gambar 4. 1 Data Udara Bekasi.....	49
Gambar 4. 2 Kode Import Library	50
Gambar 4. 3 Dataset.....	50
Gambar 4. 4 Info Dataset	51
Gambar 4. 5 Cek null (missing values) pada data.....	52
Gambar 4. 6 Labeling "count" menjadi kelas	53
Gambar 4. 7 Hasil Labeling	54
Gambar 4. 8 Menyiapkan fitur dan label untuk klasifikasi.....	55
Gambar 4. 9 Normalisasi.....	55
Gambar 4. 10 Pembagian Data.....	56
Gambar 4. 11 Random Grid	56
Gambar 4. 12 Membuat 3 model algoritma	57
Gambar 4. 13 K-Fold Cross Validation	58
Gambar 4. 14 Confusion Matrix	59
Gambar 4. 15 Confusion Matrix menggunakan algoritma SVM.....	59
Gambar 4. 16 Confusion Matrix menggunakan algoritma XGBoost	60
Gambar 4. 17 Confusion Matrix menggunakan algoritma Random Forest	62
Gambar 4. 18 Kode Analisis Distribusi	63
Gambar 4. 19 Grafik Performa 3 Model	63
Gambar 4. 20 Akurasi algoritma SVM	65
Gambar 4. 21 Analisis Kesalahan algoritma SVM	65
Gambar 4. 22 Akurasi algoritma XGBoost.....	66
Gambar 4. 23 Analisis Kesalahan XGBoost.....	66
Gambar 4. 24 Akurasi pada algoritma Random Forest.....	67
Gambar 4. 25 Analisis kesalahan pada algoritma Random Forest.....	68
Gambar 4. 26 Import library untuk deploy	69
Gambar 4. 27 Pembuatan model_udara	69
Gambar 4. 28 Tampilan website Streamlit.....	70
Gambar 4. 29 Grafik distribusi PM2.5 dan PM10	71
Gambar 4. 30 Line Chart rata-rata tahunan PM2.5 dan PM10	72
Gambar 4. 77 Line Chart rata-rata tahunan PM2.5 dan PM10	72

DAFTAR RUMUS

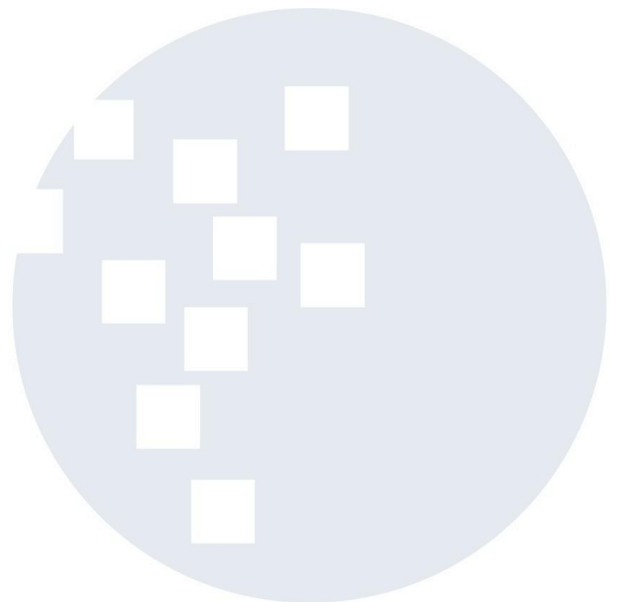
Rumus 2. 1 Rumus Hyperplane	28
Rumus 2. 2 Rumus Hyperplane (2).....	29
Rumus 2. 3 Rumus XGBoosting.....	30



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report	83
Lampiran B Form Konsultasi Bimbingan	92



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA