

**IMPLEMENTASI *RANDOM FOREST* UNTUK KLASIFIKASI  
INDEKS STANDAR PENCEMARAN UDARA  
DI DKI JAKARTA**

**SKRIPSI**

*Diajukan sebagai salah satu persyaratan untuk memperoleh gelar  
Sarjana Statistika*



Oleh  
**HANIFA HASNA**  
**NIM. 19337052**

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2024**

## PERSETUJUAN SKRIPSI

### IMPLEMENTASI *RANDOM FOREST* UNTUK KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA DI DKI JAKARTA

Nama : Hanifa Hasna  
NIM : 19337052  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 03 Juni 2024

Mengetahui:  
Kepala Departemen Statistika



Dr. Yenni Kurniawati, S.Si., M.Si  
NIP. 198402232010122005

Disetujui Oleh:  
Pembimbing



Dra. Nonong Amalita, M.Si  
NIP. 196906151993032001


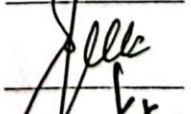

## PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Hanifa Hasna  
NIM : 19337052  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

### IMPLEMENTASI *RANDOM FOREST* UNTUK KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA DI DKI JAKARTA

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi  
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Negeri Padang

Padang, 03 Juni 2024

Tim Penguji	Nama	Tanda Tangan
Ketua	: Dra. Nonong Amalita, M.Si	
Anggota	: Dr. Dony Permana, M.Si	
Anggota	: Admi Salma, M.Si	

## SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

Nama : Hanifa Hasna  
NIM : 19337052  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul **“Implementasi *Random Forest* untuk Klasifikasi Indeks Standar Pencemaran Udara di DKI Jakarta”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan.

Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,  
Kepala Departemen Statistika,



Dr. Yenni Kurniawati, S.Si., M.Si  
NIP. 198402232010122005

Saya yang menyatakan,



Hanifa Hasna  
NIM. 19337052

# Implementasi *Random Forest* untuk Klasifikasi Indeks Standar Pencemaran Udara di DKI Jakarta

Hanifa Hasna

## ABSTRAK

Kehidupan makhluk hidup terkena dampak dari permasalahan pencemaran udara karena banyak penyakit dapat disebabkan oleh udara yang terkontaminasi, maka sangat penting untuk mengawasi tingkat polusi udara di masyarakat. Berdasarkan pengukuran Indeks Standar Pencemaran Udara yang dilakukan Dinas Lingkungan Hidup DKI Jakarta, kualitas udara di DKI Jakarta tergolong sedang hingga tidak sehat. Memburuknya kualitas udara di wilayah metropolitan Jakarta sangat berbahaya bagi manusia dan makhluk hidup. Oleh karena itu, untuk mencegah permasalahan tersebut maka dilakukan klasifikasi kualitas udara berdasarkan kandungan polutan dengan menggunakan *Random Forest* (RF). Tujuan dari penelitian ini untuk mengetahui fitur penting yang terdapat dalam pencemaran udara dan untuk mengetahui hasil akurasi yang dihasilkan oleh metode random forest.

*Random forest* merupakan metode *ensemble* yaitu pengembangan dari berapa metode decision tree dengan menggunakan *bootstrap sampling*. Validasi dari penelitian ini menggunakan tingkat *error rate* OOB terkecil dan *Feature Importance*. Klasifikasi *random forest* pada penelitian ini menggunakan data Indeks Standar Pencemaran Udara DKI Jakarta Tahun 2022 yang diperoleh dari Dinas Lingkungan Hidup dan Kehutanan berdasarkan enam kandungan udara, yaitu partikel debu (PM<sub>10</sub>), partikel debu (PM<sub>2,5</sub>), sulfur dioksida (SO<sub>2</sub>), karbon monoksida (CO), ozon (O<sub>3</sub>), dan nitrogen dioksida (NO<sub>2</sub>).

Hasil penelitian menunjukkan bahwa model *Random Forest* dengan parameter terbaik (5000 pohon dan *mtry*=2) berhasil mengklasifikasikan data ISPU dengan tingkat akurasi sebesar 99,17% dan laju kesalahan sebesar 0,83%. Polutan PM<sub>2,5</sub> teridentifikasi sebagai fitur terpenting yang mempengaruhi kualitas udara di DKI Jakarta, diikuti oleh PM<sub>10</sub>, O<sub>3</sub>, CO, NO<sub>2</sub>, dan SO<sub>2</sub>. Kesimpulan dari penelitian ini menunjukkan bahwa metode *Random Forest* efektif dalam mengklasifikasikan kualitas udara dan dapat digunakan sebagai alat bantu dalam pengambilan keputusan terkait pengendalian polusi udara di DKI Jakarta. Penelitian ini juga memberikan rekomendasi untuk penelitian lebih lanjut dalam penerapan metode *Random Forest* pada kasus-kasus lain yang serupa .

Kata Kunci: *Feature Importance*, OOB, *Random Forest*

# **Implementation of Random Forest for Air Pollution Standard Index Classification in DKI Jakarta**

**Hanifa Hasna**

## **ABSTRACT**

Living creatures are affected by air pollution problems because many diseases can be caused by contaminated air, so it is very important to monitor the level of air pollution in society. Based on Air Pollution Standard Index measurements carried out by the DKI Jakarta Environmental Service, the air quality in DKI Jakarta is classified as moderate to unhealthy. Deteriorating air quality in the Jakarta metropolitan area is very dangerous for humans and living things. Therefore, to prevent this problem, air quality classification is carried out based on pollutant content using Random Forest (RF). The aim of this research is to determine the important features contained in air pollution and to determine the accuracy results produced by the random forest method.

Random forest is an ensemble method, namely the development of several decision tree methods using bootstrap sampling. Validation of this research uses the smallest OOB error rate and Feature Importance. Random forest classification in this study uses data from the 2022 DKI Jakarta Air Pollution Standard Index obtained from the Environment and Forestry Service based on six air contents, namely dust particles (PM10), dust particles (PM2.5), sulfur dioxide (SO<sub>2</sub>), carbon monoxide (CO), ozone (O<sub>3</sub>), and nitrogen dioxide (NO<sub>2</sub>).

The research results show that the Random Forest model with the best parameters (5000 trees and mtry=2) succeeded in classifying ISPU data with an accuracy rate of 99.17% and an error rate of 0.83%. PM<sub>2.5</sub> pollutant was identified as the most important feature affecting air quality in DKI Jakarta, followed by PM<sub>10</sub>, O<sub>3</sub>, CO, NO<sub>2</sub>, and SO<sub>2</sub>. The conclusion of this research shows that the Random Forest method is effective in classifying air quality and can be used as a tool in decision making regarding air pollution control in DKI Jakarta. This research also provides recommendations for further research in applying the Random Forest method to other similar cases.

**Keywords:** Feature Importance, OOB, Random Forest

## KATA PENGANTAR

Segala puji dan rasa syukur penulis ucapkan kepada Allah SWT yang telah memberikan rahmat iman, islam dan kesehatan. Serta telah memberkahi nikmat pikiran sehingga penulis dapat menyelesaikan Skripsi dengan judul “Implementasi *Random Forest* untuk Klasifikasi Indeks Standar Pencemaran Udara di DKI Jakarta”. Skripsi ini merupakan salah satu persyaratan untuk memperoleh gelar Sarjana Statistika. Sholawat dan salam penulis ucapkan untuk Nabi Muhammad Salallahu Alaihi Wassalam.

Dalam penyusunan dan penulisan skripsi ini penulis banyak mendapatkan bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh sebab itu, dalam kesempatan ini penulis dengan senang hati menyampaikan banyak terima kasih kepada:

1. Ibu Dra. Nonong Amalita, M.Si., selaku dosen pembimbing akademik dan pembimbing skripsi yang telah meluangkan banyak waktu dan tenaga untuk memberikan ilmu, saran, dan arahan dalam menyelesaikan penelitian ini.
2. Bapak Dr. Doni Permana, M.Si., dan Ibu Admi Salma, M.Si., Dosen Penguji Skripsi yang telah memberikan masukan dan saran demi kesempurnaan skripsi ini.
3. Ibu Dr. Yenni Kurniawati, M.Si., Kepala Departemen Statistika sekaligus Koordinator Program Studi Sarjana Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.

4. Bapak dan Ibu Dosen dan Karyawan Departemen Statistika FMIPA UNP yang telah membimbing dan berbagi ilmu pengetahuan kepada penulis selama berada di bangku perkuliahan.
5. Teristimewa teruntuk kedua orang tua dan keluarga yang telah memberikan semangat, nasehat, dukungan, dan bantuan moril maupun materi kepada penulis.
6. Semua sahabat, teman, serta rekan-rekan seperjuangan yang memberikan semangat kepada penulis.

Semoga dorongan, bimbingan dan bantuan yang diberikan kepada penulis menjadi amal ibadah yang diridhai Allah SWT. Penulis menyadari bahwa skripsi ini masih terdapat kekurangan. Oleh karena itu, masukan dan saran yang bersifat membangun dari berbagai pihak sangat diharapkan demi kesempurnaan penulisan dimasa mendatang. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan sumbangan pemikiran bagi setiap pembaca dan dapat bermanfaat bagi semua pihak.

Padang, Mei 2024

Penulis

Hanifa Hasna



## DAFTAR ISI

<b>ABSTRAK</b> .....	i
<b>KATA PENGANTAR</b> .....	iii
<b>DAFTAR ISI</b> .....	v
<b>DAFTAR TABEL</b> .....	vii
<b>DAFTAR GAMBAR</b> .....	viii
<b>DAFTAR LAMPIRAN</b> .....	ix
<b>BAB I PENDAHULUAN</b> .....	1
A. Latar Belakang .....	1
B. Batasan Masalah.....	7
C. Rumusan Masalah .....	7
D. Tujuan Penelitian .....	7
E. Manfaat Penelitian .....	8
<b>BAB II KERANGKA TEORITIS</b> .....	9
A. <i>Recursive Partitioning</i> atau <i>Tree Based Method</i> .....	9
B. <i>Classification and Regression Tree (CART)</i> .....	12
C. <i>Algoritma Random Forest</i> .....	16
D. Pengukuran Tingkat Akurasi Klasifikasi .....	24
E. Indeks Standar Pencemaran Udara (ISPU) .....	26
<b>BAB III METODOLOGI PENELITIAN</b> .....	30
A. Jenis Penelitian.....	30
B. Jenis dan Sumber Data Penelitian .....	30
C. Variabel Penelitian .....	30
D. Struktur Data Penelitian .....	31

E. Teknik Analisis Data.....	31
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>34</b>
A. Eksplorasi Data .....	34
B. Pembentukan Pohon.....	36
C. Fitur <i>Importance</i> (Fitur Terpenting).....	37
D. Confusion Matrix .....	39
E. Pembahasan.....	40
<b>BAB V PENUTUP.....</b>	<b>43</b>
A. Kesimpulan .....	43
B. Saran.....	43
<b>DAFTAR PUSTAKA .....</b>	<b>44</b>
<b>LAMPIRAN.....</b>	<b>44</b>

## DAFTAR TABEL

Tabel	Halaman
1. Kategori Indeks Standar Pencemaran Udara (ISPU) .....	3
2. Istilah Pada Decision Tree .....	10
3. Perbedaan Algoritma CART dan Random Forest.....	16
4. Kriteria Tingkat Kesalahan Prediksi .....	23
5. Kriteria Tingkat Akurasi Prediksi .....	23
6. Confusion Matrix Klasifikasi Dua Kelas .....	25
7. Klasifikasi Akurasi.....	26
8. Rentang Indeks Standar Pencemaran Udara .....	28
9. Nilai Batas Indeks Standar Pencemaran Udara .....	29
10. Variabel Penelitian.....	30
11. Struktur Data Penelitian .....	31
12. Analisis Deskriptif Variabel Prediktor.....	34
13. Tingkat Akurasi Setiap Ntree.....	36
14. Feature Importance Scores.....	37
15. Confusion Matrix .....	39
16. Recall, Specificity, dan Accuracy .....	40

## DAFTAR GAMBAR

Gambar	Halaman
1. Kondisi Pencemaran Udara di DKI Jakarta tahun 2022 .....	4
2. Struktur Decision Tree .....	11
3. Ilustrasi Struktur Pohon CART.....	13
4. Ilustrasi Struktur Pohon CART.....	15
5. Ilustrasi Struktur Random Forest .....	17
6. Pembagian Sampel Bootstrap dan OOB pada Data Sampel Asli. ....	21
7. Diagram Alir Algoritma Random Forest .....	33
8. Diagram Lingkaran ISPU DKI Jakarta tahun 2022 .....	35

## DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Indeks Standar Pencemaran Udara DKI Jakarta tahun 2022.....	48
2. Syntax Random Forest menggunakan software Python .....	57
3. Output Random Forest.....	59

## **BAB I PENDAHULUAN**

### **A. Latar Belakang**

Udara memiliki peranan penting untuk segala aspek makhluk hidup di bumi, baik itu manusia, hewan atau tumbuhan. Udara merupakan sumber yang tidak terbatas dan sumber daya milik bersama yang dapat memengaruhi kehidupan manusia serta makhluk hidup lainnya. Tanpa adanya udara, makhluk hidup tidak ada yang bisa bertahan hidup. Udara adalah suatu campuran gas alam yang mengelilingi permukaan bumi. Udara tersusun dari banyaknya gas, diantaranya Nitrogen 78%, Oksigen 20%, Argon 0,93%, Karbon Dioksida 0,30%, dan gas-gas lainnya (Fardiaz, 1992). Udara bersih merupakan udara yang tidak bercampur dengan zat-zat atau gas-gas yang merugikan seperti debu, Karbon Dioksida (CO<sub>2</sub>), Nitrogen Dioksida (NO<sub>2</sub>), dan gas lainnya (Amalia dkk, 2022).

Menurut Peraturan Pemerintah (PP) No. 41 Tahun 1999 menyatakan bahwa udara dapat bermanfaat bagi pelestarian fungsi lingkungan hidup, sehingga udara perlu dipelihara, dijaga dan dijamin mutunya melalui pengendalian pencemaran udara. Pencemaran udara diartikan dengan turunnya kualitas udara sehingga udara mengalami penurunan mutu dalam penggunaannya. Kecenderungan menurunnya kualitas udara dapat diakibatkan oleh faktor alam dan manusia. Faktor alam yang dapat mencemari udara salah satunya adalah aktivitas vulkanik gunung. Sedangkan faktor manusia yang paling dominan adalah aktivitas transportasi dan perindustrian seiring dengan pertumbuhan penduduk (Agista dkk, 2020). Berdasarkan pemantauan kualitas udara yang dilakukan oleh Amerika Serikat *Air Quality Index*

(AQI) pada Juni 2022 DKI Jakarta menduduki *ranking* pertama kota paling berpolusi di Indonesia (Nurdalia, 2023).

DKI Jakarta adalah ibu kota negara Indonesia sebagai pusat bisnis, pemerintahan, dan kebudayaan. Provinsi DKI Jakarta terbagi menjadi 5 wilayah kotamadya dan 1 kabupaten administratif, yaitu Kota Jakarta Utara, Jakarta Selatan, Jakarta Barat, Jakarta Timur, Jakarta Pusat dan Kabupaten Kepulauan Seribu. DKI Jakarta merupakan tempat berdirinya kantor-kantor pusat Badan Usaha Milik Negara (BUMN), perusahaan swasta, dan perusahaan asing yang dapat menyebabkan turunnya kualitas udara melalui kegiatan penduduk, kegiatan perindustrian dan transportasi. Kualitas udara Jakarta yang buruk menjadi perbincangan publik dan menghiasi pemberitaan media massa.

Industri manufaktur adalah industri yang terkait dengan produksi barang fisik melalui pengolahan bahan mentah menjadi produk jadi menggunakan mesin, tenaga kerja, dan proses produksi yang terstruktur (Badan Pusat Statistik, 2022). Aktivitas industri manufaktur memberikan dampak kepada lingkungan masyarakat seperti asap pabrik dan limbah kimia lainnya. Dinas Lingkungan Hidup DKI Jakarta menyampaikan bahwa transportasi menjadi penyumbang polusi udara di Jakarta (CNBC Indonesia, 2023). Tercatat dalam “Jakarta Dalam Angka 2021” terjadi peningkatan pada jumlah kendaraan sebesar 3%-6% per tahunnya. Menurut Badan Pusat Statistik (BPS) terjadi peningkatan transportasi di DKI Jakarta sebesar 850.554 kendaraan pada tahun 2022 dibandingkan dengan tahun 2021.

Dinas Lingkungan Hidup (DLH) DKI Jakarta membuat 5 stasiun pemantauan kualitas udara di wilayah DKI Jakarta yaitu Stasiun Bundaran HI, Jagakarsa, Kebun

Jeruk, Kelapa Gading, dan Lubang Buaya. Stasiun tersebut berguna untuk memantau kandungan polutan utama penyebab udara tercemar. Kementerian Lingkungan Hidup dan Kehutanan (KLHK) telah mengeluarkan Peraturan Menteri LHK No.14 Tahun 2020 tentang Indeks Standar Pencemaran Udara (ISPU). Terdapat enam kandungan polutan penyebab udara tercemar yaitu Partikulat ( $PM_{10}$ ), Partikulat ( $PM_{2,5}$ ), Nitrogen Dioksida ( $NO_2$ ), Sulfur Dioksida ( $SO_2$ ), Karbon Monoksida (CO), dan Ozon ( $O_3$ ).

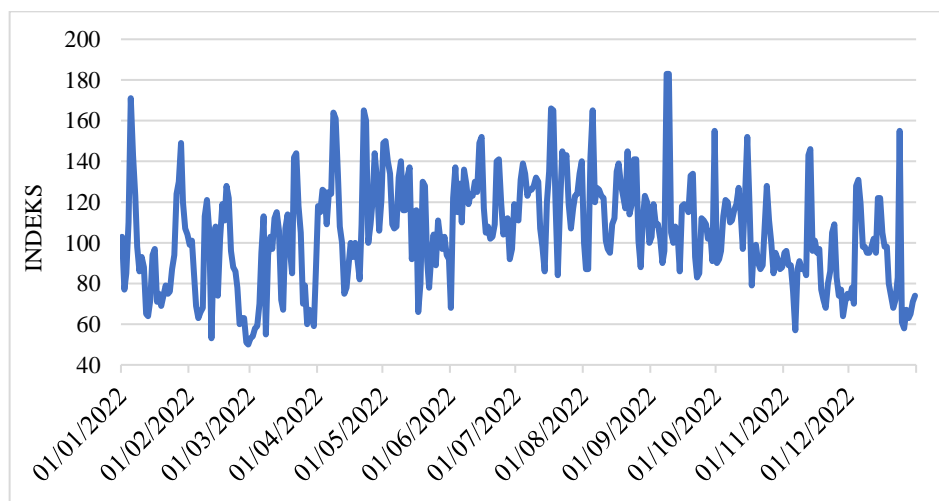
Partikulat (PM) adalah partikel yang ditemukan di udara, termasuk debu, jelaga, kotoran, asap, dan tetesan cairan. Sulfur Dioksida ( $SO_2$ ) merupakan senyawa yang berasal dari pembakaran industri minyak bumi, batu bara, letusan gunung merapi dan diperoleh juga dari oksidasi bijih sulfida. Karbon monoksida (CO) merupakan hasil pembakaran tidak sempurna antara karbon dengan oksigen pada pembakaran batu bara, minyak bumi dan gas pada industri. Ozon ( $O_3$ ) merupakan bentuk gas yang sangat reaktif. Sumber utama ozon dari kegiatan manusia dalam ruangan berasal dari mesin fotokopi, pembersih udara, elektrostatis, dan udara luar. Nitrogen dioksida kebanyakan terdapat dari hasil asap kendaraan bermotor dan bahan bakar industri (Simarmata dkk, 2022). Kandungan polutan penyebab udara tercemar dapat dikategorikan berdasarkan Indeks Standar Pencemaran Udara yang dapat dilihat pada Tabel 1.

**Tabel 1.** Kategori Indeks Standar Pencemaran Udara (ISPU)

<b>Indeks</b>	<b>Kategori</b>
1-50	Baik
51-100	Sedang
101-200	Tidak Sehat
201-300	Sangat Tidak Sehat
$\geq 301$	Berbahaya



Kementerian Lingkungan Hidup dan Kehutanan (KLHK) mengatakan bahwa ISPU merupakan angka tanpa satuan yang digunakan untuk menggambarkan kondisi mutu udara ambien di lokasi tertentu berdasarkan dampak terhadap kesehatan manusia dan makhluk hidup lainnya. Tujuan disusunnya ISPU agar memberikan kemudahan dari keseragaman informasi mutu udara ambien kepada masyarakat di lokasi dan waktu tertentu serta sebagai bahan pertimbangan dalam melakukan upaya pengendalian pencemaran udara baik bagi pemerintah pusat maupun daerah. *Air Quality Indeks (AQI) Jakarta* mengatakan bahwa udara di Jakarta termasuk ke dalam kondisi tidak sehat sehingga disarankan untuk mengenakan masker saat bepergian. Oleh karena itu, dilakukan visualisasi kondisi kualitas udara di DKI Jakarta yang disajikan pada Gambar 2.



(Sumber: Dinas Lingkungan Hidup, 2022)

**Gambar 1.** Kondisi Pencemaran Udara di DKI Jakarta tahun 2022

Berdasarkan Gambar 1, ISPU DKI Jakarta tertinggi sebesar 183 terjadi pada tanggal 8 dan 9 September 2022 yang berada di rentang 101-200 termasuk dalam kategori “Tidak Sehat”. Sedangkan ISPU DKI Jakarta terendah sebesar 50 terjadi pada tanggal 28 Februari 2022 termasuk dalam kategori “Sehat”. Kategori “Sehat”

hanya satu hari terjadi selama setahun, selebihnya kategori “Tidak Sehat” yang mendominasi udara DKI Jakarta tahun 2022. Oleh karena itu, analisis kualitas udara di DKI Jakarta perlu dilakukan. Analisis ini dilakukan untuk mengidentifikasi kualitas udara berdasarkan polutan pencemaran udara seperti Partikulat ( $PM_{10}$ ), Partikulat ( $PM_{2,5}$ ), Nitrogen Dioksida ( $NO_2$ ), Sulfur Dioksida ( $SO_2$ ), Karbon Monoksida (CO), dan Ozon ( $O_3$ ).

Penelitian terkait permasalahan ISPU DKI Jakarta telah dilakukan Nurdalia (2023) yaitu mengklasifikasi ISPU di DKI Jakarta tahun 2021 menggunakan algoritma *naive bayes* dan *K-Nearest Neighbor* (KNN). Hasil akurasi yang diperoleh sebesar 91,78% untuk algoritma *naive bayes* dan 87,67% untuk algoritma KNN. Selanjutnya, Penelitian Kirono (2022) melakukan klasifikasi terhadap ISPU di DKI Jakarta tahun 2021 menggunakan algoritma *naive bayes* menghasilkan akurasi sebesar 91%. Berdasarkan penelitian terkait ISPU DKI Jakarta terdahulu, maka dilakukan pengklasifikasian terhadap ISPU DKI Jakarta tahun 2022. Algoritma yang digunakan dalam penelitian ini yaitu algoritma *random forest*.

*Random forest* adalah salah satu algoritma pembelajaran mesin yang sering digunakan dalam tugas klasifikasi. Klasifikasi adalah salah satu tugas umum dalam pembelajaran mesin yang bertujuan untuk memprediksi kelas atau label dari data yang diberikan berdasarkan fitur-fitur yang ada. Klasifikasi adalah proses menempatkan objek atau konsep tertentu ke dalam satu *set* kategori berdasarkan sifat objek atau konsep yang bersangkutan (Larose, 2014). *Random forest* merupakan sebuah metode *ensemble*. Metode *ensemble* merupakan cara untuk meningkatkan akurasi metode klasifikasi dengan cara mengkombinasikan metode

klasifikasi (Han, 2012). *Random forest* digunakan untuk menghasilkan model prediksi yang akurat dan stabil.

*Random forest* membentuk hasil model yang memberikan informasi terkait hubungan variabel prediktor terhadap klasifikasi serta menjabarkan secara spesifik nilai dalam suatu variabel yang memberikan pengaruh terhadap klasifikasi. *Random forest* dapat memberikan hasil *error* yang lebih rendah, memberikan hasil yang bagus dalam klasifikasi, dapat mengatasi data *training* dalam jumlah sangat besar secara efisien, dan metode yang efektif untuk mengestimasi data hilang (Breiman, 2001). Oleh karena itu, *random forest* sangat cocok digunakan dalam masalah klasifikasi.

Penelitian terkait implementasi algoritma *random forest* telah dilakukan Primajaya (2018) yaitu pengklasifikasian untuk memprediksi curah hujan dengan menggunakan metode *random forest*. Hasil akurasi yang diperoleh dari metode *random forest* sebesar 99,45%. Selanjutnya, penelitian Syukron (2018) yang membuat sistem pengklasifikasian untuk penilaian kredit dengan menggunakan *dataset German Credit*. Metode yang diterapkan pada sistem tersebut adalah *random over-under sampling random forest* yang dapat meningkatkan kinerja akurasi sebesar 14,1% dengan nilai akurasi sebesar 0,901 atau 90,1%.

Berdasarkan paparan yang telah diuraikan, maka dilakukan penelitian mengenai pengklasifikasian *random forest* terhadap Indeks Standar Pencemaran Udara (ISPU) DKI Jakarta tahun 2022, sehingga penelitian ini berjudul **“Implementasi *Random Forest* untuk Klasifikasi Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta”**.

## **B. Batasan Masalah**

Batasan pada penelitian ini adalah klasifikasi Indeks Standar Pencemaran Udara (ISPU) DKI Jakarta tahun 2022 yang menggambarkan data harian. Algoritma yang digunakan adalah algoritma *random forest* dengan polutan udara dalam perhitungan ISPU yaitu Partikulat (PM<sub>10</sub>), Partikulat (PM<sub>2,5</sub>), Nitrogen Dioksida (NO<sub>2</sub>), Sulfur Dioksida (SO<sub>2</sub>), Karbon Monoksida (CO), dan Ozon (O<sub>3</sub>).

## **C. Rumusan Masalah**

Rumusan masalah dari penelitian ini sebagai berikut.

1. Bagaimana urutan fitur terpenting dalam mengklasifikasikan ISPU di DKI Jakarta tahun 2022 dengan menggunakan metode *random forest*?
2. Berapa tingkat kesalahan prediksi dan tingkat akurasi yang dihasilkan metode *random forest* dalam mengklasifikasi ISPU di DKI Jakarta tahun 2022?

## **D. Tujuan Penelitian**

Adapun tujuan dari penelitian ini sebagai berikut.

1. Mengetahui urutan fitur terpenting yang menjadi polutan utama pemicu terjadinya pencemaran udara di DKI Jakarta tahun 2022.
2. Mengetahui tingkat kesalahan prediksi dan tingkat akurasi yang dihasilkan oleh metode *random forest* dalam mengklasifikasi ISPU di DKI Jakarta tahun 2022.

### **E. Manfaat Penelitian**

Adapun manfaat yang diharapkan dari hasil penelitian ini sebagai berikut.

1. Bagi penulis, penelitian ini dapat menjadi pengalaman dalam menganalisis data serta menambah ilmu dan pemahaman dalam melakukan penerapan metode *random forest* terhadap permasalahan pada suatu data.
2. Bagi pemerintah, penelitian ini diharapkan dapat menjadi salah satu pertimbangan pemerintah dalam membuat suatu kebijakan terkait penanggulangan dan pengurangan pencemaran udara berdasarkan klasifikasi ISPU di DKI Jakarta tahun 2022.
3. Bagi pembaca, penelitian ini diharapkan dapat menjadi pedoman atau referensi sebagai alat pertimbangan dalam melakukan penelitian sejenis serta menambah wawasan tentang kajian mengenai metode terkait.