

**ANALISIS PERBANDINGAN *RANDOM FOREST* DAN
SUPPORT VECTOR MACHINES (SVM) DALAM
MENGKLASIFIKASIKAN KUALITAS UDARA JAKARTA**

TUGAS AKHIR

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana (S1)
Sistem Informasi**



**UNIVERSITAS
BAKRIE**

KHAIRUNISSA MAUDINA

1202002030

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN ILMU KOMPUTER

UNIVERSITAS BAKRIE

2025

HALAMAN PERNYATAAN ORISINALITAS

**Tugas Akhir ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.**

Nama : Khairunissa Maudina

NIM : 1202002030

Tanda Tangan :



Tanggal : 28 Februari 2025

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Khairunissa Maudina
NIM : 1202002030
Program Studi : Sistem Informasi
Fakultas : Fakultas Teknik dan Ilmu Komputer
Judul Skripsi : Analisis Perbandingan *Random Forest* dan *Support Vector Machines* (SVM) dalam Mengklasifikasikan Kualitas Udara Jakarta.

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

DEWAN PENGUJI

Pembimbing 1 : Zakiul Fahmi Jailani, S.Kom., M.Sc.

()

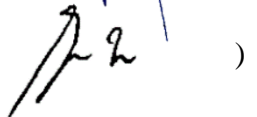
Pembimbing 2 : Dita Nurmadewi, S.Kom., M.Kom.

()

Penguji 1 : Dr. Elin Cahyaningsih, S.Kom., M.MSi.

()

Penguji 2 : Dr. Shidiq Al Hakim, S.T, M.Eng.

()

Ditetapkan di : Jakarta

Tanggal : 28 Februari 2025

UNGKAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat, hidayah, dan kemudahan, sehingga penulis dapat menyelesaikan penelitian ini dengan judul "Analisis Perbandingan *Random Forest* dan *Support Vector Machines* (SVM) dalam Mengklasifikasikan Kualitas Udara Jakarta". Dalam kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada pihak-pihak yang telah memberikan dukungan, bimbingan, serta bantuan dalam proses penyelesaian skripsi ini.

1. Bapak Zakiul Fahmi Jailani, S.Kom., M.Sc., selaku Pembimbing 1, yang telah memberikan bimbingan, arahan, serta dukungan yang sangat berarti selama penelitian ini.
2. Ibu Dita Nurmadewi, S.Kom., M.Kom., selaku Pembimbing 2, atas bimbingan, kritik, dan saran yang membangun yang sangat membantu dalam memperbaiki kualitas penelitian ini.
3. Prof. Dr. Hoga Saragih, S.T., M.T., IPM., CIRR., MIEEE., M.Th., Ph.D., selaku Ketua Program Studi Sistem Informasi Universitas Bakrie, atas dukungan dan fasilitas yang diberikan selama penulis menjalani studi di program ini.
4. Dinas Lingkungan Hidup DKI Jakarta, yang telah memberikan akses data kualitas udara yang sangat penting bagi kelancaran penelitian ini.
5. Dr. Elin Cahyaningsih, S.Kom., M.MSi. dan Dr. Shidiq Al Hakim, S.T., M.Eng., selaku Penguji 1 dan Penguji 2, yang telah memberikan masukan yang sangat berharga dalam penyusunan skripsi ini.
6. Staff Program Studi Sistem Informasi Universitas Bakrie, yang telah membantu dalam mengurus segala proses administrasi selama penyelesaian skripsi ini. Terima kasih atas pelayanan yang baik dan profesional.
7. Keluarga penulis yang, terutama kedua orang tua yang selalu memberikan doa, dukungan, dan kasih sayang yang tiada henti.
8. Karima Mahesa, Agung Kurniawan, dan Hilmi Syarira, yang telah memberikan semangat moral dan dukungan yang luar biasa selama proses penelitian ini.

9. Kirana Azzahra, Wanda Arofana, Mita Yesica, dan Naura Luviezka, yang selalu memberi dorongan semangat dan motivasi untuk terus maju dan tidak menyerah.
10. Teman-teman seperjuangan penulis: Yusuf Ikhsan, Yusran Rizky, Alfina Regita, Fela Elviana, Ergi Cahya, dan Nhisha Aulia, yang telah menemani dan memberi dukungan sepanjang perjalanan skripsi ini.
11. Terima kasih juga disampaikan kepada pihak-pihak yang tidak dapat disebutkan satu per satu, namun yang telah memberikan kontribusi, dukungan, dan bantuan dalam menyelesaikan penelitian ini.

Jakarta, 28 Februari 2025



Khairunissa Maudina

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama : Khairunissa Maudina
NIM : 1202002030
Program Studi : Sistem Informasi
Fakultas : Fakultas Teknik Ilmu Komputer
Jenis Tugas Akhir : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul: "Analisis Perbandingan *Random Forest* dan *Support Vector Machines* (SVM) dalam Mengklasifikasikan Kualitas Udara Jakarta".

Beserta perangkat yang ada (jika diperlukan). Dengan **Hak Bebas Royalti Noneksklusif** ini, Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Dibuat di : Jakarta

Pada Tanggal : 28 Februari 2025

Yang Menyatakan



Khairunissa Maudina

**ANALISIS PERBANDINGAN RANDOM FOREST DAN SUPPORT
VECTOR MACHINES (SVM) DALAM MENGLASIFIKASIKAN
KUALITAS UDARA JAKARTA**

Khairunissa Maudina

ABSTRAK

Polusi udara merupakan masalah besar di kota-kota seperti Jakarta, dengan dampak signifikan terhadap kesehatan masyarakat. Indeks Standar Pencemaran Udara (ISPU) digunakan untuk mengukur kualitas udara, namun tidak memberikan gambaran langsung tentang risiko kesehatan. Penelitian ini bertujuan untuk mengonversi data ISPU Jakarta menjadi *Air Quality Health Index* (AQHI), yang mempertimbangkan dampak kesehatan dari polutan utama seperti PM_{2.5}, PM₁₀, SO₂, CO, O₃, dan NO₂. Dengan menggunakan algoritma machine learning, yaitu *Support Vector Machine* (SVM) dan *Random Forest* (RF), penelitian ini bertujuan untuk membandingkan kinerja kedua model tersebut dalam mengklasifikasikan kualitas udara berdasarkan AQHI potensi risiko kesehatan akibat polusi udara.

Hasil penelitian menunjukkan bahwa model *Random Forest* (RF) lebih unggul dibandingkan SVM, dengan akurasi RF sebesar 98% dan akurasi SVM sebesar 96%, serta nilai *precision*, *recall*, dan *F1-score* yang lebih tinggi pada RF. Model AQHI yang dikembangkan dapat memberikan prediksi risiko kesehatan yang lebih relevan dan berpotensi digunakan untuk sistem peringatan dini terkait kualitas udara di Jakarta, terutama untuk kelompok rentan seperti anak-anak, lansia, dan individu dengan kondisi pernapasan atau kardiovaskular.

Kata Kunci: AQHI, ISPU, *Random Forest*, *Support Vector Machine*, *Machine learning*

**COMPARISON ANALYSIS OF RANDOM FOREST AND SUPPORT
VECTOR MACHINES (SVM) IN CLASSIFYING JAKARTA AIR QUALITY**

Khairunissa Maudina

ABSTRACT

Air pollution is a major issue in cities like Jakarta, with significant impacts on public health. The Indeks Standar Pencemaran Udara (ISPU) is used to measure air quality, but it does not provide a direct picture of health risks. This study aims to convert Jakarta's ISPU data into the Air Quality Health Index (AQHI), which considers the health impacts of major pollutants such as PM_{2.5}, PM₁₀, SO₂, CO, O₃, and NO₂. Using machine learning algorithms, namely Support Vector Machine (SVM) and Random Forest (RF), this study seeks to compare the performance of both models in classifying air quality based on AQHI and the potential health risks posed by air pollution.

The results show that the Random Forest (RF) model outperforms SVM, with an accuracy of 98% for RF and 96% for SVM, as well as higher precision, recall, and F1-score values for RF. The developed AQHI model can provide more relevant health risk predictions and has the potential to be used for early warning systems related to air quality in Jakarta, particularly for vulnerable groups such as children, the elderly, and individuals with respiratory or cardiovascular conditions.

Keywords: AQHI, ISPU, Random Forest, Support Vector Machine, Machine learning

DAFTAR ISI

HALAMAN JUDUL	1
HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN.....	iii
UNGKAPAN TERIMAKASIH	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR SKRIP	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	5
1.4 Batasan Masalah.....	5
1.5 Tujuan Penelitian.....	5
1.6 Manfaat Penelitian.....	5
1.7 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Polusi Udara	7
2.2 Indeks Standar Pencemar Udara (ISPU) dan <i>Air Qualiry Index (AQI)</i>	7
2.3 <i>Air Quality Health Index (AQHI)</i>	9

2.3.1 Perhitungan AQHI	10
2.3.2 Kategori Risiko Berdasarkan AQHI	10
2.3.3 Keunggulan AQHI	11
2.4 <i>Machine Learning</i>	12
2.5 Algoritma <i>Random Forest</i>	13
2.6 Algoritma <i>Support Vector Machines (SVM)</i>	15
2.7 Metode <i>Evaluation Metrics</i> Untuk Klasifikasi	16
2.7.1 <i>Confusion Matrix</i>	17
2.7.2 Akurasi (<i>Accuracy</i>)	18
2.7.3 Presisi (<i>Precision</i>)	18
2.7.4 <i>Recall</i> (Sensitivitas atau <i>True Positive Rate</i>)	18
2.7.5 <i>F1-Score</i>	19
BAB III METODOLOGI PENELITIAN	17
3.1 Kerangka Penelitian	17
3.1.1 Studi Literatur	17
3.1.2 <i>Data Preparation</i>	17
3.1.3 <i>Model Training</i>	18
3.1.4 <i>Model Evaluation</i>	18
3.1.5 Penyusunan laporan	18
3.2 Objek Penelitian	18
3.3 Model Analisis Data	20
3.4 Alokasi Waktu Penelitian	20
BAB IV HASIL DAN PEMBAHASAN	21
4.1 <i>Data Preparation</i>	21
4.1.1 Penggabungan Data Menggunakan <i>Excel</i>	21
4.1.2 <i>Data Cleaning</i>	22

4.1.3 Konversi ISPU ke AQHI	22
4.2 Analisis Eksplorasi Data (<i>Exploratory Data Analysis</i>)	24
4.2.1 <i>Data Overview</i>	24
4.2.2 <i>Correlation Analysis</i>	24
4.2.3 Hubungan Antar Variabel (<i>Pairwise Relationships</i>)	26
4.2.4 <i>Numerical Data Analysis</i>	26
4.2.5 Analisis Fitur Kategorikal (<i>Categorical Feature Analysis</i>).....	28
4.2.6 Penanganan Outliers	30
4.3 <i>Feature Engineering and Encoding</i>	33
4.4 Pemisahan Data (<i>Data Splitting</i>)	33
4.5 <i>Model Training</i> dan Evaluasi	34
4.6 Pembahasan	37
BAB V KESIMPULAN DAN SARAN	38
5.1 Kesimpulan.....	38
5.2 Saran	39
DAFTAR PUSTAKA	40
LAMPIRAN.....	43

DAFTAR GAMBAR

Gambar 2. 1 Model Pembelajaran <i>Machine Learning</i>	12
Gambar 2. 2 Model Pembelajaran <i>Random Forest</i>	14
Gambar 2. 3 Model Pembelajaran SVM	15
Gambar 3. 1 Kerangka Penelitian	17
Gambar 4. 1 <i>Correlation Matrix</i>	25
Gambar 4. 2 Distribusi Numerik Polutan.....	27
Gambar 4. 3 <i>Boxplot Categorical</i>	29
Gambar 4. 4 Distribusi Setelah Penanganan <i>Outliers</i>	32
Gambar 4. 5 <i>Boxplot</i> Setelah Penanganan <i>Outliers</i>	33

DAFTAR TABEL

Tabel 2. 1 Konversi Nilai Konsentrasi Parameter ISPU	8
Tabel 2. 2 Kategori Indeks Standar Pencemar Udara (ISPU)	9
Tabel 2. 3 <i>Confusion Matrix</i>	17
Tabel 2. 4 Penelitian Terdahulu	13
Tabel 3. 1 Data Polutan yang Digunakan dalam Penelitian.....	19
Tabel 4. 1 Hasil Evaluasi Model RF dan SVM.....	36

DAFTAR SKRIP

Skrip 4. 1 Konversi Nilai ISPU menjadi AQHI.....	23
Skrip 4. 2 Analisis Korelasi	24
Skrip 4. 3 <i>Pairwise Relationships</i>	26
Skrip 4. 4 Analisis Deskriptif Numerik.....	27
Skrip 4. 5 Fitur Kategorikal	29
Skrip 4. 6 Penanganan <i>Outliers</i>	32
Skrip 4. 7 Pemisahan <i>Data Training</i> dan <i>Testing</i>	34
Skrip 4. 8 <i>Model Training</i>	35
Skrip 4. 9 Evaluasi Model.....	36

DAFTAR LAMPIRAN

Lampiran 1 Alokasi Waktu Penelitian	43
Lampiran 2 Surat Izin Penelitian Universitas Bakrie.....	44
Lampiran 3 Surat Pemberian Izin oleh Dinas Lingkungan Hidup Jakarta.....	45