

**KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA  
(ISPU) DI DKI JAKARTA MENGGUNAKAN ALGORITMA  
*RANDOM FOREST***

**SKRIPSI**



Oleh :  
**DIMAS DZAKY DANISWARA**  
**20083010006**

**PROGRAM STUDI SAINS DATA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR**

**2024**

**CLASSIFICATION OF AIR POLLUTION STANDARD INDEX  
(ISPU) IN DKI JAKARTA USING RANDOM FOREST  
ALGORITHM**

**UNDERGRADUATES THESIS**



**By :  
DIMAS DZAKY DANISWARA  
20083010006**

**STUDY PROGRAM OF DATA SCIENCE  
FACULTY OF COMPUTER SCIENCE  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR**

**2024**

**LEMBAR PENGESAHAN  
SKRIPSI**

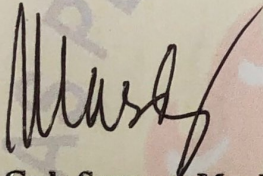
JUDUL : KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA  
(ISPU) DI DKI JAKARTA MENGGUNAKAN ALGORITMA  
*RANDOM FOREST*

DISUSUN OLEH : DIMAS DZAKY DANISWARA  
NPM : 20083010006

Telah diseminarkan dalam Ujian Skripsi pada:  
hari Selasa, tanggal 2 Januari 2024

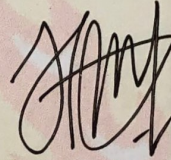
Menyetujui,

Dosen Pembimbing 1



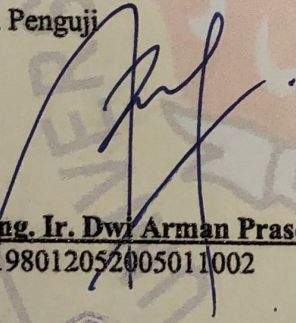
Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU  
NIP. 197006192021211009

Dosen Pembimbing 2



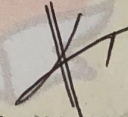
Aviolla Terza Damaliana, S.Si., M.Stat  
NIP. 199408022022032015

Dosen Penguji



Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU  
NIP. 198012052005011002

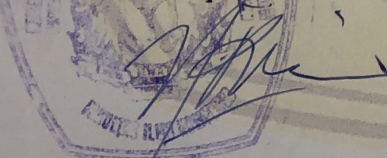
Dosen Penguji



Kartika Maulida Hindravani, S.Kom., M.Kom  
NIP. 199209092022032009

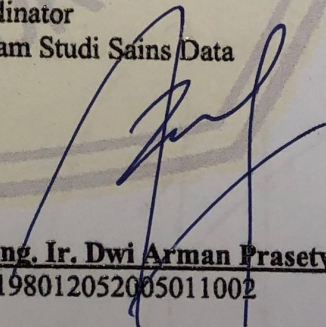
Mengetahui,

Dekan  
Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT  
NIP. 19681126199403 2001

Koordinator  
Program Studi Sains Data



Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU  
NIP. 198012052005011002



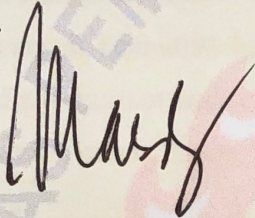
**APPROVAL SHEET  
UNDERGRADUATES THESIS**

**TITLE** : CLASSIFICATION OF AIR POLLUTION STANDARD INDEX (ISPU) IN DKI JAKARTA USING RANDOM FOREST ALGORITHM  
**STUDENT NAME** : DIMAS DZAKY DANISWARA  
**NPM** : 20083010006

Has been presented in the Undergraduates Thesis exam on:  
day Tuesday, date 2<sup>nd</sup> January 2024

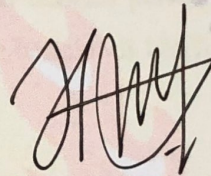
Approval,

1<sup>st</sup> Advisor



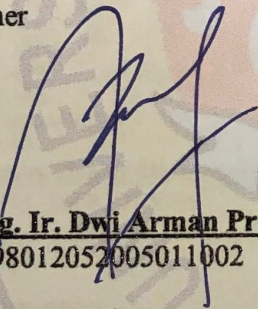
**Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU**  
NIP. 197006192021211009

2<sup>nd</sup> Advisor/ Co-Advisor



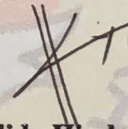
**Aviolla Terza Damaliana, S.Si., M.Stat**  
NIP. 199408022022032015

Examiner



**Dr. Eng. Ir. Dwi Arman Prasetva, ST., MT., IPU**  
NIP. 198012052005011002

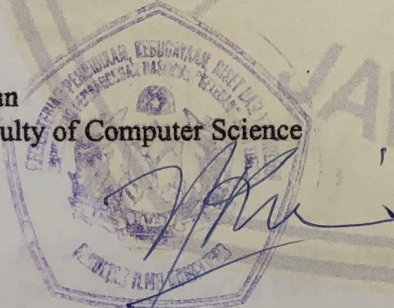
Examiner



**Kartika Maulida Hindrayani, S.Kem., M.Kom**  
NIP. 199209092022032009

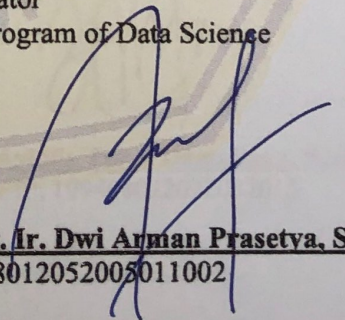
Acknowledged,

Dean  
Faculty of Computer Science



**Prof. Dr. Ir. Novirina Hendrasarie, MT**  
NIP. 19681126199403 2001

Coordinator  
Study Program of Data Science



**Dr. Eng. Ir. Dwi Arman Prasetva, ST., MT., IPU**  
NIP. 198012052005011002



## PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

Nama Lengkap Mahasiswa : Dimas Dzaky Daniswara  
NPM : 20083010006  
Program Studi : Sains Data  
Dosen Pembimbing : Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU  
Aviolla Terza Damaliana, S.Si., M.Stat

dengan ini menyatakan bahwa Skripsi dengan judul "**KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA (ISPU) DI DKI JAKARTA MENGGUNAKAN ALGORITMA *RANDOM FOREST***" adalah hasil karya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Surabaya, 12 Januari 2024

Mahasiswa,



**Dimas Dzaky Daniswara**  
NPM. 20083010006

Mengetahui,  
Dosen Pembimbing 1

**Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU**  
NIP. 197006192021211009

Dosen Pembimbing 2

**Aviolla Terza Damaliana, S.Si., M.Stat**  
NIP. 199408022022032015

# **KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA (ISPU) DI DKI JAKARTA MENGGUNAKAN ALGORITMA *RANDOM FOREST***

Nama : Dimas Dzaky Daniswara  
NPM : 20083010006  
Program Studi : Sains Data  
Dosen Pembimbing : Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU  
Aviolla Terza Damaliana, S.Si., M.Stat

## **ABSTRAK**

Kualitas udara merupakan salah satu faktor penting yang mempengaruhi kesehatan dan kesejahteraan manusia. Udara yang tercemar dapat menyebabkan berbagai penyakit pernapasan, kardiovaskular, dan kanker. Sebagai ibukota negara Indonesia, Jakarta adalah provinsi di Indonesia yang memiliki tingkat pencemaran udara tertinggi, hal ini berdampak pada penurunan kualitas udara di Jakarta. Salah satu cara untuk menginformasikan kualitas udara kepada masyarakat adalah dengan menggunakan Indeks Standar Pencemar Udara (ISPU). ISPU adalah angka tanpa satuan yang menggambarkan kondisi kualitas udara ambien di lokasi tertentu. ISPU dihitung berdasarkan konsentrasi beberapa parameter pencemar udara seperti; partikulat mater (PM10), sulfur dioksida (SO<sub>2</sub>), nitrogen dioksida (NO<sub>2</sub>), karbon monoksida (CO), dan ozon (O<sub>3</sub>). Untuk mengklasifikasikan ISPU, diperlukan suatu metode yang dapat mempelajari pola dari data sensor yang mengukur konsentrasi polutan udara. Oleh karena itu, penulis ingin melakukan penelitian mengenai perbandingan algoritma *Random Forest* dan *Support Vector Machine*. Penelitian ini dilakukan menggunakan data indeks standar pencemaran udara (ISPU) DKI Jakarta tahun 2021 dengan jumlah sebanyak 1837 data. Data yang tersedia dibagi menjadi data *train* dan data *test*, dimana 80% digunakan sebagai data *train* dan 20% sebagai data *test*. Setelah dilakukan pembagian data, dilakukan pelatihan model menggunakan data *train* dan pengujian model menggunakan data *test*. Model juga diuji validasi menggunakan *k-fold cross validation*. Pada uji coba dan *validation* dengan menggunakan algoritma *Random Forest*, diperoleh tingkat akurasi model dan rata-rata *validation* yang unggul dengan persentase 99% dan 95%. Sementara itu, untuk algoritma *Support Vector machine* mendapatkan akurasi model dan rata-rata *validation* dengan presentase 94% dan 91%.

**Kata kunci:** *Pencemaran Udara, ISPU, Klasifikasi, Machine Learning, Support Vector Machine, Random Forest, Confusion Matrix.*

# CLASSIFICATION OF AIR POLLUTION STANDARD INDEX (ISPU) IN DKI JAKARTA USING RANDOM FOREST ALGORITHM

Name : Dimas Dzaky Daniswara  
NPM : 20083010006  
Study Program : Sains Data  
Advisor : Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU  
Aviolla Terza Damaliana, S.Si., M.Stat

## ABSTRACT

Air quality is one of the most important factors affecting human health and well-being. Polluted air can cause various respiratory, cardiovascular and cancer diseases. As the capital city of Indonesia, Jakarta is the province in Indonesia that has the highest level of air pollution, this has an impact on the decline in air quality in Jakarta. One way to inform the public about air quality is by using the Air Pollutant Standard Index (ISPU). ISPU is a unitless number that describes the ambient air quality conditions at a particular location. ISPU is calculated based on the concentration of several air pollutant parameters such as; particulate matter (PM10), sulfur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>), carbon monoxide (CO), and ozone (O<sub>3</sub>). To classify ISPU, a method is needed that can learn patterns from sensor data that measure the concentration of air pollutants. Therefore, the author would like to conduct research on the comparison of Random Forest and Support Vector Machine algorithms. This research was conducted using data on the DKI Jakarta air pollution standard index (ISPU) in 2021 with a total of 1837 data. The available data is divided into train data and test data, where 80% is used as train data and 20% as test data. After data division, model training is carried out using train data and model testing using test data. The model was also tested for validation using k-fold cross validation. In testing and validation using the Random Forest algorithm, the model accuracy rate and validation average are superior with a percentage of 99% and 95%. Meanwhile, the Support Vector Machine algorithm obtained model accuracy and average validation with a percentage of 94% and 91%.

**Keywords:** *Air Pollution, ISPU, Classification, Machine Learning, Support Vector Machine, Random Forest, Confusion Matrix.*

## KATA PENGANTAR

Dengan memanjatkan puji syukur kepada Tuhan Yang Maha Esa, yang telah melimpahkan rahmat dan karunia-Nya kepada kita semua, Penulis dengan senang hati menyampaikan laporan skripsi berjudul "Klasifikasi Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta Menggunakan Algoritma *Random Forest*". Penulis menyadari bahwa laporan ini berhasil disusun berkat bimbingan dan petunjuk dari Tuhan Yang Maha Esa, serta dukungan yang diberikan oleh berbagai pihak, di antaranya:

1. Orang tua dan keluarga selalu memberikan dukungan dan doa.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional "Veteran" Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur
4. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional "Veteran" Jawa Timur.
5. Bapak Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU selaku Dosen Pembimbing 1.
6. Ibu Aviolla Terza Damaliana, S.Si., M.Stat selaku Dosen Pembimbing 2.
7. Bapak dan Ibu dosen Program Studi Sains Data UPN "Veteran" Jawa Timur yang sudah berkenan untuk memberikan waktu untuk berkontribusi dalam penelitian ini.
8. Teman-teman Sains Data angkatan 2020 yang senantiasa memberikan dukungan dan panduan dalam menyelesaikan tugas hingga penyelesaian skripsi ini.

Bantuan dan dukungan dari pihak-pihak tersebut menjadi pilar utama kesuksesan penulis dalam menyelesaikan laporan skripsi ini. Semoga laporan ini bermanfaat dan memberikan kontribusi positif pada pengembangan ilmu pengetahuan. Penulis memohon maaf atas segala kekurangan dalam laporan ini, serta mengharapkan kritik dan saran yang membangun untuk perbaikan di masa mendatang. Terima kasih.

Surabaya, 12 Januari 2024  
Penulis

Dimas Dzaky Daniswara



## DAFTAR ISI

LEMBAR PENGESAHAN .....	iii
APPROVAL SHEET.....	iv
PERNYATAAN ORISINALITAS .....	v
ABSTRAK .....	vi
ABSTRACT .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI .....	ix
DAFTAR TABEL .....	xi
DAFTAR GAMBAR.....	xii
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
BAB 2 TINJAUAN PUSTAKA.....	6
2.1. <i>Review</i> Penelitian Sebelumnya.....	6
2.2. Pencemaran Udara.....	8
2.3. Indeks Standar Pencemaran Udara (ISPU).....	8
2.4. <i>Data Mining</i> .....	10
2.5. Klasifikasi.....	11
2.6. <i>Random Forest</i> (RF).....	12
2.7. <i>Support Vector Machine</i> (SVM).....	14
2.8. <i>Confusion Matrix</i> .....	15

2.9.	<i>K-fold Cross Validation</i> .....	16
BAB 3 METODOLOGI .....		17
3.1.	Sumber Data .....	17
3.2.	Diagram Alir Penelitian.....	18
3.3.	Jadwal Penelitian .....	23
BAB 4 HASIL DAN PEMBAHASAN .....		24
4.1.	Pengumpulan Data.....	24
4.2.	Visualisasi Data .....	24
4.3.	<i>Preprocessing Data</i> .....	27
4.3.1.	Penggabungan Data .....	27
4.3.2.	Seleksi Data .....	31
4.3.3.	<i>Data Cleaning</i> .....	32
4.3.4.	<i>Label Encoding</i> .....	34
4.3.5.	Normalisasi Data .....	35
4.4.	<i>Split Data Train dan Test</i> .....	36
4.5.	Proses Klasifikasi .....	37
4.5.1.	Klasifikasi <i>Random Forest</i> .....	37
4.5.2.	Klasifikasi <i>Support Vector Machine</i> .....	39
4.6.	Evaluasi dan Perbandingan Kinerja Model Klasifikasi.....	42
4.7.	<i>Deployment</i> .....	44
BAB 5 KESIMPULAN DAN SARAN .....		47
5.1.	Kesimpulan.....	47
5.2.	Saran .....	48
DAFTAR PUSTAKA.....		49
LAMPIRAN .....		52
BIODATA PENULIS.....		55

## DAFTAR TABEL

<b>Tabel 2.1.</b> Tabel <i>Review</i> Penelitian Sebelumnya .....	6
<b>Tabel 2.2.</b> Batas ISPU .....	9
<b>Tabel 2.3.</b> Kategori ISPU .....	10
<b>Tabel 2.4.</b> Tabel <i>Confusion Matrix</i> .....	15
<b>Tabel 3.1.</b> Struktur Data ISPU .....	17
<b>Tabel 3.2.</b> Jadwal Kegiatan Penelitian .....	23
<b>Tabel 4.1.</b> Sampel Data ISPU .....	24
<b>Tabel 4.2.</b> Data bulan Juni sebelum perbaikan .....	28
<b>Tabel 4.3.</b> Data bulan Juli sebelum perbaikan .....	28
<b>Tabel 4.4.</b> Data bulan Juni sesudah perbaikan .....	29
<b>Tabel 4.5.</b> Data bulan Juli sesudah perbaikan .....	29
<b>Tabel 4.6.</b> Hasil Penggabungan Data .....	30
<b>Tabel 4.7.</b> Variabel ISPU .....	31
<b>Tabel 4.8.</b> Hasil Seleksi Data .....	32
<b>Tabel 4.9.</b> Dataset sebelum <i>label encoding</i> .....	35
<b>Tabel 4.10.</b> Dataset sesudah <i>label encoding</i> .....	35
<b>Tabel 4.11.</b> Hasil normalisasi .....	36
<b>Tabel 4.12.</b> <i>Confusion Matrix Random Forest</i> .....	38
<b>Tabel 4.13.</b> <i>Confusion Matrix Support Vector Machine</i> .....	40
<b>Tabel 4.14.</b> Perbandingan <i>Accuracy, Precision, Recall, dan F1-score</i> .....	42
<b>Tabel 4.15.</b> Hasil <i>validation</i> .....	43
<b>Tabel 4.16.</b> Tabel sampel pengujian aplikasi .....	45



## DAFTAR GAMBAR

<b>Gambar 3.1.</b> Diagram Alir Penelitian .....	18
<b>Gambar 4.1.</b> Barplot nilai rata-rata parameter per stasiun .....	25
<b>Gambar 4.2.</b> Barplot jumlah kategori dalam setahun.....	26
<b>Gambar 4.3.</b> Barplot nilai rata-rata pencemar udara pada kategori .....	26
<b>Gambar 4.4.</b> Pengecekan nilai kosong .....	32
<b>Gambar 4.5.</b> Pengecekan nilai kosong setelah <i>Cleaning</i> .....	33
<b>Gambar 4.6.</b> Boxplot sebelum <i>handling outlier</i> .....	33
<b>Gambar 4.7.</b> Boxplot sesudah <i>handling outlier</i> .....	34
<b>Gambar 4.8.</b> Hasil pengujian <i>Random Forest</i> .....	37
<b>Gambar 4.9.</b> Hasil pengujian <i>Support Vector Machine</i> .....	40
<b>Gambar 4.10.</b> <i>Deployment</i> .....	44
<b>Gambar 4.11.</b> Uji coba kategori baik .....	45
<b>Gambar 4.12.</b> Uji coba kategori sedang.....	46
<b>Gambar 4.13.</b> Uji coba kategori tidak sehat.....	46