

**ANALISIS PERBANDINGAN KLASIFIKASI
KUALITAS UDARA DI JAKARTA MENGGUNAKAN
ALGORITMA *K-NEAREST NEIGHBOR* (KNN) DAN
RANDOM FOREST CLASSIFIER.**

SKRIPSI SARJANA SISTEM INFORMASI

Oleh :

Nicholas Wiriastama

217006516016



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL
2024 - 2025**

**ANALISIS PERBANDINGAN KLASIFIKASI
KUALITAS UDARA DI JAKARTA MENGGUNAKAN
ALGORITMA *K-NEAREST NEIGHBOR* (KNN) DAN
RANDOM FOREST CLASSIFIER.**

SKRIPSI SARJANA SISTEM INFORMASI

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem
Informasi dari Fakultas Teknologi Komunikasi dan Informatika

Oleh :

Nicholas Wiriastama

217006516016



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL
2024 - 2025**

HALAMAN PENGESAHAN

TUGAS AKHIR

**ANALISIS PERBANDINGAN KLASIFIKASI
KUALITAS UDARA DI JAKARTA MENGGUNAKAN
ALGORITMA K-NEAREST NEIGHBOR DAN
RANDOM FOREST CLASSIFIER**



Nicholas Wiriastama

217006516016

Dosen Pembimbing

(Ir. Endah Tri Esti Handayani, MMSI)

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

ANALISIS PERBANDINGAN KLASIFIKASI KUALITAS UDARA DI JAKARTA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN RANDOM FOREST CLASSIFIER

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 28 Februari 2025



Nicholas Wiriastama

217006516016

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

ANALISIS PERBANDINGAN KLASIFIKASI KUALITAS UDARA DI JAKARTA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN RANDOM FOREST CLASSIFIER

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2024-2025 pada tanggal 26 Februari Tahun 2025

Dosen Pembimbing

Ir. Endah Tri Esti Handayani, MMSI

NIDN. 0110140842

Ketua Program Studi

Dr. Andrianingsih, S.Kom., MMSI

NIDN. 030309790

**LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG
DIREVISI**

Nama : Nicholas Wiriastama
NPM : 217006516016
Fakultas/Akâdemi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : 26 Februari 2025

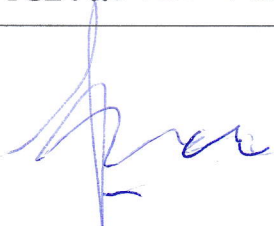
JUDUL DALAM BAHASA INDONESIA :

ANALISIS PERBANDINGAN KLASIFIKASI KUALITAS UDARA DI JAKARTA
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN RANDOM
FOREST CLASSIFIER

JUDUL DALAM BAHASA INGGRIS :

COMPARATIVE ANALYSIS OF AIR QUALITY CLASSIFICATION IN JAKARTA
USING K-NEAREST NEIGHBOR AND RANDOM FOREST CLASSIFIER
ALGORITHMS

TANDA TANGAN DAN TANGGAL

Pembimbing	Ka. Prodi	Mahasiswa
TGL : 28 Februari 2025	TGL : 28 Februari 2025	TGL : 28 Februari 2025
 (Dr. Endah Tri Esti Hidayati, M.Si)	 (Dr. A. Andrianingsih, M.Si)	 (Nicholas Wiriastama)

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, saya yang bertanda tangan di bawah ini:

Nama : Nicholas Wiriastama

NIM : 217006516016

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Fakultas Teknologi Komunikasi dan Informatika, Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalti Free Right*) atas karya ilmiah saya yang berjudul:

ANALISIS PERBANDINGAN KLASIFIKASI KUALITAS UDARA DI JAKARTA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN RANDOM FOREST CLASSIFIER.

Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Fakultas Teknologi Komunikasi dan Informatika berhak menyimpan, mengalih media/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.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 28 Februari 2025

Yang menyatakan



(Nicholas Wiriastama)

KATA PENGANTAR

Segala hormat, puji syukur dan kemuliaan bagi Tuhan Yesus Yang Maha Esa dengan kasih dan penyertaanNya telah menuntun setiap langkah dalam penyusunan skripsi ini. Tanpa kasih dan kekuatanNya, penulis tidak akan mampu menyelesaikan Skripsi dengan judul **“Analisis Perbandingan Klasifikasi Kualitas Udara Di Jakarta Menggunakan Algoritma K-Nearest Neighbor (Knn) Dan Random Forest Classifier”**. Skripsi ini disusun sebagai salah satu syarat kelulusan guna memperoleh gelar Sarjana Strata Satu (S1) dari Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional.

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan dan motivasi dari berbagai pihak selama proses penyusunan skripsi ini. Pada kesempatan ini penulis mengungkapkan rasa terima kasih setulus-tulusnya kepada:

1. Bapak Dr. El Amry Bernawi Putera, M.A. selaku Rektor Universitas Nasional.
2. Ibu Ir. Endah Tri Esti Handayani, MMSI selaku Wakil Dekan juga Dosen Pembimbing Skripsi yang dengan penuh kesabaran telah meluangkan waktu untuk memberikan bimbingan, arahan, dan masukan berharga selama proses penyusunan skripsi ini.
3. Bapak Dr. Agung Triayudi, S.Kom., M.Kom selaku Dekan Fakultas Teknologi Komunikasi dan Informatika dan Ibu Dr. Andrianingsih, S.Kom, MMSI selaku Kaprodi Sisem Informasi.
4. Para Dosen Pembimbing Akademik yang senantiasa memberikan bimbingan, arahan, dan motivasi sepanjang masa perkuliahan.
5. Bapak/Ibu Dosen Pengajar Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional yang telah dengan tulus membagikan ilmu dan wawasan yang bermanfaat selama perkuliahan.

6. Seluruh Staf Sekretariat Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional yang membantu dalam administrasi.
7. Kedua Orang Tua tercinta yang dengan penuh kasih telah memberikan semangat, doa juga dukungan dalam berbagai bentuk yang tak terhitung.
8. Pihak pengurus serta staff YAAB ORBIT HAH yang memberikan dukungan dan bantuan dalam menunjang proses perkuliahan dalam menyelesaikan studi ini.
9. Pengurus dan rekan-rekan Persekutuan Oikoumene Universitas Nasional yang telah menjadi keluarga rohani bagi saya selama masa perkuliahan yang mendukung dan memberikan semangat dalam proses penyusunan skripsi.
10. Rekan seperbimbingan Ibu Endah yang saling mendukung, berbagi pengetahuan, dan memberikan semangat bersama menyelesaikan skripsi, bersamaan dengan rekan mahasiswa angkatan 2021 Program Studi Sistem Informasi Universitas Nasional.

Penulis menyadari bahwasanya masih banyak kekurangan dalam penyajian, pemilihan kata-kata maupun pembahasan hasil skripsi ini. Oleh karena itu, dengan kerendahan hati penulis mengharapkan saran, kritik yang membangun dari semua pihak untuk perbaikan skripsi ini.

Akhir kata, kepada seluruh pihak yang telah memberikan dukungan, bantuan, dan motivasi dalam penyusunan skripsi ini, penulis mengucapkan terima kasih banyak. Semoga skripsi ini dapat memberikan manfaat baik bagi pengembangan ilmu pengetahuan maupun bagi pihak-pihak yang berkepentingan serta dapat menjadi referensi yang berguna dan dapat diterima dengan baik bagi pembaca yang memberikan kontribusi yang berarti dalam bidang analisis data.

Jakarta, Februari 2025

Nicholas Wiriastama

ABSTRAK

Kualitas udara merupakan faktor krusial dalam menjaga kesehatan masyarakat, terutama di kota besar seperti Jakarta, yang menghadapi tantangan serius akibat tingginya aktivitas manusia dan polusi. Penurunan kualitas udara di Jakarta sebagian besar bersumber dari aktivitas masyarakat, yang berpotensi menimbulkan dampak serius terhadap kesehatan, termasuk penyakit pernapasan yang dapat berujung pada kematian. Oleh karena itu, diperlukan upaya analisis kualitas udara secara sistematis. Penelitian ini bertujuan untuk mengklasifikasikan kategori kualitas udara di Jakarta serta membandingkan kinerja kedua algoritma yakni K-Nearest Neighbor (KNN) dan Random Forest. Dataset yang digunakan berupa Indeks Standar Pencemar Udara (ISPU) tahun 2022–2023 dengan enam parameter utama: PM10, PM2.5, SO₂, CO, O₃, dan NO₂ yang dikumpulkan dari stasiun pemantauan di Jakarta. Implementasi dilakukan menggunakan Google Colab dengan evaluasi model berdasarkan metrik akurasi, precision, recall, dan F1-score, serta optimasi menggunakan K-Fold Cross Validation dan GridSearch. Hasil penelitian menunjukkan bahwa model Random Forest memiliki akurasi lebih tinggi, yaitu 99%. Sementara itu, KNN memperoleh akurasi sebesar 96%.

Kata kunci : Data Mining, Klasifikasi, KNN, Random Forest, Kualitas Udara

ABSTRACT

Air quality is a crucial factor in maintaining public health, especially in big cities like Jakarta, which face serious challenges due to high human activity and pollution. The decline in air quality in Jakarta is largely due to community activities, which have the potential to cause serious impacts on health, including respiratory diseases that can lead to death. Therefore, a systematic air quality analysis effort is needed. This study aims to classify air quality categories in Jakarta and compare the performance of the two algorithms, namely K-Nearest Neighbor (KNN) and Random Forest. The dataset used is the Air Pollution Standard Index (ISPU) for 2022–2023 with six main parameters: PM10, PM2.5, SO₂, CO, O₃, and NO₂ collected from monitoring stations in Jakarta. The implementation was carried out using Google Colab with model evaluation based on accuracy, precision, recall, and F1-score metrics, and optimization using K-Fold Cross Validation and GridSearch. The results showed that the Random Forest model had a higher accuracy of 99%. Meanwhile, KNN obtained an accuracy of 96%.

Keywords : Data Mining, Classification, KNN, Random Forest, Air Quality

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah	3
1.5 Kontribusi.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Landasan Teori	5
2.1.1 Data Mining	5
2.1.2 Machine Learning.....	5
2.1.3 Classification	6
2.1.3.1 K-Nearest Neighbor	7
2.1.3.2 Random Forest	8
2.1.4 Google Colab	9

2.1.5 Kualitas Udara	9
2.1.6 Indeks Standar Pencemaran Udara	10
2.2 Studi Literatur.....	11
BAB III METODOLOGI PENELITIAN	17
3.1 Waktu dan Lokasi Penelitian.....	17
3.2 Penentuan Obyek dan Subyek Penelitian.....	17
3.3 Fokus Penelitian	17
3.4 Metode Pengumpulan Data.....	18
3.4.1 Informasi Data.....	19
3.5 Desain Penelitian.....	20
3.6 Metode Klasifikasi.....	23
3.6.1 <i>K-Nearest Neighbor</i>	23
3.6.2 <i>Random Forest</i>	26
3.7 Analisis Evaluasi	29
3.7.1 Confussion Matrix.....	29
3.8 Perangkat Penelitian.....	31
BAB IV HASIL PEMBAHASAN	32
4.1 Pengumpulan Dataset.....	32
4.2 <i>Data Preparation</i>	33
4.3 Perhitungan Manual Algoritma	36
4.3.1 K-NN	36
4.3.2 <i>Random Forest</i>	38
4.4 Penerapan dalam Google Colab	41
4.4.1 Memuat Dataset dan Library	41
4.4.2 Persiapan Data	42

4.4.3 Penerapan <i>K-Nearest Neighbors</i>	46
4.4.4 Penerapan <i>Random Forest</i>	50
4.5 Analisis dan Evaluasi Hasil	53
4.6 Tampilan Dashboard Kualitas Udara.....	57
BAB 5 KESIMPULAN	59
5.1 Kesimpulan	59
5.2 Saran	60
DAFTAR PUSTAKA	61
LAMPIRAN.....	64
Lampiran 1. Hasil Turnitin.....	64
Lampiran 2. Submit Jurnal.....	65
Lampiran 3. Bimbingan	65
DAFTAR GAMBAR	
Gambar 2.1 Ilustrasi KNN	8
Gambar 2.2 Ilustrasi Random Forest	9
Gambar 2.3 Kategori ISPU	10
Gambar 3.1 Sumber Data.....	18
Gambar 3.2 Dataset ISPU	19
Gambar 3.3 Desain Penelitian.....	20
Gambar 3.4 Missing Values	21
Gambar 3.5 Fitur Dataset	22
Gambar 3.6 Tree 1 Random Forest	29
Gambar 4.1 Tampilan Dataset Awa.....	32
Gambar 4.2 Tabel Missing Values.....	34
Gambar 4.3 Tabel Tidak ada Data.....	34
Gambar 4.4 Tabel Data Stasiun.....	35

Gambar 4.5 Tabel Fitur Dataset	35
Gambar 4.6 Tampilan Akhir Dataset.....	36
Gambar 4.7 Library Yang Diperlukan.....	41
Gambar 4.8 Pemilihan Kolom	42
Gambar 4.9 Cleaning Data.....	42
Gambar 4.10 Cek Tipe Data.....	43
Gambar 4.11 Transformasi Tipe Data	44
Gambar 4.12 Ubah Tipe Data	44
Gambar 4.13 Hasil Transformasi Data.....	44
Gambar 4.14 Data Tidak Seimbang	45
Gambar 4.15 Hasil Penyeimbangan Data	46
Gambar 4.16 Library Penerapan Model.....	46
Gambar 4.17 Membagi X dan Y	47
Gambar 4.18 Split dan Scalling	47
Gambar 4.19 Penerapan KNN	48
Gambar 4.20 Performa KNN	48
Gambar 4.21 K-Fold KNN.....	49
Gambar 4.22 K Optimal.....	50
Gambar 4.23 KNN Setelah K-Fold.....	50
Gambar 4.24 Penerapan Random Forest.....	51
Gambar 4.25 Performa Random Forest	51
Gambar 4.26 RF dengan GridSearch.....	52
Gambar 4.27 Hasil Klasifikasi	53
Gambar 4.28 Parameter Pencemar Dominan	54
Gambar 4.29 Confusion Matrix KNN.....	55
Gambar 4.30 Confusion Matrix Random Forest.....	56
Gambar 4. 31 Dashboard ISPU	58

DAFTAR TABEL

Tabel 2.1 Studi Literatur	11
Tabel 3.1 Timeline.....	17

Tabel 3.2 Deskripsi Dataset.....	19
Tabel 3.3 Dataset Sampel KNN	25
Tabel 3.4 Data Test KNN	25
Tabel 3.5 Perhitungan Jarak KNN	25
Tabel 3.6 Dataset Sampel Random Forest	27
Tabel 3.7 Sampling Acak Random Forest.....	27
Tabel 3.8 Tabel Kontigensi Random Forest	28
Tabel 3.9 Tabel Kontigensi Node 1	28
Tabel 3.10 Data Test Random Forest	29
Tabel 3.11 Confussion Matrix	30
Tabel 3.12 Perangkat Penelitian	31
Tabel 4.1 Atribut Dataset.....	33
Tabel 4.2 Sampel KNN	36
Tabel 4.3 Data Tes KNN	37
Tabel 4.4 Tabel Penghitungan Jarak.....	38
Tabel 4.5 Sampel Random Forest	38
Tabel 4.6 Sampling Acak Random Forest.....	39
Tabel 4.7 Tabel Kontigensi Root Random Forest	39
Tabel 4.8 Tabel Kontigensi Node 1 Random Forest	39
Tabel 4.9 Data Test Random Forest	40