

DAFTAR ISI

DAFTAR ISI	i
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	1
BAB 2. TARGET LUARAN	2
BAB 3. METODE RISET	3
3.1 Jenis Penelitian	3
3.2 Tahapan Penelitian	3
BAB 4. HASIL YANG DICAPAI	5
BAB 5. POTENSI HASIL	7
5.1 Manfaat Penelitian	7
5.2 Target Publikasi Ilmiah	7
5.3 Peluang Perolehan Kekayaan Intelektual	8
BAB 6. RENCANA TAHAPAN BERIKUTNYA	9
DAFTAR PUSTAKA	10
LAMPIRAN	11
Lampiran 1. Penggunaan Dana	11
Lampiran 2. Bukti-Bukti Pendukung Kegiatan	11
Lampiran 3. Formulir Penilaian Laporan Kemajuan Pelaksanaan	11
Lampiran 4. Formulir Penilaian PKP2 (Presentasi)	12

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

TARGET LUARAN

Adapun target luaran yang ingin kami capai adalah sebagai berikut:

- a. Menghasilkan model prediktif yang dapat mengestimasi tingkat pencemaran udara di Jakarta dengan akurasi yang tinggi pada tahun yang mendatang berdasarkan tren historis.
- b. Dapat melakukan aksi pencegahan terhadap memburuknya kualitas udara dan menyediakan informasi yang berguna untuk meningkatkan kesadaran masyarakat Jakarta berdasarkan model prediksi yang telah dibuat.
- c. Dapat menyajikan temuan dan metodologi penelitian melalui laporan atau publikasi ilmiah yang memungkinkan pengetahuan ini tersebar luas dan dapat digunakan oleh pihak yang berkepentingan.

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 SO₂) 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*

Dataset kualitas udara diverifikasi sebelum memulai pengembangan model prediktif menggunakan *machine learning*. Dalam riset ini, empat algoritma *machine learning*, seperti *Linear Regression*, *K-Nearest Neighbors*, *Decision Tree*, dan *Random Forest*, diterapkan untuk memprediksi tingkat pencemaran udara berdasarkan tren historis dan faktor-faktor yang teridentifikasi. Tahapan pengembangan model mencakup *feature encoding* dan pemisahan data menjadi *training set* dan

testing set. *Feature encoding* mengubah variabel kategorikal ke bentuk numerik agar algoritma *machine learning* dapat memahaminya. Data dibagi menjadi *training set* untuk melatih model dan *testing set* untuk mengevaluasi kinerjanya, mencegah *overfitting* yang bisa mempengaruhi performa model pada data baru. Dengan pendekatan ini, model dapat menghadapi kompleksitas dunia nyata dan memberikan prediksi yang lebih handal.

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

Berikut adalah hasil prediksi model menggunakan *machine learning* dengan empat metode:

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. 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

Persentase hasil yang dicapai oleh kelompok kami terhadap keseluruhan data yang menjadi target kegiatan sudah mencapai 100%.

BAB 5

POTENSI HASIL

5.1. Manfaat Penelitian

Adapun manfaat yang akan dicapai melalui penelitian ini meliputi:

- a. Dapat memberikan pemahaman mendalam tentang faktor-faktor penyebab polusi udara di Jakarta.
- b. Dapat meningkatkan kesadaran masyarakat tentang dampak kesehatan yang diakibatkan oleh polusi udara.
- c. Dapat menyediakan dasar informasi bagi pemerintah dan lembaga terkait untuk merancang kebijakan pengendalian polusi udara yang efektif.

5.2. Target Publikasi Ilmiah

Berikut adalah target publikasi ilmiah yang relevan dengan penelitian kami:

a. Masyarakat Umum

Publikasi dapat diakses dan dimengerti oleh masyarakat umum dengan menjelaskan pentingnya pengendalian pencemaran udara bagi kesehatan mereka sehingga dapat dilakukan pencegahan agar pencemaran udara tidak semakin memburuk.

b. Pemerintah

Publikasi memberikan wawasan tentang hasil penelitian serta implikasinya terhadap kebijakan dan regulasi pemerintah dalam mengelola kualitas udara. Bersinggungan dengan kebijakan, publikasi ini dapat membantu pemerintah untuk menganalisis dan menyusun kebijakan dalam pengambilan keputusan yang lebih matang.

c. Industri

Publikasi menjelaskan implikasi penelitian terhadap industri, khususnya terkait dengan industri yang berkontribusi terhadap pencemaran udara. Selanjutnya, publikasi juga dapat memberikan saran dan inspirasi untuk praktik yang lebih ramah lingkungan.

d. Lembaga Internasional

Publikasi yang mengeksplorasi implikasi global dari penelitian tersebut, memberikan wawasan kepada lembaga internasional tentang pentingnya mitigasi pencemaran udara dan penggunaan model prediktif untuk perlindungan lingkungan dan kesehatan masyarakat secara global.

5.3. Peluang Perolehan Hak Kekayaan Intelektual

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

Dengan ini, kita dapat melakukan aksi pencegahan terhadap memburuknya kualitas udara Jakarta. Publikasi artikel bertujuan untuk meningkatkan kesadaran terhadap masalah ini dengan harapan dapat mengembangkan model yang telah dibuat untuk menjadi model yang lebih efektif dan akurat dalam hasil prediksinya.

BAB 6

RENCANA TAHAPAN BERIKUTNYA

Tahapan berikutnya setelah mendapatkan hasil dari keempat metode prediksi (*Linear Regression*, *Decision Tree*, *Random Forest*, dan *KNN*) adalah melakukan perbandingan antara masing-masing metode prediksi tersebut. Perbandingan ini dilakukan dengan tujuan untuk mengevaluasi dan menentukan metode prediksi yang paling optimal dan tepat digunakan untuk setiap kolom atau variabel dalam dataset.

Proses perbandingan akan melibatkan analisis mendalam terhadap performa masing-masing metode prediksi. Kami akan mengevaluasi kinerja setiap metode dalam memprediksi nilai-nilai pada kolom-kolom tertentu, seperti *pm10*, *pm25*, *so2*, *co*, *o3*, *no2*, dan variabel lainnya yang terdapat dalam dataset. Evaluasi ini akan mempertimbangkan metrik-metrik evaluasi seperti *Mean Squared Error (MSE)* dan *Root Mean Squared Error (RMSE)*.

Selain itu, kami juga akan melakukan analisis terhadap kelebihan dan kekurangan dari setiap metode prediksi. Kami akan meninjau aspek-aspek seperti kecepatan komputasi, kemampuan dalam menangani data yang kompleks, serta kemudahan dalam interpretasi hasil prediksi. Analisis ini akan membantu kami untuk memahami konteks di mana setiap metode prediksi paling baik digunakan dan keunggulan masing-masing metode dalam memprediksi nilai-nilai kualitas udara.

Hasil dari tahap perbandingan ini akan menjadi dasar dalam menentukan metode prediksi terbaik yang akan kami gunakan secara spesifik untuk setiap variabel dalam pengembangan model prediksi. Langkah ini sangat penting dalam memastikan bahwa prediksi yang dihasilkan oleh model memiliki akurasi dan keandalan yang tinggi, serta dapat memberikan informasi yang berguna dalam upaya mitigasi pencemaran udara dan perlindungan kesehatan masyarakat.

DAFTAR PUSTAKA

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

Lampiran 2. Bukti-Bukti Pendukung Kegiatan

Lampiran 3. Formulir Penilaian Laporan Kemajuan Pelaksanaan

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	10		
2	Target Luaran	15		
3	Metode Riset	25		
4	Hasil yang Dicapai	25		
5	Potensi Hasil	15		
6	Publikasi dan/atau promosi melalui media sosial	5		
7	Rencana Tahapan Berikutnya	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,

()

Lampiran 4. Formulir Penilaian PKP2 (Presentasi)

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	Target Luaran (kesesuaian luaran dan permasalahan)	10		
2	Metode Riset (kemutakhiran dan keberhasilan metode)	15		

	riset)			
3	Tingkat Kreativitas dan Ketercapaian Target Luaran (permasalahan, ketepatan solusi, kesesuaian jenis dan jumlah luaran, kesesuaian dengan catatan harian)	35		
4	Kesesuaian Pelaksanaan dan Rencana Tahapan Berikutnya (waktu pelaksanaan, bahan dan alat serta metode yang digunakan, personalia, biaya)	10		
5	Kekompakan Tim Pelaksana dan Peran Dosen Pendamping (kerjasama, pembagian tugas, mengoreksi proposal, memantau pelaksanaan, melayani konsultasi)	10		
6	Potensi Khusus (utamanya potensi menghasilkan artikel ilmiah yang baik, kemanfaatan)	20		
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,

()