

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu aspek lingkungan yang berperan penting bagi kelangsungan hidup manusia adalah udara dengan kualitas yang baik [1]. Perkembangan teknologi yang semakin pesat berdampak signifikan pada seluruh aspek kehidupan manusia, termasuk di bidang ekonomi, politik, sosial, budaya, serta pendidikan. Untuk mendukung keberlangsungan kehidupan manusia, lingkungan yang sehat memerlukan kualitas udara yang bersih dan aman untuk dihirup. Kualitas udara telah menjadi isu yang semakin krusial di wilayah perkotaan di seluruh dunia. Polusi udara menjadi perhatian serius karena dapat menyebabkan berbagai penyakit pernapasan dan masalah kesehatan lainnya. Masalah ini terutama dipicu oleh peningkatan aktivitas kendaraan bermotor, industri konstruksi, pembakaran lahan, penumpukan sampah, serta perubahan komposisi kualitas udara secara signifikan.

Kualitas udara yang buruk biasanya menunjukkan tingkat polusi yang tinggi, sementara kualitas udara yang baik menandakan konsentrasi polutan yang rendah. Polutan udara terdiri dari berbagai senyawa kimia, di antaranya partikulat (*particulate matter*), sulfur dioksida (SO₂), karbon monoksida (CO), ozon (O₃), dan nitrogen dioksida (NO₂). Kadar unsur dalam senyawa tersebut dapat berdampak pada kesehatan masyarakat, terutama penyakit pernafasan, dan bahkan dapat menyebabkan kematian [2].

DKI Jakarta, sebagai salah satu kota metropolitan terbesar di dunia, dihadapkan pada tantangan besar dalam upaya mengendalikan polusi udara dari berbagai sumber. Kualitas udara di Jakarta, sebuah perhatian utama pemerintah dan masyarakat selama beberapa dekade terakhir. Jumlah kendaraan bermotor meningkat dan pesatnya pertumbuhan aktivitas industri adalah penyebab tingginya konsentrasi polutan udara di kota. Dalam beberapa penelitian, tingkat pencemaran udara di Jakarta sering kali melebihi ambang batas yang ditetapkan oleh WHO, yang menunjukkan bahwa kualitas udara di Jakarta sering dikelompokkan sebagai tidak sehat. Pengukuran kualitas udara di Jakarta dilakukan secara berkala oleh Dinas Lingkungan Hidup Provinsi DKI Jakarta. Hasil pengukuran ini

didokumentasikan dalam bentuk Indeks Standar Pencemar Udara (ISPU), yang merupakan indikator dari hasil pemantauan kualitas udara, yang merupakan indikator dari hasil pemantauan kualitas udara. ISPU memberikan gambaran mengenai tingkat kebersihan atau pencemaran udara serta dampaknya terhadap kesehatan manusia [3].

Dalam hal mitigasi dampak negatif dari adanya kualitas udara yang buruk, diperlukan agar pemerintah memiliki sistem yang mampu memprediksi dan mengklasifikasikan kualitas udara dengan akurat. Dalam hal ini, data mining memiliki peran penting dalam proses pengelolaan data kualitas udara sehingga dapat diperoleh informasi yang bermakna. *Cross-Industry Standard Process for Data Mining* (CRISP-DM) adalah suatu metodologi yang umum diterapkan dalam proyek data mining. CRISP-DM merupakan pendekatan berbasis model proses pengolahan data yang sering dimanfaatkan oleh para ahli untuk menyelesaikan permasalahan [4]. Metodologi ini memberikan kerangka kerja proses data mining meliputi enam fase utama, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Setiap fase dalam metodologi CRISP-DM dirancang untuk menawarkan pendekatan yang sistematis dan terstruktur dalam proses analisis data, termasuk dalam studi kualitas udara. Penerapan metodologi CRISP-DM dalam analisis kualitas udara memberikan sejumlah dampak positif yang signifikan terhadap proses dan hasil analisis.

Selain metodologi, pemilihan algoritma yang tepat juga menjadi faktor penting dalam keberhasilan klasifikasi kualitas udara. Terdapat beberapa algoritma yang digunakan dalam penelitian terdahulu, salah satunya yaitu penelitian untuk menganalisis klasifikasi data kualitas udara di DKI Jakarta dengan menggunakan algoritma C4.5. Hasil penelitian menunjukkan bahwa penerapan algoritma C4.5 dalam data mining mampu mengklasifikasikan kualitas udara ke dalam tiga kategori utama, yaitu sedang, baik, dan tidak sehat. Dari proses klasifikasi yang dilakukan, metode ini terbukti efektif dalam mengolah data kualitas udara, namun algoritma ini mempunyai keterbatasan sensitivitas algoritma terhadap perubahan data dan kemungkinan *overfitting* pada dataset dengan jumlah atribut yang besar [5].

Penelitian selanjutnya yaitu menganalisis klasifikasi kualitas udara di DKI Jakarta menggunakan Algoritma *Decision Tree* dan *Support Vector Machine* (SVM). Hasil akurasi dalam penelitian ini memperlihatkan bahwa algoritma *Decision Tree* menunjukkan kinerja yang lebih unggul dibandingkan algoritma SVM dalam mengklasifikasikan kualitas udara. Namun, algoritma ini cenderung mengalami *overfitting*, sementara SVM kurang optimal dalam menangani data dengan margin kelas yang berdekatan [6].

Tahun 2023, beberapa peneliti juga menggunakan algoritma *Random Forest* untuk mengklasifikasikan kualitas udara di DKI Jakarta. Hasil penelitian menunjukkan bahwa variasi jumlah pohon keputusan serta perbandingan rasio antara data *training* dan data *testing* berpengaruh terhadap performa klasifikasi. Meskipun *Random Forest* mampu memberikan hasil yang baik, algoritma ini masih memiliki keterbatasan, seperti kompleksitas komputasi yang tinggi dan sensitivitas terhadap jumlah parameter yang digunakan [1].

Dari beberapa penelitian terdahulu yang telah diuraikan, terlihat bahwa meskipun berbagai algoritma telah diterapkan, masih terdapat kelemahan yang mempengaruhi akurasi dan generalisasi model. Oleh karena itu, penelitian ini akan menggunakan algoritma *Gradient boosting*, yang dikenal mampu mengatasi *overfitting* lebih baik dibandingkan *Decision Tree*, memiliki performa lebih stabil dibandingkan *Random Forest*, serta dapat meningkatkan akurasi dengan memperbaiki kesalahan prediksi secara bertahap melalui pendekatan *ensemble learning*. Algoritma ini merupakan salah satu algoritma pembelajaran mesin yang sering digunakan untuk masalah klasifikasi karena kemampuannya dalam mengolah data dengan jumlah besar dan variabel yang kompleks. Algoritma *Gradient boosting* adalah metode untuk tugas regresi dan klasifikasi yang membangun model prediktif dengan menggabungkan sejumlah model lemah secara bertahap [7]. Algoritma *Gradient Boosting* ini juga mampu meminimalkan kesalahan prediksi dengan membangun model secara bertahap dan mengoptimalkan kinerjanya. Penelitian ini mengimplementasikan metodologi CRISP-DM untuk klasifikasi kualitas udara di Kota Jakarta dengan menggunakan algoritma *Gradient boosting*. Dengan menggunakan algoritma ini, hasil prediksi kualitas udara dapat diperoleh dengan tingkat akurasi yang lebih tinggi, sehingga

dapat membantu Dinas Lingkungan Hidup di wilayah DKI Jakarta untuk mengambil keputusan dalam mengembangkan kebijakan yang efektif untuk mengurangi dampak polusi udara dan petunjuk untuk mencegah masalah kualitas udara kepada orang-orang di sekitar wilayah DKI Jakarta.

1.2 Perumusan Masalah

Dalam penelitian ini, perumusan masalah menjadi tiga bagian, yaitu identifikasi permasalahan, metode usulan, dan penyusunan pertanyaan penelitian. Identifikasi permasalahan bertujuan untuk lebih memahami masalah utama yang menjadi fokus analisis. Pengusulan metode dilakukan dengan memilih pendekatan atau teknik yang terbaru untuk penyelesaian masalah yang telah diidentifikasi. Penyusunan pertanyaan penelitian dilakukan untuk memastikan bahwa tujuan penelitian dapat dicapai secara sistematis.

1.2.1 Permasalahan

Permasalahan kualitas udara di Jakarta menjadi isu lingkungan yang krusial dan berdampak besar terhadap kesehatan masyarakat maupun tingkat kualitas hidup. Penelitian-penelitian sebelumnya telah banyak menganalisis masalah ini dengan menggunakan algoritma seperti *Decision Tree*, *Support Vector Machine* (SVM), dan *Random Forest* untuk melakukan klasifikasi kualitas udara di Jakarta. Namun, pendekatan-pendekatan tersebut masih memiliki beberapa kelemahan, seperti kecenderungan *overfitting*, ketidak-optimalan dalam menangani margin kelas yang berdekatan, serta kompleksitas komputasi yang tinggi. Ketidakseimbangan data (*imbalance class*) pada dataset kualitas udara sering kali tidak diperhatikan, di mana kategori polusi tertentu memiliki sampel lebih sedikit, menyebabkan model kurang akurat dalam memprediksi tingkat polusi udara yang justru paling berisiko bagi kesehatan masyarakat.

1.2.2 Metode Usulan

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan algoritma *Gradient boosting*. Algoritma ini dipilih karena kemampuannya yang lebih stabil dalam menangani *overfitting* serta potensinya dalam meningkatkan akurasi klasifikasi melalui pendekatan *ensemble learning*.

Penelitian ini dilakukan dalam kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) untuk mengelola proses analisis data.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan yang telah diidentifikasi, beberapa pertanyaan penelitian yang dapat dirumuskan adalah sebagai berikut:

1. Fitur-fitur apa saja yang paling berpengaruh terhadap pencemaran udara di Jakarta berdasarkan data yang tersedia?
2. Bagaimana penerapan metode CRISP-DM dapat mendukung analisis kualitas udara di Jakarta?
3. Seberapa tinggi tingkat akurasi model prediksi kualitas udara yang dibangun menggunakan metode *Gradient boosting*?
4. Rekomendasi kebijakan apa yang dapat disusun untuk meningkatkan kualitas udara di Jakarta berdasarkan hasil analisis ini?

1.3 Tujuan dan Manfaat

Dalam penelitian ini, tujuan dan manfaat memiliki peran yang penting, karena dapat memberikan arah yang jelas dan nilai bagi penelitian, serta menjelaskan signifikansi penelitian dalam memberikan solusi yang berpotensi dapat dampak positif.

1.3.1 Tujuan Penelitian

Penelitian ini memiliki 4 tujuan dalam menganalisis kualitas udara di Kota Jakarta:

1. Mengidentifikasi fitur-fitur yang paling signifikan terhadap pencemaran udara di Jakarta berdasarkan data yang tersedia menggunakan analisis berbasis algoritma *Gradient boosting*.
2. Mengevaluasi efektivitas penerapan metode CRISP-DM dalam mendukung proses analisis data kualitas udara, mulai dari pemahaman data hingga pembuatan model prediktif.
3. Mengukur tingkat akurasi model prediksi kualitas udara yang dibangun dengan algoritma *Gradient boosting* untuk memastikan keandalan hasil prediksi.

4. Menyusun rekomendasi kebijakan berbasis data untuk meningkatkan kualitas udara di Jakarta berdasarkan temuan analisis fitur penting dan hasil prediksi model.

1.3.2 Manfaat Penelitian

Hasil analisis dan model yang dikembangkan, dapat digunakan oleh pemerintah sebagai dasar dalam pengambilan keputusan untuk kebijakan pengelolaan kualitas udara. Penelitian juga bisa dijadikan referensi bagi penelitian-penelitian selanjutnya yang berhubungan dengan analisis kualitas udara dan menerapkan algoritma atau metode data mining yang lainnya.

1.4 Batasan Masalah

Pada penelitian ini, peneliti membuat batasan masalah untuk mencegah kesalahan selama penelitian. Batasan masalah tersebut adalah sebagai berikut:

1. Objek penelitian dalam penelitian ini adalah kualitas udara di Kota DKI Jakarta.
2. Data yang digunakan dalam penelitian ini berfokus pada dataset ISPU DKI Jakarta pada periode tahun 2010-2021. Data yang diambil sebanyak 21.915 data.
3. Dataset yang digunakan terdiri dari 5 fitur dan 3 kelas. Fitur-fitur yang digunakan dalam penelitian ini terdapat dalam dataset antara lain PM10, SO₂, CO, O₃, NO₂. Kelas kategori terdiri dari 3 kelas yaitu sedang, baik, tidak sehat.
4. Hasil rekomendasi untuk masyarakat dan pemerintah sebagai kebijakan dalam mengatasi masalah kualitas udara.

1.5 Sistematika Penulisan

Adapun sistematika penulisan dalam proposal penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Menjelaskan latar belakang masalah, perumusan masalah, serta sistematika penulisan

BAB II TINJAUAN PUSTAKA

Membahas teori-teori yang mendasari penelitian, termasuk metode CRISP-DM, algoritma *Gradient boosting*, serta penelitian terkait tentang analisis kualitas udara.

BAB III METODOLOGI PENELITIAN

Menjelaskan metode pengumpulan data, teknik analisis data dengan pendekatan CRISP-DM, teknik klasifikasi data dengan algoritma *Gradient boosting*, serta tahapan pelaksanaan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Menjelaskan mengenai hasil serta pembahasan dari penelitian yang telah dilaksanakan.

BAB V PENUTUP

Menjelaskan mengenai hasil kesimpulan dan saran dari penelitian

REFERENSI

Berisi daftar referensi dari penelitian yang telah dilakukan