

## ANALISIS POLUSI UDARA DAN PREDIKSI PEMODELAN

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

Data Science, BINUS University, Jakarta Barat, Jakarta, Indonesia

Abram Setyo Prabowo, S.Kom, M.Cs.: kimberly.dewi@binus.ac.id

### ABSTRAK

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, *Machine Learning*, *Artificial Intelligence*, Jakarta, dan Python.

### ABSTRACT

*This study highlights the increased air pollution in Jakarta against the backdrop of high levels of air contamination and its serious impact on public health. The main objective of this research is to develop a predictive model using machine learning methods to project future air pollution levels. In the literature review, the concepts of machine learning, the use of Python programming language, and several machine learning modeling algorithms such as linear regression, K-Nearest Neighbors, decision tree, and random forest are discussed. The research methodology includes the collection and analysis of air quality datasets, model development, testing, evaluation, and error analysis. The results indicate that this predictive model can provide accuracy estimates using metrics like Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) for air pollution levels. The particular potential of this research lies in its contribution to air pollution mitigation and raising public awareness. The conclusion summarizes that this model is effective in providing projections of future air quality, supporting air pollution control policies, and emphasizing the need for more proactive preventive measures.*

*Keywords: Pollution, Machine Learning, Artificial Intelligence, Jakarta, and Python.*

## Pendahuluan

Polusi udara merupakan pencemaran udara yang disebabkan oleh makhluk hidup yang mengakibatkan zat maupun komponen lainnya masuk ke dalam udara. 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 juga 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 serta kerusakan ekosistem.

Manfaat khusus yang ingin dicapai melalui penelitian ini mencakup pemahaman yang mendalam tentang faktor-faktor penyebab polusi udara di Jakarta, peningkatan kesadaran masyarakat mengenai dampak kesehatan yang disebabkan oleh polusi udara, dan penyediaan dasar informasi bagi pemerintah dan lembaga terkait untuk merancang kebijakan pengendalian polusi udara yang efektif.

Urgensi penelitian ini muncul dari tingkat polusi udara yang mengkhawatirkan di Jakarta, yang memiliki dampak serius terhadap kesehatan masyarakat dan lingkungan. Oleh karena itu, penelitian ini mendesak dalam memberikan pemahaman lebih dalam tentang penyebab dan dampak polusi udara, sambil mengembangkan solusi berbasis data untuk mengatasi masalah krusial ini.

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.

## Metode

Penelitian ini dilaksanakan selama lima bulan, mulai dari September hingga Januari 2023, dengan lokasi secara nomaden antara BINUS University dan tempat tinggal masing-masing anggota tim. Kami menggunakan laptop pribadi sebagai alat untuk mendukung pelaksanaan PKM-RE.

Dataset yang digunakan diperoleh dari *Kaggle*, terfokus pada Indeks Standar Pencemar Udara (ISPU) dari lima stasiun pemantau kualitas udara (SPKU) di Provinsi DKI Jakarta selama tahun 2021. Dataset ini berisi 11 kolom yang menyimpan informasi penting mengenai pencemaran udara di Jakarta.

Metode penelitian menggunakan pendekatan kuantitatif dengan pengembangan model prediktif menggunakan *machine learning*. Tahapannya mencakup riset, pengujian model, evaluasi, analisis error, dan cara penafsiran model prediksi. Proses pengembangan model melibatkan implementasi empat algoritma *machine learning*: *linear regression*, *K-Nearest Neighbors*, *decision tree*, dan *random forest*. Data dipecah menjadi *training set* dan *testing set* untuk menghindari *overfitting*, sehingga model dapat memberikan prediksi yang lebih handal.

Evaluasi model menggunakan data validasi untuk membandingkan hasil prediksi dengan data aktual. Matrik evaluasi, seperti *Mean Squared Error (MSE)* dan *Root Mean Squared Error (RMSE)*, digunakan untuk mengukur akurasi model. Analisis error dilakukan untuk mengidentifikasi pola kesalahan, mengoreksi model, dan memperbaiki parameter yang belum optimal.

Hasil akhirnya adalah model prediksi yang dapat memberikan estimasi tingkat pencemaran udara di Jakarta berdasarkan tren historis. Model ini berperan dalam mitigasi pencemaran udara dan perlindungan kesehatan masyarakat Jakarta.

## Hasil dan Pembahasan

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

### a. *Linear Regression*

Tabel 1. Hasil Prediksi Model dengan Metode *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*

Tabel 2. Hasil Prediksi Model dengan Metode *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*

Tabel 3. Hasil Prediksi Model dengan Metode *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*Tabel 4. Hasil Prediksi Model dengan Metode *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:

Tabel 5. Hasil Prediksi Model Metode Terbaik untuk 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

Hasil dari penelitian kami adalah sebuah model prediksi yang mampu mengestimasi tingkat pencemaran udara di Jakarta dengan memperhatikan pola dari masa lalu. Model yang telah kami buat memiliki peran penting dalam mengurangi pencemaran udara dan menjaga kesehatan masyarakat di Jakarta

## Kesimpulan

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.

## Ucapan Terima Kasih

Dalam kesempatan ini, kami ingin mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Abram Setyo Prabowo, S.Kom, M.Cs. karena telah memberikan kami kesempatan untuk menyalurkan kreativitas kami dan membimbing selama penelitian ini berlangsung. Kami juga ingin berterima kasih kepada BINUS University yang telah menyalurkan program kreativitas mahasiswa dari pemerintah kepada kami. Kami juga ingin berterima kasih kepada Kaggle sebagai sumber *dataset* yang telah digunakan oleh kami untuk melakukan penelitian. Kami juga mengucapkan terima kasih kepada teman-teman seangkatan yang berasal dari program studi yang sama atas bantuannya dan juga semua pihak yang sudah disebut maupun yang belum disebutkan yang telah membantu kami dalam pelaksanaan penelitian dan penyusunan laporan ini.

## Kontribusi Penulis

Kimberly Kayla Dewi (Penulis Satu):

Menjabat sebagai ketua kelompok, Kimberly Kayla Dewi memegang peran sentral dalam kesuksesan proyek penelitian ini. Ia bertanggung jawab penuh atas pelaksanaan PKM-RE, penyusunan proposal, dan artikel ilmiah. Mulai dari perencanaan hingga implementasi. Kontribusinya mencakup pengamatan mendalam terhadap setiap anggota kelompok, tahap penelitian dan penyusunan laporan, dan memastikan bahwa setiap langkah proyek terarah sesuai dengan tujuan yang telah ditetapkan.

David Paul Ong (Penulis Dua):

Bertanggung jawab dalam pembuatan dan penjalanan pemodelan prediksi dengan penuh dedikasi dan mengumpulkan data yang digunakan sesuai dengan metodologi yang telah ditetapkan. Data yang diperolehnya menjadi dasar utama untuk

menyokong temuan penelitian. Hasilnya dari model dengan metode *decision tree* dan *random forest* digunakan untuk pembuatan penelitian.

Fiona Maharani Nugraha (Penulis Tiga):

Bertanggung jawab sepenuhnya atas pengetikan laporan akhir proyek penelitian. Fiona Maharani Nugraha menunjukkan dedikasi tinggi dalam merinci setiap aspek penelitian dan memastikan bahwa informasi yang terkandung dalam laporan disusun dengan jelas dan tepat. Perannya bukan hanya sebatas pengetikan, tetapi juga mencakup pengamatan mendalam terhadap hasil penelitian, memastikan bahwa setiap temuan terdokumentasi dengan akurat dan terstruktur.

Jennifer Patricia (Penulis Empat):

Bertanggung jawab dalam pembuatan dan penjalanan pemodelan prediksi dengan penuh dedikasi dan mengumpulkan data yang digunakan sesuai dengan metodologi yang telah ditetapkan. Data yang diperolehnya menjadi dasar utama untuk menyokong temuan penelitian. Hasilnya dari model dengan metode *linear regression* dan *KNN* digunakan untuk pembuatan penelitian.

Wella Priscillia (Penulis Lima):

Bertanggung jawab penuh atas pengetikan laporan kemajuan proyek penelitian ini. Kontribusinya melibatkan ketelitian dalam mengetik dan menyusun informasi yang terkandung dalam laporan, memastikan bahwa semua aspek penelitian termasuk pengamatan dan temuan terdokumentasi secara akurat.

Abram Setyo Prabowo, S.Kom, M.Cs. (Dosen Pendamping):

Sebagai dosen pendamping, Abram Setyo Prabowo memiliki peran strategis dalam arahan riset keseluruhan. Beliau terlibat dalam memberikan panduan metodologis dan memberikan arahan intelektual. Selain itu, Abram Setyo Prabowo turut bertanggung jawab atas keseluruhan proyek penelitian, memastikan bahwa metodologi dan analisis data sesuai dengan standar ilmiah. Abram Setyo Prabowo juga berkontribusi signifikan dalam penyelesaian akhir naskah/manuskrip, memastikan konsistensi dan kesesuaian dengan tujuan penelitian.

## Daftar Pustaka

- Alpaydin, E. (2021). *Machine Learning*. Massachusetts Institute of Technology. Massachusetts. USA. Dikutip 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.