

BAB I

PENDAHULUAN

1.1 Latar Belakang

Udara terdiri dari kombinasi gas alam seperti nitrogen, oksigen, argon, dan karbon dioksida. Udara memainkan peran penting dalam kehidupan berbagai spesies organisme hidup, meski tak kasat mata, namun kehadirannya dapat dirasakan. Udara bersih memiliki beberapa manfaat, antara lain menjaga kesehatan pernafasan, mengurangi risiko penyakit kronis, memperpanjang umur, meningkatkan stamina dan konsentrasi, serta meningkatkan suasana hati (Kirono et al., 2022).

Pencemaran udara terjadi ketika zat-zat pencemar, baik fisik, kimiawi, maupun biologis, masuk ke dalam udara pada tingkat yang membahayakan kesehatan manusia, hewan, dan tumbuhan. *World Health Organization (WHO)* melaporkan bahwa pencemaran udara mengakibatkan jutaan kematian setiap tahunnya dan menyebabkan kerusakan lingkungan, terutama di wilayah perkotaan besar seperti DKI Jakarta, serta menimbulkan berbagai masalah kesehatan, seperti penyakit pernafasan dan kardiovaskular. Lalu lintas, operasi industri, dan aktivitas pengelolaan limbah merupakan kontributor utama polusi udara, yang berdampak negatif terhadap kualitas udara, merusak ekosistem, dan mengancam kesehatan organisme hidup dalam jangka panjang (Wiranata et al., 2023).

Kondisi atmosfer suatu wilayah tertentu berkaitan dengan kualitas udara yang dipengaruhi oleh berbagai faktor. Pencemaran udara dinilai dengan menggunakan indikator tanpa satuan yang disebut Indeks Standar Pencemar Udara (ISPU). Tujuan ISPU adalah memberikan informasi kualitas udara yang terpercaya kepada masyarakat dan menjadi acuan pemerintah dalam pengelolaan pencemaran udara (Budianita et al., 2024).

Udara mengandung berbagai macam polutan di antaranya partikel debu (pm10 dan pm2.5), sulfur dioksida (so2), karbon monoksida (co), ozon permukaan (o3), dan nitrogen dioksida (no2). Bahan kimia dalam jumlah besar dapat membahayakan kesehatan masyarakat. Oleh karena itu, sangat penting untuk mengukur dan menilai kualitas udara untuk mendukung implementasi kebijakan pemerintah (Nababan et al., 2023).

Dinas Lingkungan Hidup Provinsi DKI Jakarta membutuhkan proses pengolahan data yang disebut *data mining* untuk mendapatkan informasi yang dibutuhkan. *Data mining* melibatkan analisis data dalam jumlah besar untuk mendeteksi pola, tren, dan hubungan yang dapat memberikan wawasan berharga. Metode ini menggabungkan berbagai teknik dari berbagai bidang, termasuk pembelajaran mesin, pemrosesan sinyal, database, statistik, dan analisis data spasial atau temporal (Setiawan & Triayudi, 2022). Kualitas udara di DKI Jakarta dapat diukur dengan menerapkan *data mining* yang menggunakan algoritma klasifikasi yaitu memeriksa elemen data dan mengorganisasikannya ke dalam kelompok tertentu. Dalam kategorisasi ini, algoritma seperti C4.5 dan Naïve Bayes digunakan (Valentino Jayadi et al., 2023).

Algoritma klasifikasi pohon keputusan C4.5 sering digunakan dalam metode ekstraksi data dan informasi. Algoritma C4.5 dipilih karena keunikannya yang membedakannya dengan algoritma lain berdasarkan beberapa ciri utama, seperti mudah dipahami karena menghasilkan aturan yang jelas dalam bentuk pohon keputusan serta bekerja dengan baik untuk dataset yang memiliki struktur aturan berbasis batasan nilai (Astriyani et al., 2023). Selain itu, penelitian ini menggunakan Naïve Bayes. Algoritma Naïve Bayes, berdasarkan probabilitas dan statistik, digunakan untuk menentukan probabilitas suatu peristiwa di kemudian hari berdasarkan pengalaman sebelumnya. Klasifikasi Naïve Bayes menyatakan bahwa terdapat perbedaan antara keberadaan suatu fitur di suatu kelas dengan kehadirannya di kelas lain (Apridiansyah et al., 2021). Algoritma Naïve Bayes dipilih karena pendekatan

probabilistik yang dapat menangani variasi dalam data dan mengasumsikan independensi antar atribut (Meena et al., 2024).

Penelitian sebelumnya telah menggunakan berbagai algoritma klasifikasi untuk menilai tingkat ISPU di berbagai wilayah geografis. Sebagai contoh pada tahun 2022, penelitian di Tangerang Selatan menggunakan algoritma K-Nearest Neighbor (KNN) serta Naïve Bayes untuk memprediksi polusi udara, sehingga menghasilkan akurasi KNN sebesar 94,44% dan Naïve Bayes 86,11%. Penelitian lain di DKI Jakarta pada tahun 2023 menggunakan algoritma Naïve Bayes dengan hasil penelitian mencapai 91,96%.

Ilustrasi ketiga yaitu penelitian di Bandung akan mengkaji tiga algoritma yaitu Naïve Bayes, KNN dan Support Vector Machine (SVM) pada tahun 2024, dengan tingkat akurasi Naïve Bayes sebesar 87,50%, KNN sebesar 85% dan SVM sebesar 92,50%. Pada tahun yang sama, penelitian di Daerah Istimewa Yogyakarta menggunakan algoritma C4.5, yang menunjukkan akurasi luar biasa yaitu 99,94%. Penelitian lain yang dilakukan pada tahun 2024 tentang klasifikasi kualitas udara di kota-kota besar di Indonesia melalui penggunaan algoritma K-Nearest Neighbor dan Naïve Bayes menghasilkan akurasi KNN sebesar 95,13% dan akurasi Naïve Bayes sebesar 95,97%.

Kendati demikian, kelima penelitian ini memiliki keterbatasan substansial terkait dengan jumlah data yang digunakan, dan data yang digunakan tidak sepenuhnya mewakili keadaan sebenarnya. Hal ini dapat mempengaruhi validitas dan keakuratan hasil klasifikasi. Sehubungan dengan itu, penelitian ini bertujuan untuk mengumpulkan data yang lebih besar, lebih representatif, dan lebih relevan dengan situasi sebenarnya, guna meningkatkan akurasi dan mengatasi kendala tersebut, serta memperbarui data yang tersedia.

Dengan demikian, penulis merasa tertarik untuk melakukan penelitian ini sebagai bagian dari Tugas Akhir dengan tajuk "Klasifikasi Kualitas Udara Jakarta Berdasarkan Indeks Standar Pencemar Udara Menggunakan Algoritma C4.5 dan Naïve Bayes". Diharapkan pihak-pihak yang berkepentingan dapat menerapkan temuan-temuan penelitian untuk mengatasi permasalahan pencemaran udara dalam menjaga kualitas udara.

1.2 Rumusan Masalah

Dalam konteks penelitian, beberapa rumusan masalah diajukan dalam analisis mengenai klasifikasi kualitas udara di DKI Jakarta. Topik-topik yang dibahas dalam analisis ini adalah sebagai berikut:

1. Bagaimana tingkat akurasi yang dicapai ketika mengklasifikasikan kualitas udara di DKI Jakarta menggunakan algoritma C4.5 dan Naïve Bayes?
2. Atribut apa saja yang mempengaruhi hasil klasifikasi kualitas udara di DKI Jakarta dengan menggunakan kedua algoritma tersebut?
3. Bagaimana perbandingan hasil antara algoritma C4.5 dan Naïve Bayes berdasarkan tingkat akurasi dalam mengklasifikasikan kualitas udara di DKI Jakarta?

1.3 Tujuan Penelitian

Tujuan dari penelitian adalah untuk menerapkan algoritma C4.5 dan Naïve Bayes dalam hal-hal berikut:

1. Mengidentifikasi dan mempelajari kualitas udara di wilayah DKI Jakarta berdasarkan data dari Indeks Standar Pencemar Udara (ISPU).
2. Mempelajari perbandingan hasil algoritma C4.5 dan Naïve Bayes untuk menilai klasifikasi tingkat kualitas udara di DKI Jakarta.
3. Memberikan informasi bermanfaat mengenai kualitas udara di DKI Jakarta kepada masyarakat dan pemangku kepentingan berdasarkan hasil klasifikasi.

1.4 Batasan Masalah

Membatasi masalah sangat penting untuk menetapkan batasan topik yang diteliti agar dapat lebih fokus pada masalah yang sedang dibahas:

1. Data penelitian menggunakan sumber dari halaman situs Jakarta Rendah Emisi dan Udara Jakarta periode September 2024.
2. Penelitian berfokus pada data Indeks Standar Pencemar Udara (ISPU) untuk wilayah dki1 (Jalan Bundaran HI, Jalan Cempaka Putih, dan Stasiun Gambir) mewakili Jakarta Pusat, dki2 (Jalan Kelapa Gading, Rusun

Penjaringan, dan Jakarta International Stadium) mewakili Jakarta Utara, dki3 (Jalan Jagakarsa, Jalan Pasar Minggu, dan Jalan Fatmawati) mewakili Jakarta Selatan, dki4 (Jalan Lubang Buaya, GOR Ciracas, dan Taman Delonix) mewakili Jakarta Timur, dan dki5 (Jalan Kebon Jeruk, Kota Tua, dan Terminal Kalideres) mewakili Jakarta Barat.

3. Atribut yang digunakan oleh kedua sumber data meliputi lokasi, waktu, partikel debu (pm10 dan pm2.5), sulfur dioksida (so2), karbon monoksida (co), ozon permukaan (o3), nitrogen dioksida (no2), dan ispu_status.
4. Metode analisis dalam penelitian menerapkan algoritma C4.5 dan Naïve Bayes.
5. Nilai *accuracy*, *precision*, *recall*, dan *f1-score* digunakan untuk menilai keakuratan klasifikasi.

1.5 Kontribusi

Kontribusi dari penelitian ini adalah untuk mengklasifikasikan kualitas udara di wilayah DKI Jakarta dengan menggunakan algoritma C4.5 dan Naïve Bayes. Sebagai langkah untuk mengklasifikasikan kualitas udara secara akurat, penelitian ini dibuat dengan mengevaluasi data ISPU. Dengan demikian, hasil penelitian ini dapat digunakan oleh berbagai pemangku kepentingan seperti masyarakat, pemerintah, dan organisasi lingkungan hidup, untuk memperoleh pemahaman yang lebih cepat dan tepat mengenai situasi kualitas udara. Informasi ini dimaksudkan untuk membantu mengembangkan peraturan lingkungan hidup di wilayah DKI Jakarta.