

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Kualitas udara menjadi salah satu isu lingkungan yang semakin mendapat perhatian dalam beberapa tahun terakhir, khususnya di kawasan perkotaan dengan mobilitas dan aktivitas ekonomi yang tinggi. Peningkatan jumlah kendaraan, pertumbuhan industri, serta kepadatan penduduk menyebabkan konsentrasi polutan di udara terus meningkat. DKI Jakarta sebagai pusat pemerintahan dan aktivitas nasional termasuk wilayah yang menghadapi tekanan pencemaran udara yang cukup serius (World Air Quality Report 2024).

Data menunjukkan bahwa rata-rata konsentrasi partikulat halus PM_{2.5} di Jakarta mencapai 57 $\mu\text{g}/\text{m}^3$. Angka tersebut jauh melampaui batas aman tahunan yang direkomendasikan sebesar 15 $\mu\text{g}/\text{m}^3$. Kondisi ini mengindikasikan bahwa kualitas udara Jakarta belum berada pada tingkat yang aman dan memerlukan upaya pengendalian yang lebih konsisten

Dampak pencemaran udara tidak hanya terlihat pada aspek lingkungan, tetapi juga pada kesehatan masyarakat. Paparan jangka panjang terhadap polutan seperti PM_{2.5}, PM₁₀, NO₂, SO₂, CO, dan O₃ dapat meningkatkan risiko gangguan pernapasan serta penyakit kardiovaskular. Di kota dengan tingkat kepadatan tinggi seperti Jakarta, paparan tersebut terjadi secara berulang, sehingga potensi risikonya menjadi semakin besar (Health Effects Institute (2023).

Di Indonesia, pemantauan kualitas udara dilakukan melalui Indeks Standar Pencemar Udara (ISPU). Indeks ini dihitung berdasarkan 6 parameter polutan yang kemudian diklasifikasikan ke dalam beberapa kategori kualitas udara. Meskipun sistem ini telah diterapkan secara nasional, data ISPU memiliki karakteristik yang cukup kompleks, termasuk hubungan yang tidak linear antarparameter serta distribusi kategori yang tidak selalu seimbang (Kementerian Lingkungan Hidup Dan Kehutanan (2022).

Selain itu, ada keterbatasan jumlah stasiun pemantauan dan variasi kondisi lingkungan antarwilayah turut memengaruhi konsistensi data. Pola

pencemaran udara juga cenderung fluktuatif, mengikuti dinamika aktivitas manusia dan kondisi meteorologis. Hal ini menyebabkan pendekatan statistik konvensional sering kali kurang memadai untuk menangkap pola secara menyeluruh (Sun et al., 2023).

Perkembangan metode *machine learning* ini yang akan membuka peluang untuk menganalisis data kualitas udara dengan pendekatan yang lebih adaptif. Teknik berbasis *ensemble learning*, khususnya pada *gradient boosting*, dinilai mampu menangani data berdimensi tinggi serta hubungan nonlinier secara lebih efektif. Light Gradient Boosting Machine (LightGBM) dan Categorical Boosting (CatBoost) merupakan dua algoritma yang sering digunakan dalam analisis data lingkungan dan menunjukkan hasil yang kompetitif (Mahmood et al., 2024)

Meskipun demikian, beberapa penelitian menunjukkan bahwa kinerja kedua algoritma tersebut dapat berbeda tergantung pada karakteristik dataset dan konteks wilayah penelitian. Oleh karena itu, diperlukan evaluasi komparatif dalam satu skema eksperimen yang sama untuk mengetahui algoritma mana yang paling sesuai dalam klasifikasi kualitas udara berbasis ISPU di DKI Jakarta. Atas dasar tersebut, penelitian ini difokuskan pada analisis perbandingan kinerja LightGBM dan CatBoost untuk menentukan model klasifikasi yang paling optimal.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini dirumuskan dalam beberapa pertanyaan sebagai berikut:

1. Bagaimana karakteristik dan distribusi data kualitas udara berbasis ISPU di wilayah DKI Jakarta berdasarkan parameter polutan utama?
2. Bagaimana perbandingan kinerja algoritma LightGBM dan CatBoost dalam mengklasifikasikan kualitas udara berbasis data ISPU?
3. Algoritma mana yang menunjukkan performa paling optimal berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score?

1.3 Tujuan Penelitian

Tujuan pada penelitian saya ini adalah:

1. Mengidentifikasi dan menganalisis karakteristik serta distribusi data pada kualitas udara berbasis ISPU di wilayah DKI Jakarta.
2. Mengevaluasi dan membandingkan performa pada algoritma LightGBM dan CatBoost dalam proses klasifikasi kualitas udara.
3. Menentukan algoritma yang paling optimal berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

1.4 Batasan Masalah

Agar penelitian tetap terarah dan tidak melebar, ruang lingkup penelitian ini dibatasi pada hal-hal berikut:

1. Penelitian difokuskan pada wilayah DKI Jakarta menggunakan data Indeks Standar Pencemar Udara (ISPU).
2. Data yang digunakan berupa data numerik periode 2020–2025.
3. Parameter polutan yang dianalisis meliputi PM_{2.5}, PM₁₀, NO₂, SO₂, CO, dan O₃.
4. Algoritma yang digunakan terbatas pada LightGBM dan CatBoost tanpa melibatkan model pembanding lainnya.
5. Evaluasi performa model menggunakan metrik akurasi, presisi, recall, dan F1-score dengan skema *train–test split*.
6. Penelitian tidak mencakup analisis data non-numerik atau pendekatan selain klasifikasi berbasis ISPU.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi akademik dan praktis melalui analisis klasifikasi kualitas udara berbasis ISPU menggunakan algoritma *LightGBM* dan *CatBoost*.

1. Akademis

- a. Menambah referensi ilmiah terkait penerapan algoritma LightGBM dan CatBoost dalam klasifikasi kualitas udara berbasis ISPU.
- b. Mendukung pengembangan kajian *ensemble learning* dalam bidang lingkungan dan analisis data.
- c. Menjadi dasar pertimbangan bagi penelitian selanjutnya yang mengkaji pemodelan kualitas udara berbasis data sensor.

2. Praktis

- a. Memberikan gambaran kondisi kualitas udara di DKI Jakarta berdasarkan hasil klasifikasi data ISPU.
- b. Menyajikan perbandingan performa model yang dapat digunakan sebagai referensi dalam pengembangan sistem pemantauan kualitas udara.
- c. Mendukung proses pengambilan keputusan berbasis data dalam pengelolaan kualitas udara perkotaan.

3. Kebijakan

- a. Menjadi bahan pertimbangan bagi instansi pemerintah dan lembaga lingkungan dalam mengevaluasi kondisi kualitas udara di wilayah DKI Jakarta.
- b. Memberikan dukungan berbasis data bagi perumusan kebijakan pengendalian pencemaran udara dan peningkatan kualitas lingkungan.

1.6 Kontribusi.

Penelitian ini memberikan kontribusi pada aspek akademik, metodologis, dan praktis melalui penerapan serta evaluasi komparatif algoritma LightGBM dan CatBoost dalam klasifikasi kualitas udara berbasis data ISPU.

1. Kontribusi Akademik

- a. Menghadirkan analisis perbandingan kinerja LightGBM dan CatBoost dalam konteks klasifikasi ISPU di DKI Jakarta.
- b. Menambah referensi penelitian di bidang *machine learning* dan lingkungan, khususnya terkait evaluasi model klasifikasi kualitas udara di Indonesia.

2. Kontribusi Metodologis

- a. Menyusun alur kerja klasifikasi kualitas udara berbasis ISPU yang terstruktur, mulai dari tahap prapemrosesan hingga evaluasi model.
- b. Menawarkan pendekatan evaluasi yang dapat diterapkan kembali pada wilayah dengan karakteristik data yang serupa.

3. Kontribusi Praktis

- a. Menghasilkan analisis yang dapat dimanfaatkan dalam pengembangan sistem informasi pemantauan kualitas udara berbasis data.
- b. Memberikan gambaran kondisi kualitas udara yang lebih terukur sehingga dapat membantu pemangku kepentingan dalam memahami dinamika pencemaran udara.

