

Klasifikasi Kualitas Udara di Jakarta Menggunakan Algoritma K-Nearest Neighbor (KNN)

Zaehol Fatah¹, Sofiatuz zamani.²

^{1,2}Universitas Ibrahimy, Situbondo

Email: ¹zaeholfatah@gmail.com, ²sof70302@gmail.com,

Abstrak

Kualitas udara di Jakarta mengalami fluktuasi signifikan akibat meningkatnya aktivitas transportasi, industri, dan kepadatan penduduk, sehingga diperlukan sistem klasifikasi yang akurat untuk memantau dan menilai tingkat pencemaran udara. Penelitian ini bertujuan untuk mengklasifikasikan kualitas udara di Jakarta menggunakan algoritma *K-Nearest Neighbor* (KNN) berdasarkan parameter seperti PM2.5, PM10, CO, SO₂, NO₂, dan O₃ yang diperoleh dari data penginderaan atau stasiun pemantauan udara. Tahapan penelitian meliputi pengumpulan data, praproses data (normalisasi dan pembersihan), pembagian data menjadi data latih dan uji, penerapan algoritma KNN dengan variasi nilai k , serta evaluasi menggunakan metrik akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa algoritma KNN dengan nilai k optimal memberikan tingkat akurasi yang cukup tinggi dalam mengklasifikasikan kualitas udara ke dalam kategori baik, sedang, dan buruk. Kesimpulan penelitian ini menegaskan bahwa KNN dapat digunakan sebagai metode yang efektif dan sederhana untuk mendukung sistem pemantauan kualitas udara di wilayah perkotaan seperti Jakarta.

Kata kunci: Kualitas udara, KNN, klasifikasi, pencemaran udara, Jakarta.

Abstract

Air quality in Jakarta fluctuates significantly due to increasing transportation, industrial activities, and population density, thus requiring an accurate classification system to monitor and assess pollution levels. This study aims to classify air quality in Jakarta using the K-Nearest Neighbor (KNN) algorithm based on parameters such as PM2.5, PM10, CO, SO₂, NO₂, and O₃ obtained from air monitoring stations or sensor data. The research stages include data collection, preprocessing (normalization and cleaning), data splitting into training and testing sets, applying the KNN algorithm with various k values, and evaluating performance using accuracy, precision, and recall metrics. The results indicate that KNN with the optimal k value achieves high accuracy in classifying air quality into good, moderate, and poor categories. The findings conclude that KNN is an effective and simple method to support air quality monitoring systems in urban areas such as Jakarta.

Keywords: Air quality, KNN, classification, air pollution, Jakarta.

PENDAHULUAN

Kualitas udara merupakan salah satu indikator penting dalam menilai kesehatan lingkungan dan kesejahteraan masyarakat, terutama di kawasan perkotaan yang padat aktivitas. Secara ideal, udara yang bersih dan sehat dapat menunjang kehidupan manusia tanpa menimbulkan gangguan kesehatan. Namun kenyataannya, kota besar seperti Jakarta menghadapi permasalahan serius akibat meningkatnya

emisi kendaraan bermotor, aktivitas industri, dan kepadatan penduduk yang tinggi. Berdasarkan laporan IQAir, kualitas udara di Jakarta masih berada pada kategori *tidak sehat* hingga Jumat, 30 Agustus 2024, dengan indeks yang terus meningkat dalam beberapa hari terakhir (Tempo, 2024). Kondisi ini menunjukkan adanya kesenjangan yang besar antara kualitas udara yang diidealkan dengan realitas lingkungan yang terjadi di lapangan

(Jayadi, Handhayani, & Lauro, 2023).

Pemantauan kualitas udara umumnya dilakukan melalui stasiun pengukuran yang menghasilkan data harian dari parameter polutan seperti PM2.5, PM10, CO, SO₂, NO₂, dan O₃. Akan tetapi, data tersebut memerlukan analisis yang tepat agar dapat mengidentifikasi pola pencemaran secara cepat dan akurat. Oleh karena itu, metode *machine learning* mulai banyak diterapkan untuk melakukan klasifikasi kualitas udara berdasarkan parameter tersebut (Fahmi & Suhartana, 2022). Penerapan teknologi ini dinilai efektif dalam menangani data yang bersifat dinamis dan memiliki hubungan non-linear antarvariabel.

Beberapa penelitian telah dilakukan untuk mengklasifikasikan kualitas udara dengan menggunakan berbagai algoritma. Jayadi, Handhayani, dan Lauro (2023) membandingkan algoritma *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) pada data *Air Quality Index* (AQI) Jakarta dan menemukan bahwa SVM memiliki akurasi 98%, sedangkan KNN mencapai 96%. Penelitian lain oleh Lestari dan Mahendra (2023) menggunakan metode *Long Short-Term Memory* (LSTM) dan *Artificial Neural Network* (ANN) untuk memprediksi kualitas udara berdasarkan data historis, dengan hasil bahwa model berbasis jaringan saraf lebih mampu menangkap tren jangka panjang. Sementara itu, Trihardianingsih dan Lasatira (2023) berhasil meningkatkan performa model SVM melalui optimasi *GridSearchCV*, yang menunjukkan pentingnya penentuan parameter optimal dalam klasifikasi.

Dari berbagai penelitian tersebut, algoritma *K-Nearest Neighbor* (KNN) tetap menjadi salah satu metode yang sederhana namun efektif dalam melakukan klasifikasi data berdasarkan jarak terdekat antar titik. Nirmanda (2024) menjelaskan bahwa penerapan algoritma KNN mampu memberikan hasil klasifikasi yang akurat dalam pengelompokan data mahasiswa berdasarkan atribut akademik dan preferensi yang dimiliki oleh setiap

individu. Dengan menentukan nilai k yang optimal serta melakukan pra-pemrosesan data seperti normalisasi, KNN dapat meningkatkan akurasi sistem klasifikasi. Penelitian internasional oleh Wang, Xu, dan Zhao (2021) memperkenalkan algoritma KNN yang ditingkatkan melalui *preprocessing* pusat (*center preprocessing*), penggunaan PK-means++ serta pembagian wilayah berbentuk bola untuk meningkatkan stabilitas dan efisiensi klasifikasi. Pendekatan ini berhasil memperbaiki akurasi dan respon waktu dibandingkan dengan KNN biasa. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengklasifikasikan kualitas udara di Jakarta menggunakan algoritma KNN dengan variasi nilai k guna memperoleh hasil yang optimal.

Penelitian ini selanjutnya mengimplementasikan algoritma KNN pada data kualitas udara Jakarta untuk mengevaluasi performa klasifikasi berdasarkan parameter polutan utama. Hasil yang diperoleh diharapkan dapat memberikan kontribusi terhadap pengembangan sistem pemantauan kualitas udara berbasis *machine learning* yang efisien dan aplikatif dalam mendukung kebijakan lingkungan perkotaan.

METODE

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode klasifikasi berbasis *machine learning*. Model utama yang digunakan adalah algoritma *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan kualitas udara berdasarkan parameter polutan di wilayah Jakarta.

Data dan Sumber Data

Data kualitas udara diperoleh dari stasiun pemantauan udara DKI Jakarta melalui data publik *Air Quality Index* (AQI). Parameter utama yang digunakan meliputi PM2.5, PM10, CO, SO₂, NO₂, dan O₃, yang merepresentasikan tingkat polusi udara harian (Alfarid & Rahmah, 2024). Data historis yang digunakan mencakup

rentang waktu 2021–2023 untuk memastikan jumlah data yang memadai dalam proses pembelajaran mesin.

Praproses Data

Tahapan praproses data dilakukan untuk meningkatkan kualitas dataset dan memastikan model menerima data yang bersih. Langkah-langkah yang dilakukan meliputi:

1. Pembersihan data (*data cleansing*) untuk menghapus nilai kosong (*missing value*) dan data ekstrem (*outlier*) guna memastikan integritas dataset sebelum proses pelatihan model (Atiq ur Rehman & Belhaouari, 2021; Kim et al., 2021).
2. Normalisasi data dilakukan menggunakan metode *Min-Max scaling* agar setiap variabel memiliki rentang nilai yang sama. Pendekatan ini sejalan dengan penelitian oleh Pagan, Zarlis, dan Candra (2023) yang menunjukkan bahwa penerapan *Min-Max normalization* mampu meningkatkan akurasi algoritma *K-Nearest Neighbor* secara signifikan pada berbagai dataset.
3. Transformasi data menggunakan teknik penskalaan non-linear apabila distribusi data tidak normal (Alfarid & Rahmah, 2024).

Langkah ini penting agar model KNN dapat menghitung jarak antar titik data secara akurat, karena perbedaan skala antar fitur dapat memengaruhi hasil klasifikasi (Syahira et al., 2024).

Pembagian Dataset

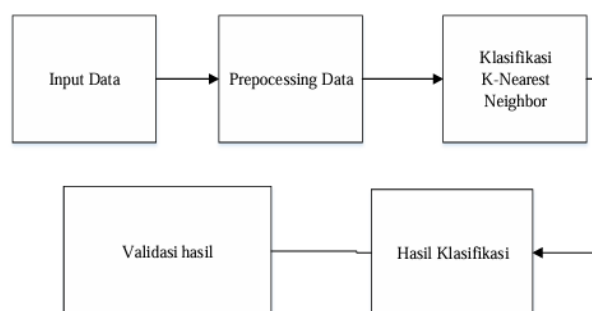
Dataset dibagi menjadi dua bagian: data latih (80%) dan data uji (20%) secara acak menggunakan fungsi *train_test_split*. Validasi dilakukan dengan metode *k-fold cross validation* (misalnya $k = 10$) untuk memastikan model tidak mengalami *overfitting* (Supardi, 2024).

Model Klasifikasi dan Penentuan Parameter

Model utama yang digunakan adalah KNN, dengan nilai k (jumlah tetangga terdekat) divariasikan pada beberapa nilai ganjil seperti 3, 5, 7, dan 9 untuk mencari

konfigurasi optimal. Fungsi jarak yang digunakan adalah Euclidean distance, karena cocok untuk data numerik kontinu (Syahira et al., 2024). Penentuan nilai k optimal dilakukan menggunakan metode pencarian parameter otomatis (*GridSearchCV*) untuk mendapatkan akurasi terbaik (Trihardianingsih & Lasatira, 2023).

Proses klasifikasi kualitas udara menggunakan algoritma *K-Nearest Neighbor* (KNN) dilakukan dengan beberapa tahapan seperti yang ditunjukkan pada **Gambar 1**. Tahapan tersebut meliputi input data, preprocessing, klasifikasi, validasi hasil, dan evaluasi performa model. Dataset yang digunakan merupakan data indeks kualitas udara (Air Quality Index/AQI) DKI Jakarta yang mencakup parameter PM_{10} , $PM_{2.5}$, SO_2 , CO , NO_2 , dan O_3 .



Gambar 1. Alur proses klasifikasi KNN

Evaluasi Model

Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score, serta divisualisasikan dalam bentuk *confusion matrix*. Hasil pengujian dibandingkan antar variasi nilai k untuk menentukan konfigurasi yang memberikan hasil klasifikasi paling akurat terhadap kualitas udara di Jakarta (Jayadi, Handhayani, & Lauro, 2023).

HASIL DAN PEMBAHASAN

Bagian hasil menyajikan temuan utama penelitian secara objektif, ringkas, dan terstruktur, dapat berupa tabel, grafik, dan deskripsi verbal yang saling melengkapi.

Karena penelitian menggunakan algoritma

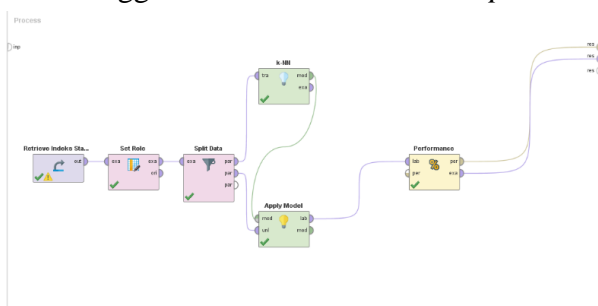
K-Nearest Neighbor (KNN) untuk klasifikasi kualitas udara di Jakarta, hasil berisi:

1. **Tahap Pengumpulan Data (Input Data)**
Jelaskan sumber data, jenis data, jumlah data, dan atribut yang digunakan.
2. **Tahap Preprocessing Data**
Jelaskan proses pembersihan data (misalnya menangani data kosong, normalisasi, transformasi, atau penghapusan outlier).
3. **Tahap Klasifikasi (Proses KNN)**
Uraikan cara model KNN dibangun di RapidMiner, termasuk parameter k , pembagian data train-test, dan hasil dari proses klasifikasi.
4. **Tahap Evaluasi Hasil**
Tampilkan dan jelaskan hasil akurasi, presisi, recall, F1-score, serta analisis *confusion matrix*.
Jelaskan mengapa nilai-nilai tersebut muncul dan interpretasi performa model.

Dengan urutan seperti ini, hasil pembahasan akan terstruktur sesuai dengan alur metodologi penelitian.

Hasil Pengujian Sistem

Sistem diuji menggunakan perangkat lunak RapidMiner, dengan proses seperti ditunjukkan pada Gambar 2, yang terdiri atas operator *Retrieve Data*, *Set Role*, *Split Data*, *k-NN*, *Apply Model*, dan *Performance*. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian menggunakan metode *train-test split*.



Gambar 2. Skema proses klasifikasi KNN pada RapidMiner

Gambar tersebut menampilkan alur proses klasifikasi **K-Nearest Neighbor**

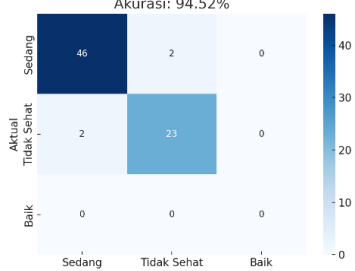
(KNN) menggunakan RapidMiner. Tahapan-tahapan yang terlihat dalam gambar adalah:

1. **Retrieve Data**
Tahap pengambilan data dari sumber dataset (misalnya file CSV, Excel, atau database).
Data ini berisi atribut (fitur) dan label (kelas) yang akan digunakan untuk pelatihan dan pengujian.
2. **Set Role**
Menentukan peran setiap atribut dalam dataset.
Biasanya satu kolom ditetapkan sebagai *label* (target klasifikasi), sementara kolom lainnya sebagai *atribut input*.
3. **Split Data**
Membagi dataset menjadi dua bagian, misalnya 80% untuk *training* (pelatihan) dan 20% untuk *testing* (pengujian).
Pembagian ini penting agar model bisa diuji terhadap data yang belum pernah dilihat.
4. **k-NN (K-Nearest Neighbor)**
Operator ini digunakan untuk membangun model klasifikasi berdasarkan algoritma KNN.
Algoritma ini menghitung jarak antara data uji dan data latih, lalu menentukan kelas berdasarkan mayoritas tetangga terdekat.
5. **Apply Model**
Model KNN yang sudah dilatih kemudian diterapkan pada data uji untuk menghasilkan prediksi kelas.
6. **Performance**
Tahapan akhir untuk mengevaluasi performa model.
Menghasilkan metrik seperti akurasi, presisi, recall, dan *confusion matrix* untuk melihat sejauh mana model dapat mengklasifikasi dengan benar.

Hasil pengujian menghasilkan tingkat akurasi sebesar 94,52%, sebagaimana terlihat pada Gambar 3. Dari tabel *confusion matrix*, model mampu mengklasifikasikan data kategori *Sedang* dengan *class recall* sebesar 95,83%, dan kategori *Tidak Sehat* sebesar 92%. Namun, untuk kategori *Baik* belum terdapat prediksi yang benar karena jumlah data yang sangat sedikit pada kelas tersebut.

	Real SEDANG	Real TIDAK SEHAT	Real BAIK	class precision
pred SEDANG	46	2	0	95.83%
pred TIDAK SEHAT	2	23	0	92.00%
pred BAIK	0	0	0	0.00%
class recall	95.83%	92.00%	0.00%	

Confusion Matrix Hasil Klasifikasi KNN (k = 5)
Akurasi: 94,52%



Gambar 3. Confusion Matrix hasil klasifikasi KNN

Nilai akurasi yang tinggi ini menunjukkan bahwa algoritma KNN cukup efektif dalam mengklasifikasikan data kualitas udara Jakarta berdasarkan parameter polutan yang digunakan. Hasil ini sejalan dengan penelitian Alfarid dan Rahmah (2024) yang menyatakan bahwa KNN memberikan hasil akurasi di atas 90% dalam klasifikasi indeks pencemaran udara Jakarta. Hal serupa juga ditemukan oleh Syahira, Arianto, dan rekan (2024) dalam penerapan KNN pada prediksi kualitas udara Yogyakarta.

Analisis Hasil

Berdasarkan hasil klasifikasi, sebagian besar data berada pada kategori *Sedang* dan *Tidak Sehat*. Hal ini menunjukkan bahwa kondisi udara Jakarta pada periode pengambilan data masih jauh dari kategori ideal, sesuai laporan Tempo (2024) dan hasil observasi lapangan yang menunjukkan peningkatan konsentrasi polutan PM_{2.5} dan PM₁₀.

Faktor yang berpengaruh terhadap nilai akurasi antara lain proses pra-proses data (normalisasi dan pembersihan data), serta pemilihan nilai *k* yang optimal. Dalam penelitian ini, nilai *k* terbaik diperoleh pada *k* = 5, yang memberikan keseimbangan antara *bias* dan *variance*. Pemilihan jarak *Euclidean* juga memberikan hasil yang stabil karena data bersifat numerik kontinu (Supardi, 2024).

Kinerja model dapat ditingkatkan dengan penambahan jumlah data dari berbagai stasiun pemantauan serta penggunaan algoritma lain seperti *Support Vector Machine* atau *Random Forest* sebagai pembanding, sebagaimana dilakukan oleh Jayadi, Handhayani, dan Lauro (2023) yang memperoleh akurasi hingga 98% pada klasifikasi kualitas udara Jakarta.

Pembahasan

Hasil pengujian algoritma *K-Nearest Neighbor* (KNN) dalam klasifikasi kualitas udara di Jakarta menunjukkan tingkat akurasi sebesar 94,52% dengan nilai *class recall* tertinggi pada kategori “Sedang” sebesar 95,83% dan “Tidak Sehat” sebesar 92%, sedangkan kategori “Baik” memiliki nilai *recall* sebesar 0%. Hasil ini mengindikasikan bahwa model KNN mampu mengenali sebagian besar data dengan baik, terutama pada dua kategori dominan yang memiliki distribusi data terbesar. Nilai akurasi yang tinggi memperlihatkan bahwa algoritma KNN efektif digunakan dalam mengelompokkan data kualitas udara berdasarkan parameter polutan utama seperti PM_{2.5}, PM₁₀, CO, SO₂, NO₂, dan O₃.

Temuan ini sejalan dengan hasil penelitian oleh Ramadhan dan Triayudi (2024) dalam *SAGA: Journal of Technology and Information Systems*, yang melaporkan bahwa algoritma KNN mencapai tingkat akurasi hingga 95,13% dalam klasifikasi kualitas udara di Jakarta berdasarkan data ISPU, menunjukkan konsistensi kinerja tinggi dalam mendeteksi tingkat pencemaran udara. Kinerja model yang tinggi tersebut sejalan dengan temuan

Syahira et al. (2024) yang menyatakan bahwa algoritma KNN memberikan hasil akurasi di atas 90% dalam prediksi kualitas udara di Yogyakarta, terutama karena kemampuannya dalam menangani data non-linear dan multivariat. Selain itu, Alfarid dan Rahmah (2024) juga menunjukkan bahwa KNN memiliki performa yang kompetitif dibandingkan SVM dalam klasifikasi data Air Quality Index (AQI) Jakarta, dengan tingkat kesalahan relatif kecil pada kategori “Sedang”. Hal ini menunjukkan konsistensi efektivitas algoritma KNN dalam permasalahan klasifikasi kualitas udara di wilayah perkotaan.

Namun, nilai *recall* nol pada kategori “Baik” menandakan adanya ketidakseimbangan jumlah data (class imbalance) yang signifikan. Kasus seperti ini menyebabkan model cenderung bias terhadap kelas dengan jumlah data yang lebih besar. Fenomena ini sesuai dengan penjelasan Supardi (2024) bahwa akurasi model KNN sangat dipengaruhi oleh distribusi dataset dan nilai parameter k yang digunakan. Jika satu kelas memiliki jumlah data terlalu sedikit, maka jarak antar titik data dalam ruang fitur menjadi kurang representatif, sehingga model gagal mengenali pola dari kelas minoritas.

Selain itu, hasil penelitian ini juga memperkuat temuan Jayadi, Handhayani, dan Lauro (2023) bahwa walaupun algoritma SVM sedikit lebih unggul dalam hal akurasi, KNN tetap memiliki keunggulan dalam kesederhanaan komputasi dan efisiensi implementasi untuk sistem real-time. Oleh karena itu, penggunaan KNN dalam penelitian ini dapat menjadi alternatif yang tepat untuk sistem pemantauan kualitas udara berbasis sensor di lapangan yang membutuhkan kecepatan pemrosesan dan interpretasi hasil secara langsung.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma KNN masih relevan digunakan untuk klasifikasi kualitas udara, terutama jika dilakukan normalisasi data yang baik dan pemilihan

nilai k yang optimal. Namun demikian, peningkatan performa masih dapat dilakukan dengan teknik penyeimbangan data (*data balancing*), optimasi parameter, atau penerapan model *ensemble* yang dapat menggabungkan kekuatan beberapa algoritma.

SIMPULAN (PENUTUP)

Kesimpulan

Penelitian ini menunjukkan bahwa penerapan algoritma *K-Nearest Neighbor* (KNN) efektif dalam mengklasifikasikan kualitas udara di Jakarta berdasarkan parameter polutan utama, yakni PM2.5, PM10, CO, SO₂, NO₂, dan O₃. Dengan tingkat akurasi mencapai 94,52%, KNN terbukti mampu memberikan hasil klasifikasi yang cukup stabil untuk kategori “Sedang” dan “Tidak Sehat”. Temuan ini memperkuat pemahaman bahwa model berbasis jarak seperti KNN masih relevan digunakan dalam permasalahan lingkungan yang melibatkan data numerik dan bersifat non-linear.

Lebih dari sekadar perolehan angka akurasi, penelitian ini memaknai bahwa penerapan pembelajaran mesin sederhana dapat menjadi solusi praktis dalam sistem pemantauan kualitas udara perkotaan, khususnya di wilayah dengan keterbatasan sumber daya komputasi. Hasil ini juga menunjukkan potensi besar penggunaan KNN sebagai model pendukung pengambilan keputusan kebijakan lingkungan berbasis data.

Selain itu, hasil yang diperoleh mengindikasikan adanya keterbatasan pada kategori “Baik”, yang menunjukkan perlunya perhatian terhadap distribusi data yang tidak seimbang (*class imbalance*). Temuan ini menjadi pijakan bagi pengembangan metode klasifikasi yang lebih adaptif, baik melalui optimasi parameter, penerapan algoritma *hybrid*, maupun integrasi dengan model *ensemble learning*.

Saran

Penelitian selanjutnya disarankan untuk:

1. Menyeimbangkan jumlah data tiap kategori polusi udara dengan metode *resampling* seperti SMOTE agar model tidak bias terhadap kelas mayoritas.
2. Mengembangkan model hybrid yang mengombinasikan KNN dengan algoritma lain seperti SVM atau Random Forest untuk meningkatkan performa klasifikasi.
3. Mengintegrasikan sistem klasifikasi dengan sensor IoT dan *dashboard* berbasis web guna mendukung pemantauan kualitas udara secara real-time di wilayah perkotaan.
4. Menambah dimensi data temporal dan meteorologi (seperti suhu, kelembapan, dan arah angin) untuk memperkaya konteks prediksi kualitas udara.

Dengan demikian, penelitian ini tidak hanya memberikan kontribusi konseptual terhadap pengembangan teknologi klasifikasi berbasis *machine learning*, tetapi juga membuka peluang aplikasi nyata bagi sistem pemantauan lingkungan cerdas di masa depan.

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Ucapan terima kasih juga disampaikan kepada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, serta seluruh dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan berharga selama proses penelitian ini berlangsung.

DAFTAR PUSTAKA

- Tempo. (2024, Agustus 30). *Kualitas udara Jakarta memburuk, IQAir catat konsentrasi polutan terus meningkat*. Tempo.co. Retrieved from <https://www.tempo.co/lingkungan/kualitas-udara-jakarta-memburuk-iqair-catat-konsentrasi-polutan-terus-meningkat--14328>
- Fahmi, M., & Suhartana, I. K. G. (2022). *Perbandingan Algoritma Decision*

Tree dan Support Vector Machine dalam Prediksi Kualitas Udara. *Jurnal Nasional Teknologi Informasi dan Aplikasinya (JNATIA)*, 1(1), 21–30. <https://jurnal.harianregional.com/jnatia/id-92742>

Jayadi, V. B., Handhayani, T., & Lauro, M. D. (2023). *Perbandingan KNN dan SVM untuk Klasifikasi Kualitas Udara di Jakarta*. *Jurnal Ilmu Komputer dan Sistem Informasi*, 11(2), 91–98. <https://doi.org/10.24912/jiksi.v11i2.26006>

Lestari, I. G. A. N., & Mahendra, I. N. D. A. (2023). *Prediksi Kualitas Udara dengan Menggunakan Metode Long-Short Term Memory dan Artificial Neural Network*. *Jurnal Sistem dan Informatika (JSI)*, 17(2), 121–129. <https://doi.org/10.30864/jsi.v17i2.565>

Trihardianingsih, L., & Lasatira, G. S. (2023). *Optimasi Hyperparameter GridSearchCV pada Klasifikasi Kualitas Udara Menggunakan Support Vector Machine*. *Jurnal Informasi dan Teknologi*, 1(2), 65–72. <https://doi.org/10.30605/jitu.v1i2.65>

Syahira, N., Arianto, et al. (2024). *Prediksi kualitas udara di Yogyakarta dengan algoritma K-Nearest Neighbor*. *Jurnal Ilmiah Informatika Komputer*, 29(1). <https://doi.org/10.35760/ik.2024.v29i1.10069>

Supardi, N. E. (2024). *Klasifikasi Indeks Standar Pencemaran Udara di Jakarta dengan Metode K-Nearest Neighbors dan Artificial Neural Network*. *Jurnal Komputer dan Informatika (JKI)*, 19(2), 139–147. Retrieved from <https://journal.untar.ac.id/index.php/JKI/article/view/34593>

Nirmanda, P. (2024). *Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Konsentrasi Mahasiswa Program Studi Manajemen Universitas Muhammadiyah Makassar* [Skripsi, Universitas Muhammadiyah Makassar]. Universitas Muhammadiyah Makassar Repository.

- https://digilibadmin.unismuh.ac.id/upload/40843-Full_Text.pdf
- Wang, H., Xu, P., & Zhao, J. (2021). *Improved KNN Algorithm Based on Preprocessing of Center in Smart Cities. Complexity, 2021*, Article ID 5524388. <https://doi.org/10.1155/2021/5524388>
- Alfarid, M. R., & Rahmah, Q. H. (2024). *Implementasi prediksi kualitas udara Jakarta menggunakan K-Nearest Neighbor dan Support Vector Machine. Kohesi: Jurnal Sains dan Teknologi, 4(10)*, 91–100. Retrieved from <https://ejournal.warunayama.org/index.php/kohesi/article/view/6621/6095>
- Jayadi, V. B., Handhayani, T., & Lauro, M. D. (2023). *Perbandingan KNN dan SVM untuk Klasifikasi Kualitas Udara di Jakarta. Jurnal Ilmu Komputer dan Sistem Informasi, 11(2)*, 91–98. <https://doi.org/10.24912/jiksi.v11i2.26006>
- Syahira, N., Arianto, A., & Salsabila, R. (2024). *Prediksi kualitas udara di Yogyakarta dengan algoritma K-Nearest Neighbor. Jurnal Ilmiah Informatika Komputer, 29(1)*. <https://doi.org/10.35760/ik.2024.v29i1.10069>
- Ramadhan, D. P., & Triayudi, A. (2024). *Jakarta air quality classification based on air pollutant standard index using C4.5 and naïve Bayes algorithms. SAGA: Journal of Technology and Information Systems, 2(4)*, 311–327. <https://journal.mediadigitalpublikasi.com/index.php/saga/article/view/395>
- Pagan, M., Zarlis, M., & Candra, A. (2023). *Investigating the impact of data scaling on the K-nearest neighbor algorithm. Computer Science and Information Technologies, 4(2)*, 135–142. <https://www.iaesprime.com/index.php/csit/article/view/279>
- Atiq ur Rehman, & Belhaouari, S. B. (2021). *Unsupervised outlier detection in multidimensional data. Journal of Big Data, 8*, Article 80. <https://doi.org/10.1186/s40537-021-00469-z>
- Kim, T., Kim, J., Yang, W., Lee, H., & Choo, J. (2021). *Missing value imputation of time-series air-quality data via deep neural networks. International Journal of Environmental Research and Public Health, 18(22)*, 12213. <https://doi.org/10.3390/ijerph182212213>