

ANALISIS ALGORITMA DECISION TREE DALAM PENGKLASIFIKASIAN INDEKS PENCEMARAN UDARA KOTA JAKARTA DENGAN METODE CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING

Decision Tree Algorithm Analysis In Classification Of The Air Pollution Index In Jakarta City Using The Cross Industry Standard Process For Data Mining Method

Anisa Putri Arla Vatwa Lubu^{1,a}, Siti Mutiah², Arya Praditya³, Nerisa Rahma⁴, Prizka Rismawati Arum⁵

¹Universitas Muhammadiyah Semarang [Email: anisaputriarla02@gmail.com]

²Universitas Muhammadiyah Semarang [Email: sitimutiah09@gmail.com]

³Universitas Muhammadiyah Semarang [Email: aryapraditya389@gmail.com]

⁴Universitas Muhammadiyah Semarang [Email: nerisarha03@gmail.com]

⁵Universitas Muhammadiyah Semarang [Email: prizka.rismawati@unimus.ac.id]

^a*anisaputriarla02@gmail.com*

ABSTRAK

Kualitas udara yang bersih sangat penting untuk kelangsungan hidup manusia. Namun, DKI Jakarta saat ini menghadapi tantangan serius dengan kualitas udara terburuk di dunia, yang disebabkan oleh aktivitas manusia, termasuk industri dan penggunaan bahan bakar fosil. Dalam konteks ini, model klasifikasi, khususnya algoritma Decision Tree, dapat berperan dalam memahami faktor-faktor yang mempengaruhi kualitas udara serta mengklasifikasikan Indeks Standar Pencemar Udara (ISPU). Kajian ini bertujuan untuk menganalisis klasifikasi dengan menggunakan metode CRISP-DM guna mengidentifikasi pola dan parameter yang memengaruhi pencemaran udara. Penelitian ini mengevaluasi enam parameter, yaitu karbon monoksida (CO), sulfur dioksida (SO₂), nitrogen dioksida (NO₂), ozon (O₃), serta partikel debu PM2.5 dan PM10. Kategori level ISPU yang dianalisis meliputi Baik, Sedang, dan Tidak Sehat. Hasil penelitian menunjukkan bahwa model yang digunakan memiliki performa yang sangat baik, dengan akurasi mencapai 97,01%. Dari analisis, PM2.5 ditemukan memiliki korelasi tertinggi terhadap Indeks Standar Pencemar Udara, sementara ozon terbukti efektif dalam membedakan antara kualitas udara yang sedang dan tidak sehat.

Kata kunci: CRISP-DM, *Decision Tree*, Klasifikasi, Kualitas Udara

ABSTRACT

Clean air quality is very important for human survival. However, DKI Jakarta is currently facing serious challenges with the worst air quality in the world, which is caused by human activities, including industry and the use of fossil fuels. In this context, classification models, especially the Decision Tree algorithm, can play a role in understanding the factors that influence air quality and classifying the Air Pollution Standard Index (ISPU). This study aims to analyze classification using the CRISP-DM method to identify patterns and parameters that influence air pollution. This research evaluates six parameters, namely carbon monoxide (CO), sulfur dioxide (SO₂), nitrogen dioxide (NO₂), ozone (O₃), and PM2.5 and PM10 dust particles. The ISPU level categories analyzed include Good, Medium and Unhealthy. The research results show that the model used has very good performance, with an accuracy of 97.01%. From the analysis, PM2.5 was found to have the highest correlation to the Air Pollution Standard Index, while ozone was proven to be effective in distinguishing between moderate and unhealthy air quality.

Keywords: CRISP-DM, *Decision Tree*, Classification, Air Quality

1. PENDAHULUAN

Udara memainkan peran krusial dalam mendukung kehidupan makhluk hidup di Bumi. Proses metabolisme yang esensial bagi organisme tidak dapat berjalan tanpa adanya oksigen yang diambil dari udara [1]. Meskipun oksigen merupakan komponen utama, udara juga mengandung berbagai zat lain, seperti karbon monoksida, karbon dioksida, sulfur dioksida, nitrogen oksida, ozon, dan zat-zat lainnya [2]. Beberapa dari zat-zat ini masih dapat ditoleransi oleh tubuh selama konsentrasiannya berada di bawah

batas normal [3]. Namun, jika melebihi batas yang ditentukan, dapat menyebabkan dampak serius pada kesehatan manusia, seperti penyakit jantung, infeksi pernapasan, stroke, asma, hingga penyakit paru [4]. Salah satu faktor penyebab peningkatan konsentrasi zat-zat tersebut dalam udara adalah aktivitas manusia [5].

Kualitas udara menjadi elemen kritis dalam dinamika perkotaan, dengan dampak langsung pada kesehatan masyarakat dan keberlanjutan lingkungan. DKI Jakarta, sebagai pusat ekonomi dan urbanisasi yang berkembang pesat di Indonesia, memperlihatkan tingkat kualitas udara yang mengkhawatirkan, menjadikannya salah satu kota dengan tingkat polusi udara tertinggi di dunia [6]. Menurut pemanfaatan Kualitas Udara oleh US Air Quality Index (AQI) pada kuartal ketiga tahun 2019, Jakarta pernah mencapai peringkat teratas sebagai kota dengan kualitas udara terburuk, mencapai angka 179, yang dikategorikan sebagai tidak sehat, terutama melalui parameter PM_{2,5} dengan konsentrasi hingga 110 µg/m³ [7]. Selain itu, studi yang dilakukan oleh [8] di DKI Jakarta menunjukkan bahwa selama periode 2017 hingga 2019, kualitas udara terus memburuk, ditandai dengan peningkatan jumlah hari dengan tingkat kesehatan udara yang memburuk setiap tahunnya. Meskipun terjadi penurunan aktivitas manusia pada tahun 2020 dan 2021 akibat pandemi COVID-19, kualitas udara DKI Jakarta tetap mencapai tingkat sedang hingga tidak sehat. Penyebab utama dari kondisi ini, selain dari aktivitas manusia pada umumnya, adalah kontribusi dari pabrik industri dan pembangkit listrik berbahan bakar fosil di sepanjang wilayah administratif Jakarta [9].

Berdasarkan Keputusan Badan Pengendalian Dampak Lingkungan (Bapedal) Nomor KEP-107/Kabapedal/11/1997, pemerintah Indonesia menetapkan Indeks Standar Pencemar Udara (ISPU) sebagai acuan untuk mengevaluasi kualitas udara di suatu wilayah dan dampaknya terhadap kesehatan manusia setelah menghirup udara tersebut dalam rentang waktu beberapa jam hingga beberapa hari. Tingkatan ISPU menjadi penentu sejauh mana udara yang dihirup dapat memengaruhi kesehatan, di mana semakin tinggi level ISPU, semakin buruk kualitas udara yang terhirup oleh individu. ISPU dibagi menjadi lima level, yakni Baik, Sedang, Tidak Sehat, Sangat Tidak Sehat, dan Berbahaya [10]. Penelitian tentang implementasi data mining untuk menentukan Indeks Standar Pencemar Udara juga telah dilakukan [11] dengan menggunakan algoritma Decision Tree untuk menentukan kualitas udara di provinsi DKI Jakarta. Berdasarkan hasil pengujian menunjukkan bahwa klasifikasi dengan menggunakan Decision Tree mendapatkan nilai Akurasi sebesar 95,89%.

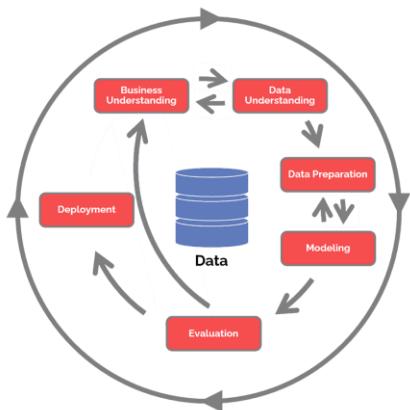
Decision Tree atau Pohon Keputusan adalah suatu model prediktif dalam analisis data yang memanfaatkan struktur berbentuk diagram pohon. Dalam diagram ini, setiap node (simpul) mewakili keputusan atau nilai-nilai tertentu, sedangkan cabang-cabangnya mencerminkan hubungan antara keputusan tersebut dengan nilai-nilai yang terlibat [12]. Pada setiap simpul keputusan, pohon tersebut bercabang sesuai dengan fitur atau atribut tertentu, dan proses ini berlanjut hingga mencapai simpul daun (leaf node) yang menghasilkan prediksi akhir atau keputusan akhir [13].

Kajian mengenai Indeks Kualitas Udara (AQI) bertujuan untuk memahami kondisi udara di suatu wilayah dan dapat memanfaatkan teknik data mining untuk melihat pola pencemaran udara. Dengan menggunakan pendekatan ini, peneliti dapat mengidentifikasi parameter-parameter yang berperan dalam meningkatkan intensitas pencemaran udara. Data yang dihasilkan akan memberikan kontribusi penting bagi pemerintah, masyarakat, dan sektor industri dalam merancang kebijakan, melindungi kesehatan, serta mengurangi emisi gas berbahaya. Semakin dini penelitian ini dilakukan, maka semakin cepat langkah-langkah penanganan polusi udara dapat diimplementasikan sehingga dapat membawa dampak positif bagi kesehatan seluruh warga.

2. METODE PENELITIAN

Penelitian ini menggunakan dataset Indeks Pencemar Standar Udara DKI Jakarta dikumpulkan dari laman situs terbuka pemerintah DKI Jakarta <http://www.data.jakarta.go.id/> berupa berkas CSV yang diseleksi per SPKU dan per daerah, yaitu Jakarta Pusat (DKI1), Jakarta Utara (DKI2), Jakarta Selatan (DKI3), Jakarta Timur (DKI4), dan Jakarta Barat (DKI5) pada tahun 2021. Penelitian ini menggunakan metodologi CRISP-DM dalam memecahkan masalah yang umum di bisnis dan penelitian misalnya da-

lam pengembangan sistem peringatan dini untuk polusi udara. Proses CRISP-DM ini dimulai dari business understanding, data understanding, data preparation, modelling, evaluation, dan deployment. Dalam penelitian ini, algoritma yang digunakan adalah Decision Tree, dengan tahap evaluation menggunakan nilai confusion matrix (accuracy), Recall, dan Precision kemudian dianalisis untuk membentuk pengetahuan baru [14]. Penelitian untuk mengklasifikasikan level Indeks Standar Polusi Udara sebagai penentu kualitas udara di DKI Jakarta menggunakan skema alur penelitian sebagai berikut: Bagian ini memuat metode penelitian yang digunakan untuk mencapai hasil penelitian. Cara atau metode penelitian mengandung uraian tentang cara-cara pelaksanaan penelitian, yang mencakup subjek penelitian, populasi dan sampel, teknik pengumpulan data, dan teknik analisis data.



Gambar 1 Metodologi CRISP-DM dengan tahapan business understanding, data understanding, data preparation, modelling, evaluation, dan deployment.

2.1 Business Understanding

Dalam tahapan ini ada beberapa hal yang dilakukan seperti memahami kebutuhan serta tujuan dari bisnis yang selanjutnya akan didefinisikan ke dalam data mining, kemudian menentukan rencana dan strategi untuk mencapai tujuan data mining. Adapun yang menjadi masalah dari penelitian ini adalah Indeks Standar Pencemaran Udara di wilayah Kota Jakarta [15]. Domain Understanding merupakan istilah lain pada proses ini, karena tidak semua subjek dalam standard process for data mining merupakan sebuah entitas bisnis [16].

2.2 Data Understanding

Tahapan data understanding berfokus pada pemahaman awal mengenai data yang dibutuhkan untuk memecahkan permasalahan bisnis yang diberikan. Tujuannya adalah untuk menentukan tujuan, mengidentifikasi data yang relevan, dan memahami kebutuhan pengguna [17].

2.3 Data Preparation

Pada tahap ini, data akan dibersihkan dengan menghilangkan data yang memiliki missing value, data duplikat, data yang tidak lengkap, dan atribut yang tidak digunakan. Selanjutnya, karena klasifikasi merupakan pembelajaran supervised learning, maka diperlukan tahap pelabelan pada atribut yang akan diklasifikasikan [18].

2.4 Modelling

Fokus tahap ini adalah pengembangan model data mining. Langkah-langkahnya mencakup pemilihan algoritma yang sesuai, pelatihan model, dan evaluasi model yang telah dibuat. Penelitian ini menggunakan model klasifikasi dengan algoritma Decision Tree. Algoritma tersebut dipilih karena dapat memproses data dalam bentuk polynomial, sesuai dengan format label ISPU. Algoritma Tree biasa dipakai untuk pengenalan pola statistik. Decision Tree terbuat dari tiga simpul yaitu leaf, lalu terdiri juga dari simpul root yang merupakan titik awal dari suatu decision tree, dan yang terakhir adalah simpul perantara yang berhubungan dengan suatu pengujian [13]. Untuk menentukan atribut mana yang paling tepat untuk menjadi akar pertama dalam algoritma decision tree diperlukan perhitungan nilai

gain dari masing-masing atribut. Nilai gain yang paling tinggi akan menjadi akar pertama. Namun, sebelum melakukan perhitungan nilai gain, diperlukan nilai entropi dari kumpulan data terlebih dahulu. Nilai entropi dapat dihitung menggunakan persamaan 1.

$$Entropy(S) = \sum_{i=1}^n -pi \cdot \log_2 pi \quad (1)$$

Dimana S merupakan himpunan kasus, n adalah jumlah partisi S , dan pi adalah proporsi S_i terhadap S . Selanjutnya nilai gain dapat dihitung menggunakan persamaan 2.

$$Gain(S, A) = Entropy(S) - \sum_{t=1}^n \frac{|S_t|}{|S|} \times Entropy(S_t) \quad (2)$$

Dengan S merupakan himpunan kasus, A adalah fitur, dan n adalah jumlah partisi atribut A . Untuk proporsi S_i terhadap S diwakili oleh $|S_t|$ dan jumlah kasus terhadap S diwakili dengan $|S|$.

Pada tahap ini juga akan dilakukan analisa untuk mengidentifikasi pengaruh pada masing-masing atribut menggunakan korelasi pearson. Korelasi pearson adalah salah satu dari pengujian korelasi yang digunakan dalam mengetahui derajat keeratan hubungan dua variabel yang memiliki interval atau rasio, berdistribusi normal, serta mengembalikan nilai koefisien korelasi dengan rentang nilai antara -1, 0 dan 1 [19]. Nilai koefisien korelasi pearson dapat dihitung menggunakan persamaan 3.

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{i=1}^n Y_i^2 (\sum_{i=1}^n Y_i)^2}} \quad (3)$$

Semakin mendekati angka 1 sebuah nilai korelasi yang dihasilkan, maka semakin kuat hubungan atribut tersebut, dan semakin mendekati angka 0 sebuah nilai korelasi maka semakin lemah hubungan atribut tersebut.

2.5. Evaluation

Tahap ini bertujuan untuk mengevaluasi kinerja model data mining dan menentukan apakah model tersebut layak digunakan [20]. Hasil analisis tersebut akan menghasilkan nilai akurasi, presisi, dan recall yang didapatkan dari confusion matrix yang dapat dilihat dari Tabel berikut ini:

Tabel 1 Confusion Matrix

		Actual Positive	Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Perhitungan nilai akurasi, presisi, dan recall dapat dihitung menggunakan persamaan 4, 5, dan 6.

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{i=1}^n Y_i^2 (\sum_{i=1}^n Y_i)^2}} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Implementasi dalam penelitian ini terkait dengan data pencemaran udara di suatu area dan dapat menggunakan *data mining* untuk melihat pola pencemaran udara, mengidentifikasi parameter-parameter apa saja yang mempengaruhi tingginya intensitas pencemaran udara di wilayah DKI Jakarta, dan

memberikan informasi yang berharga bagi pemerintah, masyarakat, dan industri. Hasil penelitian diharapkan dapat meningkatkan kesadaran akan dampak pencemaran udara, mendorong usaha pengurangannya, dan akhirnya menciptakan lingkungan yang lebih sehat dan berkelanjutan.

3.2 Data Understanding

Hasil seleksi pengumpulan dataset Indeks Pencemar Standar Udara (ISPU) di wilayah Kota Jakarta pada tahun 2021 adalah sebagai berikut:

Tabel 2 Jumlah Record Dataset Tahun 2021

Bulan	Jumlah Record dalam seluruh dataset
Januari	31
Februari	28
Maret	31
April	30
Mei	31
Juni	30
Juli	31
Agustus	31
September	30
Okttober	31
November	30
Desember	31
Total	365

3.3 Data Preparation

Dalam meningkatkan efisiensi dan keperluan klasifikasi level ISPU, perlu dilakukan penyesuaian terhadap 12 dataset yang terdiri dari 365 record. Hal ini meliputi pemilihan atribut yang relevan serta perlu dilakukan proses *data preparation* pada atribut tertentu sebagaimana dapat dilihat pada Tabel 3 berikut.

Tabel 3 Atribut Data ISPU DKI Jakarta

Atribut	Data Kosong	Proses	Keterangan
Tanggal	0	Tidak digunakan	Waktu pengambilan data ISPU
PM10	31	Digunakan	Partikel debu yang berukuran kurang dari 10 mikrometer
PM2.5	0	Digunakan	Partikel debu yang berukuran kurang dari 2,5 mikrometer
SO2	0	Digunakan	Sulfur dioksida
CO	0	Digunakan	Karbon monoksida
O3	0	Digunakan	Ozon
NO2	0	Digunakan	Nitrogen oksida
Categori	0	Digunakan	Kategori hasil perhitungan tingkat ISPU

Location	0	Tidak digunakan	Lokasi SPKU yang melakukan pengambilan data
----------	---	-----------------	---

Tabel 3 di atas, terdapat beberapa atribut yang tidak digunakan, diantaranya adalah Tanggal dan *Location*. Untuk atribut PM2,5 diperlukan tahap *preparation* untuk mengatasi data yang kosong yakni penghapusan data tersebut. Sedangkan untuk atribut *Categori* walaupun tidak terdapat data kosong namun perlu dilakukan *preparation data* untuk dilakukan proses transformasi data yang bertujuan mengubah nilai-nilai data ke dalam skala 1, 2, dan 3. Setelah itu, dilakukan pembersihan data terhadap nilai duplikat untuk mencegah terjadinya bias dalam proses klasifikasi.

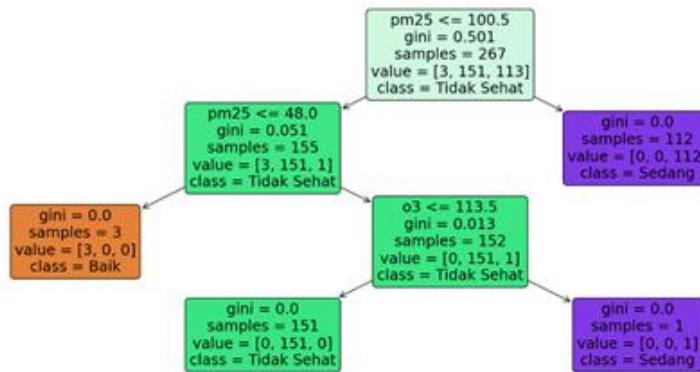
3.4 Modeling

Pada tahap ini, model klasifikasi menghasilkan pola informasi yang mendukung para pemangku kepentingan dalam memprediksi pencemaran udara serta mengidentifikasi parameter yang mempengaruhinya. Proses ini dilakukan menggunakan Google Collaboratory dan bahasa pemrograman Python, dengan mengintegrasikan 12 dataset menjadi satu yang terdiri dari 334 record setelah proses persiapan. Selanjutnya, identifikasi korelasi antar atribut dilakukan dengan menggunakan metode korelasi Pearson untuk analisis yang lebih mendalam.

	pm10	pm25	so2	co	o3	no2	max	categori
pm10	1.000000	0.779663	0.414410	0.329441	0.182792	0.404055	0.881248	0.613175
pm25	0.779663	1.000000	0.454264	0.266070	0.171012	0.391858	0.950698	0.761951
so2	0.414410	0.454264	1.000000	0.156762	0.247506	0.454816	0.440509	0.302462
co	0.329441	0.266070	0.156762	1.000000	0.068663	0.729635	0.269060	0.183227
o3	0.182792	0.171012	0.247506	0.068663	1.000000	0.115592	0.229305	0.100401
no2	0.404055	0.391858	0.454816	0.729635	0.115592	1.000000	0.377288	0.253267
max	0.881248	0.950698	0.440509	0.269060	0.229305	0.377288	1.000000	0.775921
categori	0.613175	0.761951	0.302462	0.183227	0.100401	0.253267	0.775921	1.000000

Gambar 2 Hasil analisa masing-masing atribut pada dataset Indeks Standar Pencemar Udara di wilayah Kota Jakarta pada tahun 2021

Dalam penelitian ini menggunakan 7 atribut, diantaranya adalah PM10, PM2.5, SO2, CO, O3, NO2, dan Categori. Hasil korelasi matriks ini menunjuk bahwa atribut PM2,5 memiliki korelasi yang cukup tinggi dengan indeks kualitas udara di wilayah kota Jakarta dengan nilai korelasi sebesar 0,762. Selanjutnya dilakukan proses *data mining* menggunakan algoritma *Decision Tree* untuk dijadikan model klasifikasi menggunakan *feature selection* berupa *backward elimination* untuk mereduksi informasi yang tidak relevan dan juga *holdout validation* untuk membagi dataset menjadi data latih dan data uji dengan pembagian 0,2 dimana 80% merupakan data latih dan 20% menjadi data uji. *Random state* yang digunakan adalah 42, hal ini dapat membantu memastikan hasil yang konsisten dari algoritma *machine learning*, sehingga memudahkan untuk membandingkan kinerja algoritma yang berbeda. Berikut hasil pengolahan data algoritma *Decision Tree*.



Gambar 3 Hasil pohon keputusan setelah dilakukan pengolahan data menggunakan algoritma *decision tree*

Hasil ini merupakan visualisasi pengolahan data pada algoritma *decision tree* menghasilkan 3 *nodes*, 4 *leaf*, dan 3 tingkat kedalaman. Pada *root node* memiliki *splitting* kriteria pada atribut PM_{2,5} \leq 100,5 dengan kata lain hal ini memanfaatkan atribut PM_{2,5} sebagai pemecah kondisi sebelum proses *splitting*. Berdasarkan nilai atribut O₃ \leq 113,5 dapat dilihat dari nilai *gini impurity* yang rendah pada node 3 adalah 0,013 yang kemudian membagi data menjadi dua *class* yang murni, yaitu *class* "Tidak Sehat" dan *class* "Sedang". Pada *leaf* 4 nilai *gini impurity* adalah 0,0 hal ini menunjukkan bahwa nilai O₃ (ozon) memiliki kemampuan untuk membedakan antara dua kelas tersebut dengan efektif. Nilai ozon yang tinggi menunjukkan bahwa kualitas udara tidak sehat, sedangkan nilai ozon yang rendah menunjukkan bahwa kualitas udara sedang.

3.5 Evaluation

Hasil pengujian algoritma pohon keputusan menghasilkan Tabel *confusion matrix* yang dapat dilihat pada Tabel 4 dan tabel *confusion matrix* tersebut dapat direpresentasikan ke dalam bentuk nilai akurasi, presisi, dan *recall* pada Tabel 5.

Tabel 4 Confusion Matrix Decision Tree

	TB	TS	TTS
PB	0	0	0
PS	1	43	0
PTS	0	1	22

Keterangan:

- TB : *True Baik*
- TS : *True Sedang*
- TTS : *True Tidak Sehat*
- PB : Prediksi Baik
- PS : Prediksi Sedang
- PTS : Prediksi Tidak Sehat

Tabel 5 Hasil Evaluasi Algoritma *Decision Tree*

Evaluasi	Hasil
Accuracy	0,9701
Precision	0,9850
Recall	0,9701

Dari hasil evaluasi pada tabel 5 diketahui bahwa algoritma *decision tree* menunjukkan nilai *accuracy* sebesar 97,01% dengan jumlah prediksi benar sebanyak 65 data dari 67 data yang diuji.

3.6 Deployment

Model klasifikasi data mining menggunakan algoritma Decision Tree memberikan wawasan berharga dalam memantau kualitas udara. Atribut O3 dapat dengan jelas membedakan antara kategori kualitas udara "Sedang" dan "Tidak Sehat," yang sangat membantu bagi para ahli lingkungan dan masyarakat dalam mengambil tindakan pencegahan yang diperlukan. Selain itu, PM2,5 menunjukkan korelasi yang signifikan dengan Indeks Standar Pencemar Udara (ISPU), mengindikasikan tingkat polusi yang tinggi di Jakarta. Partikel PM2,5, yang dihasilkan dari pembakaran bahan bakar dan aktivitas industri, dapat menimbulkan masalah kesehatan yang serius. Di Jakarta, ISPU PM2,5 sering kali mencapai tingkat yang tidak sehat, sehingga sangat penting untuk melakukan upaya pengurangan emisi dari sektor transportasi dan industri demi meningkatkan kualitas udara serta kesehatan masyarakat.

4. KESIMPULAN DAN SARAN

Penelitian dengan menggunakan dataset Indeks Pencemar Udara di wilayah Kota Jakarta dengan algoritma *Decision Tree* menggunakan *holdout validation* 0,2 menghasilkan *testing data* sebanyak 67 *record* yang kemudian dianalisis dan mendapatkan hasil dengan performa *accuracy* yang baik sebesar 97,01%. Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Model klasifikasi data mining dengan menggunakan algoritma Decision Tree dapat memberikan informasi yang berguna dalam memantau kualitas udara di wilayah Kota Jakarta. Model ini dapat digunakan untuk memprediksi tingkat pencemaran udara berdasarkan data-data yang tersedia seperti tingkat konsentrasi polutan.
2. Atribut yang memiliki korelasi tertinggi terhadap indeks standar pencemar udara adalah PM2,5. Hal ini menunjukkan bahwa konsentrasi PM2,5 di wilayah Kota Jakarta cukup tinggi dan menjadi salah satu faktor utama yang mempengaruhi tingginya intensitas pencemaran udara di kota tersebut.
3. Atribut O3 terbukti memiliki kemampuan untuk membedakan antara kualitas udara yang "Sedang" dan "Tidak Sehat". Hal ini menunjukkan bahwa konsentrasi ozon yang tinggi dapat menjadi indikator awal terjadinya pencemaran udara yang tidak sehat.

5. UCAPAN TERIMAKASIH

Terima kasih kepada Ibu Prizka Rismawati Arum selaku dosen pendamping yang selalu memberikan arahan, semangat, dan motivasi sehingga proposal artikel ilmiah ini dapat terselesaikan dengan baik. Terima kasih juga kepada Bapak M. Al-Haris selaku dosen pembimbing Studi Independen yang telah memberikan usulan serta masukan sehingga peneliti dapat berkesempatan untuk melakukan penelitian ini. Terakhir, terima kasih juga kepada seluruh teman-teman Statistika yang telah membantu selama proses penelitian dan pembuatan proposal artikel ilmiah.

DAFTAR PUSTAKA

- [1] N. I. Mansyur, E. H. Pudjiwati, and A. Murtilaksono, *Pupuk dan Pemupukan*. Syiah Kuala University Press, 2021.
- [2] M. S. Novelan, "Sistem Monitoring Kualitas Udara Dalam Ruangan Menggunakan Mikrokontroler dan Aplikasi Android," *Jurnal Nas. Inform. dan Teknol. Jar.*, pp. 2540–7597, 2020.
- [3] Z. Zhang *et al.*, "Recent Advances In Carbon Dioxide Utilization," *Renew. Sustain. energy Rev.*, vol. 125, 2020.
- [4] H. A. Shahriyari *et al.*, "Air Pollution And Human Health Risks: Mechanisms And Clinical Manifestations Of Cardiovascular And Respiratory Diseases," *Toxin Rev.*, vol. 41, no. 2, pp. 606–617, 2022.
- [5] A. A. Almetwally, M. Bin-Jumah, and A. A. Allam, "Ambient Air Pollution And Its Influence On Human Health And Welfare: An Overview," *Env. Sci Pollut Res*, vol. 27, pp. 24815–

24830, 2020.

- [6] S. Nurjanah, A. M. Siregar, and D. S. Kusumaningrum, “Penerapan Algoritma K-Nearest Neighbor (KNN) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta,” *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 2, pp. 71–76, 2020.
- [7] A. Oktaviani and Hustinawati, “Prediksi Rata-Rata Zat Berbahaya Di DKI Jakarta Berdasarkan Indeks Standar Pencemar Udara Menggunakan Metode Long Short-Term Memor,” *J. Ilm. Inform. Komput.*, vol. 26, no. 1, pp. 41–55, 2021.
- [8] L. Myllyvirta, I. Suarez, E. Uusivuori, and Hubert Thieriot, “Pencemaran Udara Lintas Batas Di Provinsi Jakarta, Banten, Dan Jawa Barat,” *Cent. Res. Energy Clean Air*, pp. 1–39, 2020.
- [9] S. H. D. A. Anggraeni, Y. H. Darundiati, and Tri Joko, “Analisis Konsentrasi PM10 Hasil Pengukuran Stasiun BMKG Kemayoran di Jakarta Pusat Pada Masa Pandemi COVID-19,” *Media Kesehat. Masy. Indones.*, vol. 20, no. 1, pp. 64–69, 2021.
- [10] S. S. A. Umri, “Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Dki Jakarta,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 98–104, 2021.
- [11] Irwansyah, A. D. Wiranata, and T. T. M, “Komparasi Algoritma Decision Tree, Naive Bayes Dan Knearest Neighbor Untuk Menentukan Kualitas Udara di Provinsi DKI Jakarta,” *Infotech J. Technol. Inf.*, vol. 9, no. 2, pp. 193–198, 2023.
- [12] A. H. Nasrullah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris,” *J. Ilm. Ilmu Komput. Fak. Ilmu Komput. Univ. Al Asyariah Mandar*, vol. 7, no. 2, pp. 45–51, 2021.
- [13] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, pp. 646–654, 2020.
- [14] N. C. Sastya and I. Nugraha, “Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution,” *J. Pendidik. Dan Apl. Ind.*, vol. 10, no. 2, pp. 103–115, 2023.
- [15] C. Maisyarah, E. Harytami, R. Y. Fajriatifah, and Y. H. Puspita, “Prediksi Penyakit Diabetes menggunakan Algoritma Artificial Neural Network,” *J. Data Sci. Inform.*, vol. 2, no. 2, pp. 46–52, 2022.
- [16] D. A. Kusuma and A. D. P. Wicaksono, “Analisis Klastering Dampak Lingkungan Berdasarkan Konsumsi Energi Perusahaan Berbasis Industri 4.0 Menggunakan Metode Crisp-Dm,” *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, 2023.
- [17] M. A. Hasanah, S. Soim, and Ade Silvia Handayani, “Implementasi Crisp-dm Model Menggunakan Metode Decision Tree dengan Algoritma Cart untuk Prediksi Curah Hujan Berpotensi Banjir,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021.
- [18] Y. A. Singgalen, “Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT),” *J. Media Inform. Budidarma*, vol. 7, no. 3, 2023.
- [19] T. Fu, X. Tang, Z. Cai, Y. Zuo, Y. Tang, and X. Zhao, “Correlation Research Of Phase Angle Variation And Coating Performance By Means Of Pearson’s Correlation Coefficient,” *Prog. Org. Coatings*, vol. 139, 2020.
- [20] D. A. Munawwaroh and A. H. Primandari, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree Dengan Algoritma CART Untuk Prediksi Lila Ibu Hamil Berpotensi Gizi Kurang,” *Delta J. Ilm. Pendidik. Mat.*, vol. 10, no. 2, 2022.