

## Analisis Polusi Udara dan Prediksi Pemodelan

Disusun Oleh : Kimberly Kayla Dewi, David Paul Ong, Fiona Maharani Nugraha,  
Jennifer Patricia, Wella Priscillia.

---

### RINGKASAN

Penelitian ini menyoroti peningkatan polusi udara di Jakarta dengan latar belakang tingginya tingkat pencemaran udara dan dampak seriusnya terhadap kesehatan masyarakat. Tujuan utama penelitian ini adalah mengembangkan model prediktif menggunakan metode *machine learning* untuk memproyeksikan tingkat pencemaran udara di masa depan. Dalam tinjauan pustaka, dipaparkan konsep *machine learning*, penggunaan bahasa pemrograman Python, serta beberapa algoritma metode *machine learning modelling* seperti *linear regression*, *K-Nearest Neighbors*, *decision tree*, dan *random forest*. Metode riset mencakup pengumpulan dan analisis dataset kualitas udara, pengembangan model prediktif, pengujian, evaluasi, dan analisis error. Hasilnya menunjukkan bahwa model prediktif ini dapat memberikan estimasi akurasi dari hasil metrik *Mean Squared Error (MSE)* dan *Root Squared Mean Error (RMSE)* terhadap tingkat pencemaran udara. Potensi khusus penelitian ini terletak pada kontribusinya terhadap mitigasi polusi udara dan peningkatan kesadaran masyarakat. Penutup menyimpulkan bahwa model ini efektif dalam memberikan proyeksi kualitas udara di masa depan, mendukung kebijakan pengendalian polusi udara, dan menyoroti perlunya tindakan pencegahan yang lebih proaktif.

Kata Kunci: Polusi, Pembelajaran Mesin, Kecerdasan Buatan, Jakarta, dan Python.

## DAFTAR ISI

DAFTAR ISI	i
<b>BAB 1. PENDAHULUAN</b>	1
1.1 Latar Belakang	1
1.2 Tujuan	1
<b>BAB 2. TINJAUAN PUSTAKA</b>	2
2.1 <i>Machine Learning Modelling</i>	2
2.2 Python	2
2.3 <i>Linear Regression</i>	2
2.4 <i>K-Nearest Neighbors</i>	3
2.5 <i>Decision Tree</i>	3
2.6 <i>Random Forest</i>	3
<b>BAB 3. METODE RISET</b>	4
3.1 Jenis Penelitian	4
3.2 Tahapan Penelitian	4
<b>BAB 4. HASIL YANG DICAPAI DAN POTENSI KHUSUS</b>	6
4.1 Hasil yang Dicapai	6
4.2 Potensi Khusus	8
<b>BAB 5. PENUTUP</b>	9
<b>DAFTAR PUSTAKA</b>	10
<b>LAMPIRAN</b>	11
Lampiran 1. Formulir Penilaian Laporan Akhir	11

## **BAB 1**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Polusi udara merupakan pencemaran udara yang dikarenakan oleh makhluk hidup, zat, maupun komponen yang lainnya yang masuk ke dalam udara. Polusi dapat dikarenakan oleh limbah yang diproduksi oleh pabrik maupun pembakaran bahan bakar yang diproduksi oleh kendaraan.

Polusi udara mempunyai banyak dampak ke berbagai hal, seperti terhadap flora dan fauna, lingkungan, dan juga terhadap makhluk hidup. Polusi mempunyai dampak yang besar bagi manusia, yaitu seperti penyakit yang menyerang saluran pernafasan, menyebabkan penyakit jantung, maupun kanker.

Akhir-akhir ini, tingkat polusi di Jakarta mulai meningkat, dan kasus masyarakat yang terkena penyakit pernapasan mulai meningkat. Menurut IQAir, Jakarta telah mencapai peringkat ke-35 dengan nilai 72, yang berarti bahwa kualitas udara di Jakarta sedang. Pencemaran udara berdampak sangat serius pada kesehatan masyarakat yang tinggal di Jakarta. Ini juga berdampak pada lingkungan Jakarta, seperti penurunan kualitas air dan tanah dan kerusakan ekosistem.

Dengan mengembangkan model prediktif, penggunaan *machine learning* pada dataset kualitas udara dapat memberikan landasan untuk mengembangkan model yang dapat memprediksi tingkat pencemaran udara di Jakarta di masa depan berdasarkan tren historis.

#### **1.2 Tujuan**

Tujuan utama dari penelitian ini adalah mengembangkan model prediktif menggunakan pendekatan *machine learning* untuk memproyeksikan tingkat pencemaran udara di Jakarta di masa depan. Adapun tujuan khusus yang ingin dicapai melalui penelitian ini meliputi:

- a. Menganalisis faktor-faktor utama yang menyebabkan peningkatan tingkat polusi udara di Jakarta.
- b. Mengidentifikasi dampak kesehatan masyarakat yang secara langsung terkait dengan tingkat pencemaran udara.
- c. Mengembangkan model prediktif menggunakan *machine learning* untuk memprediksi tingkat pencemaran udara di Jakarta berdasarkan tren historis.

Dengan mencapai tujuan-tujuan tersebut, diharapkan penelitian ini dapat memberikan kontribusi positif dalam upaya mitigasi pencemaran udara dan perlindungan kesehatan masyarakat di Jakarta.

## BAB 2 TINJAUAN PUSTAKA

### 2.1. *Machine Learning Modelling*

*Machine learning* adalah cabang dari kecerdasan buatan yang fokus pada pengembangan algoritma dan model komputer untuk memungkinkan sistem belajar dari data atau pengalaman sebelumnya. Proses pembuatan model, dikenal sebagai *machine learning modelling*, melibatkan langkah-langkah seperti pengumpulan dan pemrosesan data, pemilihan fitur, pemilihan algoritma, pelatihan model, evaluasi, dan penyesuaian. Tujuannya adalah menghasilkan model yang dapat menggeneralisasi dan memberikan prediksi atau keputusan yang baik pada data tak terlihat sebelumnya. Dengan menggunakan *machine learning*, sistem dapat belajar dari pola dalam data, mengidentifikasi tren, dan membuat keputusan tanpa aturan pemrograman eksplisit, memungkinkannya untuk beradaptasi dan meningkatkan kinerjanya seiring waktu dan penambahan data atau pengalaman.

### 2.2. Python

Dalam *machine learning modelling*, teknik pemrograman melibatkan penggunaan bahasa pemrograman seperti Python, serta library atau framework seperti scikit-learn, TensorFlow, atau PyTorch. Python terintegrasi dengan sistem database, mampu membaca dan mengubah berbagai jenis file, sehingga sering digunakan untuk pembuatan model *machine learning*, terutama dalam menangani data besar dan perhitungan matematika kompleks.

### 2.3 *Linear Regression*

Algoritma regresi linear adalah salah satu metode dalam statistika dan pembelajaran mesin yang digunakan untuk memodelkan hubungan linier antara satu atau lebih variabel independen (fitur) dan variabel dependen (target atau respons). Tujuan utama dari regresi linear adalah untuk menemukan garis atau bidang yang paling baik untuk mem-fitting data observasional.

Terdapat dua jenis regresi linier, yaitu:

#### a. *Simple Linear Regression*

Bentuk regresi linear yang paling sederhana dan hanya melibatkan satu variabel independen dan satu variabel dependen. Regresi linear sederhana memiliki persamaan:

$$y = \beta_0 + \beta_1 X$$

Di mana:

$y$	= variabel dependen
$x$	= variabel independen
$\beta_0$	= titik potong ( <i>intercept</i> )

$\beta_1$  = kemiringan (*slope*)

b. *Multiple Linear Regression*

Regresi ini melibatkan lebih dari satu variabel independen dan satu variabel dependen. Persamaan *multiple linear regression* adalah:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i$$

Dimana:

$y$  = variabel dependen

$x_1, x_2, \dots, x_i$  = variabel independen

$\beta_0$  = titik potong (*intercept*)

$\beta_1, \beta_2, \dots, \beta_i$  = kemiringan (*slope*)

Dalam *machine learning*, parameter-parameter ini diestimasi menggunakan metode seperti kuadrat terkecil (*least squares*) untuk menyesuaikan garis atau bidang dengan data observasional secara optimal. Regresi linear menggunakan hubungan antara titik-titik data untuk menarik garis lurus melalui semua titik. Garis ini digunakan untuk memprediksi nilai.

#### 2.4. *K-Nearest Neighbors*

*K-Nearest Neighbors* (KNN) adalah algoritma pembelajaran mesin yang efektif untuk klasifikasi dan regresi, memprediksi label atau nilai titik data baru dengan mempertimbangkan K tetangga terdekat. Pemilihan nilai K, jumlah tetangga, sangat penting dan disarankan mempertimbangkan karakteristik input data, seperti adanya outlier atau *noise*, dengan nilai K ganjil untuk menghindari ikatan dalam klasifikasi.

#### 2.5. *Decision Tree*

*Decision Tree* adalah algoritma pembelajaran mesin yang memodelkan keputusan menggunakan struktur hierarki simpul dan cabang. Dimulai dari simpul akar, pohon tumbuh dengan membagi data berdasarkan kondisi atau tes menggunakan kriteria seperti *Gini Index* atau *Mean Squared Error*. Keputusan diambil pada setiap simpul hingga mencapai kondisi berhenti.

#### 2.6. *Random Forest*

*Random Forest*, dalam pembelajaran mesin dengan pendekatan *ensemble learning*, menggabungkan prediksi dari sejumlah model *decision tree* untuk meningkatkan kinerjanya. *Decision tree*, struktur berhirarki dengan simpul dan cabang, menjadi model dasar. Melalui proses *bootstrap aggregating* (*bagging*), banyak subset acak dari data latih dibuat dengan penggantian. Setiap pohon keputusan dibangun pada subset ini dengan fitur yang dipilih secara acak untuk menghindari *overfitting*. *Voting* (klasifikasi) atau *averaging* (regresi) dilakukan pada hasil pohon untuk mencapai keputusan akhir.

## BAB 3 METODE RISET

### 3.1. Jenis Penelitian

Penelitian ini menggunakan pendekatan penelitian kuantitatif dengan metode pengembangan model prediktif menggunakan *machine learning*. Desain penelitian ini mencakup tahapan riset, pengujian dan evaluasi model, analisis error, cara penafsiran model prediksi, dan penyimpulan hasil riset.

### 3.2. Tahapan Penelitian

#### 3.2.1. Pengumpulan dan Analisis Dataset Kualitas Udara

Penelitian ini mengambil dataset dari sumber yang dapat dipercaya, yaitu Kaggle. Dataset yang digunakan dalam penelitian ini berfokus pada Indeks Standar Pencemar Udara (ISPU) dan diperoleh dari lima stasiun pemantau kualitas udara (SPKU) di Provinsi DKI Jakarta selama tahun 2021. Dataset ini membawa informasi penting terkait tingkat pencemaran udara di Jakarta. Dataset ini berisi 11 kolom. Penjelasan variabel dari dataset sebagai berikut:

- a. tanggal : Tanggal pengukuran kualitas udara
- b. stasiun : Lokasi pengukuran di stasiun
- c. pm10 : Partikulat salah satu parameter yang diukur
- d. pm25 : Partikulat salah satu parameter yang diukur
- e. so2 : Sulfida (dalam bentuk SO<sub>2</sub>) salah satu parameter yang diukur
- f. co : Karbon Monoksida salah satu parameter yang diukur
- g. o3 : Ozon salah satu parameter yang diukur
- h. no2 : Nitrogen dioksida salah satu parameter yang diukur
- i. max : Nilai ukur paling tinggi dari seluruh parameter yang diukur dalam waktu yang sama 10.
- j. critical : Parameter yang hasil pengukurannya paling tinggi
- k. kategori : Kategori hasil perhitungan indeks standar pencemaran udara

#### 3.2.2. Prosedur Riset Pengembangan Model Prediktif dengan *Machine Learning*

Setelah dataset kualitas udara terverifikasi, dilakukan pengembangan model prediktif menggunakan teknik *machine learning*. Dalam penelitian ini, empat algoritma *machine learning* akan diimplementasikan, yaitu *Linear Regression*, *K-Nearest Neighbors*, *Decision Tree*, dan *Random Forest*. Penggunaan algoritma ini bertujuan untuk memprediksi tingkat pencemaran udara berdasarkan tren historis dan faktor-faktor yang telah diidentifikasi. Proses pengembangan model prediktif melibatkan beberapa tahapan, salah satunya adalah *feature*

*encoding* dan pembagian data menjadi *training set* dan *testing set*. *Feature encoding* adalah proses mengubah variabel kategorikal dalam dataset menjadi bentuk yang dapat diinterpretasikan oleh model *machine learning*. Variabel kategorikal perlu diubah menjadi representasi numerik agar dapat dimanfaatkan oleh algoritma pembelajaran mesin.

Setelah *encoding*, data dibagi menjadi dua bagian utama: *training set* dan *testing set*. *Training set* digunakan untuk melatih model, sedangkan *testing set* digunakan untuk menguji kinerja model. Pembagian tersebut dipakai guna untuk mencegah *overfitting*, yaitu kondisi dimana model terlalu beradaptasi dengan data pelatihan dan performanya menurun pada data baru. *Training set* memberi model informasi untuk belajar pola dalam data, sementara *testing set* memberikan evaluasi independen pada kemampuan model untuk membuat prediksi akurat pada data baru yang tidak terlibat dalam pelatihan. Dengan pembagian ini, model dapat lebih baik mengatasi kompleksitas dunia nyata dan memberikan prediksi yang lebih andal.

### **3.2.3. Indikator Capaian dengan Pengujian dan Evaluasi Model**

Model yang dikembangkan akan diuji dan dievaluasi menggunakan data validasi yang tidak digunakan dalam proses pelatihan. Pengujian ini melibatkan perbandingan antara hasil prediksi model dengan data aktual. Metrik evaluasi *Mean Squared Error (MSE)* dan *Root Mean Squared Error (RMSE)* akan digunakan untuk mengukur tingkat akurasi dan keandalan model.

### **3.2.4. Analisis Error**

Setelah pengujian model menggunakan data validasi, tahapan selanjutnya adalah menganalisis error. Proses analisis error ini melibatkan identifikasi pola kesalahan, memeriksa apakah ada tren tertentu yang tidak tercakup oleh model, mengevaluasi apakah terdapat outlier atau anomali dalam *dataset* yang mungkin mempengaruhi kinerja model. Dengan memahami akar penyebab kesalahan, penelitian dapat melakukan iterasi dan penyesuaian pada model, memperbaiki parameter atau fitur yang mungkin belum optimal.

### **3.2.5. Cara Penafsiran Model Prediksi**

Tahap terakhir dalam pengembangan model prediktif adalah mendapatkan model prediksi yang optimal. Setelah melalui tahap pelatihan, pengujian, evaluasi, dan iterasi untuk mengatasi kesalahan, model yang dihasilkan dianggap siap untuk digunakan. Model prediksi ini dapat memberikan estimasi tingkat pencemaran udara di Jakarta berdasarkan faktor-faktor yang telah diidentifikasi melalui tren historis. Dengan demikian, model prediktif ini menjadi alat yang efektif dalam upaya mitigasi pencemaran udara dan perlindungan kesehatan masyarakat di Jakarta.

## BAB 4

### HASIL YANG DICAPAI DAN POTENSI KHUSUS

#### 4.1. Hasil yang Dicapai

Berikut adalah hasil prediksi model menggunakan *machine learning* dengan empat metode, yaitu *Linear Regression*, *Decision Tree*, *Random Forest*, dan *KNN*.

##### a. *Linear Regression*

No.	Variabel	MSE	RMSE
1.	pm10	209.333598315352	14.468365433432762
2.	pm25	561.2925628472929	23.69161376621046
3.	so	157.37033114209817	12.54473320330481
4.	co	16.631166103903112	4.0781326736514
5.	o3	124.3866221839405	11.152875063585197
6.	no2	117.67095271959492	10.847624289197839

Pada algoritma *linear regression*, nilai RMSE terendah diperoleh oleh variabel co dan yang tertinggi diperoleh oleh pm25.

##### b. *KNN*

No.	Variabel	MSE	RMSE
1.	pm10	227.54954545454535	15.084745455411085
2.	pm25	196.05136363636362	14.001834295418712
3.	so	103.56481818181817	10.176680115922784
4.	co	18.496909090909092	4.300803307628598
5.	o3	115.50399999999999	10.747278725333217
6.	no2	88.95745454545454	9.431725957928089

Pada algoritma *KNN*, nilai RMSE terendah diperoleh oleh variabel co dan yang tertinggi diperoleh oleh pm10.

##### c. *Decision Tree*

No.	Variabel	MSE	RMSE
1.	pm10	232.36363636363637	15.24347848634413



2.	pm25	235.46818181818182	15.344972525820365
3.	so	105.99090909090908	10.29518863794681
4.	co	32.679545454545455	5.71660261471317
5.	o3	485.27272727272725	22.028906629080055
6.	no2	119.5909090909091	10.93576284906129

Pada algoritma *decision tree*, nilai RMSE terendah diperoleh oleh variabel co dan yang tertinggi diperoleh oleh o3.

d. *Random Forest*

No.	Variabel	MSE	RMSE
1.	pm10	227.5614958565117	15.08514155904782
2.	pm25	217.5913693605715	14.75097858992994
3.	so	101.83072278744675	10.091120987652797
4.	co	22.74051056796782	4.768701140558907
5.	o3	118.46973223916233	10.884380195452671
6.	no2	70.45284026165982	8.39361902052147

Pada algoritma *random forest*, nilai RMSE terendah diperoleh oleh variabel co dan yang tertinggi diperoleh oleh pm10.

Setiap fitur memiliki metode *prediction* yang berbeda-beda untuk menghasilkan nilai error yang terendah. Berdasarkan perbandingan dari semua nilai error antar keempat metode, berikut adalah hasil yang terbaik menurut setiap fitur:

No.	Variabel	Metode	RMSE
1.	pm10	<i>Linear Regression</i>	14.468365433432762
2.	pm25	<i>KNN</i>	14.001834295418712
3.	so2	<i>Random Forest</i>	10.091120987652797
4.	co	<i>Linear Regression</i>	4.0781326736514
5.	o3	<i>KNN</i>	10.747278725333217
6.	no2	<i>Random Forest</i>	8.39361902052147

#### **4.2. Potensi Khusus**

Pengembangan model prediksi kualitas udara Jakarta dapat membantu dalam penanganan pencemaran udara yang sedang terjadi di Jakarta. Model ini dapat memprediksi kualitas udara di Jakarta setiap tahunnya.

Dengan penelitian ini, dapat dilakukan aksi pencegahan terhadap memburuknya kualitas udara Jakarta lewat publikasi artikel yang bertujuan untuk meningkatkan kesadaran masyarakat Jakarta, memberikan wawasan secara global, membantu penyusunan kebijakan dan regulasi pemerintah, serta melaksanakan praktik industri yang ramah lingkungan dengan mengembangkan model yang efektif dan akurat dalam hasil prediksinya terhadap masalah ini.

## **BAB 5**

### **PENUTUP**

Penelitian ini memberikan kontribusi signifikan dalam pemahaman dan mitigasi polusi udara di Jakarta melalui pengembangan model prediktif menggunakan pendekatan *machine learning*. Diperlukan analisis faktor-faktor utama yang berkontribusi pada peningkatan polusi udara, identifikasi dampak kesehatan masyarakat, dan pengembangan model prediktif yang menjadi langkah-langkah penting untuk meningkatkan pemahaman kita terhadap masalah ini. Melalui implementasi algoritma *machine learning*, seperti *linear regression*, *K-Nearest Neighbors*, *decision tree*, dan *random forest*, model prediktif ini mampu memberikan estimasi yang akurat terhadap tingkat pencemaran udara.

Kesimpulan dari penelitian ini adalah bahwa pengembangan model prediktif kualitas udara Jakarta dapat memberikan kontribusi positif dalam mengatasi masalah pencemaran udara. Model ini dapat menjadi alat yang efektif dalam memberikan proyeksi tingkat pencemaran udara di masa depan, memberikan pemahaman yang lebih baik terhadap faktor-faktor penyebab, dan memberikan dasar untuk merancang kebijakan pengendalian polusi udara yang lebih efektif.

Selain itu, penelitian ini memberikan potensi khusus dalam meningkatkan kesadaran masyarakat terhadap masalah polusi udara di Jakarta. Dengan publikasi ilmiah hasil penelitian dan pemahaman lebih lanjut tentang faktor-faktor penyebab polusi udara, diharapkan dapat mendorong aksi pencegahan yang lebih proaktif dan partisipasi aktif dalam upaya mitigasi polusi udara.

Saran untuk penelitian selanjutnya adalah melakukan pengembangan lebih lanjut pada model prediktif ini dengan mengintegrasikan data yang lebih detail, termasuk variabel-variabel cuaca, topografi, dan aktivitas industri yang lebih spesifik. Selain itu, perlu dilakukan penelitian lanjut terkait implementasi kebijakan berbasis hasil prediksi model untuk memberikan dampak yang lebih nyata dalam upaya pengendalian polusi udara di Jakarta.

## DAFTAR PUSTAKA

- Alpaydin, E. (2021). *Machine Learning*. Massachusetts Institute of Technology. Massachusetts. USA. Dakuten dari <https://books.google.co.id/books?hl=en&lr=&id=2nQJEAAAQBAJ&oi=fnd&pg=PR7&dq=machine+learning#v=onepage&q=machine%20learning&f=false> pada tanggal 24 Desember 2023.
- Aurélien Géron (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA: O'Reilly Media, Inc.
- Daly, A. and P. Zannetti. 2007. *Air Pollution Modeling – An Overview*. Chapter 2 of *AMBIENT AIR POLLUTION* (P. Zannetti, D. Al-Ajmi, and S. Al-Rashied, Editors). Published by The Arab School for Science and Technology (ASST) (<http://www.arabschool.org.sy>) and The EnviroComp Institute (<http://www.envirocomp.org/>).
- Vidnerová, P., & Neruda, R. (2021). *Air Pollution Modelling by Machine Learning Methods*. The Czech Academy of Sciences. Prague. Czech Republic. Dikutip dari <https://doi.org/10.3390/modelling2040035> pada tanggal 20 Desember 2023.
- Zannetti, P. (1990) Polimi.it, Air Pollution Modeling. Dikutip dari <https://guariso.faculty.polimi.it/AESA/Air%20Pollution%20Modeling.pdf> pada tanggal 26 Desember 2023.

## LAMPIRAN

Lampiran 1. Formulir Penilaian Laporan Akhir

Judul Kegiatan	:	Analisis Polusi Udara dan Prediksi Pemodelan
Bidang PKM	:	PKM-RE
Bidang Ilmu	:	<i>Artificial Intelligence</i>
NIM / Nama Ketua	:	2602190816 / Kimberly Kayla Dewi
NIM / Nama Anggota 1	:	2602199582 / Fiona Maharani Nugraha
NIM / Nama Anggota 2	:	2602135963 / Wella Priscillia
NIM / Nama Anggota 3	:	2602186472 / Jennifer Patricia
NIM / Nama Anggota 4	:	2602164433 / David Paul Ong
Perguruan Tinggi	:	BINUS University
Program Studi	:	Data Science

No	Kriteria	Bobot	Skor	Nilai
1	Pendahuluan	15		
2	Tinjauan Pustaka	20		
3	Metode Riset	25		
4	Hasil yang Dicapai dan Potensi Khusus	25		
5	Penutup (Kesimpulan dan Saran)	10		
6	Daftar Pustaka	5		
	Total	100		

Keterangan:

Nilai = Bobot x Skor; Skor (1=Buruk; 2=Sangat kurang; 3=Kurang; 5=Cukup; 6=Baik; 7=Sangat baik);

Komentar:

Jakarta,.....  
Penilai,

( )