

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kualitas udara merupakan salah satu permasalahan lingkungan yang berdampak langsung terhadap kesehatan masyarakat, khususnya di wilayah perkotaan dengan tingkat aktivitas transportasi dan industri yang tinggi [1]. Kualitas udara ditentukan oleh konsentrasi berbagai polutan seperti PM_{2.5}, PM₁₀, O₃, NO₂, dan SO₂. Semakin tinggi konsentrasi polutan tersebut di udara ambien, semakin besar potensi gangguan terhadap kesehatan masyarakat, mulai dari iritasi saluran pernapasan hingga penyakit kardiovaskular dan respiratori kronis [2], [3], [4].

Kota Medan sebagai salah satu kota metropolitan di Indonesia menghadapi tantangan serius terkait kualitas udara. Berdasarkan data dari platform pemantauan kualitas udara, konsentrasi polutan di Kota Medan seringkali menunjukkan fluktuasi yang signifikan, dipengaruhi oleh berbagai faktor seperti kepadatan lalu lintas, aktivitas industri, dan kondisi meteorologis [5]. Namun demikian, informasi kualitas udara yang tersedia bagi masyarakat saat ini hanya terbatas pada data historis dan real-time melalui platform seperti OpenAQ atau ISPU KLHK [6]. Ketiadaan model peringatan dini (early warning model) yang dapat memprediksi kualitas udara di masa depan menyebabkan masyarakat dan pemangku kebijakan hanya dapat bertindak secara reaktif setelah polusi terjadi, tanpa memiliki kesempatan untuk melakukan mitigasi proaktif [7]. Padahal, kemampuan memprediksi Indeks Kualitas Udara (AQI) untuk beberapa hari ke depan sangat krusial agar kelompok rentan (lansia, anak-anak, penderita asma) dapat mempersiapkan diri dan masyarakat dapat merencanakan aktivitas harian dengan lebih baik [4], [8].

Tantangan utama dalam membangun model prediksi kualitas udara terletak pada karakteristik data yang kompleks. Data kualitas udara Kota Medan menunjukkan dua jenis pola yang saling tumpang tindih. Pertama, terdapat hubungan linear antar polutan, misalnya antara PM_{2.5} dan PM₁₀ yang cenderung meningkat bersama-sama, serta hubungan linear antara konsentrasi polutan dengan nilai AQI. Kedua, data juga mengandung pola non-linear yang kompleks, seperti efek sinergis antar polutan dalam kondisi meteorologis tertentu (misalnya, suhu tinggi dan kelembapan rendah yang memicu pembentukan O₃ troposferik), fluktuasi ekstrem pada hari-hari dengan kepadatan lalu lintas tinggi, serta pola musiman yang tidak sepenuhnya mengikuti siklus tetap [5], [9], [10]. Dengan kata lain, data AQI Kota

Medan tidak dapat dimodelkan secara memadai oleh pendekatan linear saja, tetapi juga tidak sepenuhnya non-linear sehingga pendekatan non-linear murni pun berisiko menghasilkan variansi yang tinggi.

Berdasarkan kajian literatur, Multiple Linear Regression (MLR) merupakan metode statistik parametrik yang efektif sebagai model baseline untuk menangkap hubungan linear antar variabel prediktor dengan target AQI [9]. MLR memiliki keunggulan dalam hal interpretabilitas, di mana koefisien regresi memberikan informasi eksplisit mengenai arah dan besaran pengaruh setiap variabel [11]. Namun, MLR memiliki keterbatasan struktural karena gagal menangkap hubungan non-linear dan sensitif terhadap data pencilan (*outlier*) [12]. Sebaliknya, *Random Forest Regressor* (RFR) merupakan algoritma *machine learning* berbasis *ensemble learning* yang mampu menangani hubungan non-linear, robust terhadap *outlier*, dan tidak memerlukan normalisasi data [10]. RFR bekerja dengan membangun banyak pohon keputusan dan menggabungkan prediksinya melalui mekanisme *averaging*, sehingga menghasilkan prediksi yang lebih stabil [13]. Namun, RFR memiliki kelemahan berupa kecenderungan *overfitting* pada data dengan noise tinggi serta interpretabilitas yang rendah [14].

Pendekatan meta-ensemble *stacking* menawarkan solusi untuk menggabungkan kelebihan dan mengoreksi kelemahan masing-masing model. *Stacking* (*stacked generalization*) adalah teknik ensemble lanjutan yang menggabungkan prediksi dari beberapa model dasar (*base models*) menggunakan model pembelajaran tingkat tinggi yang disebut *meta-model* [14]. Berbeda dengan metode ensemble konvensional seperti *bagging* yang menggabungkan prediksi dengan mekanisme statis (rata-rata), *stacking* mempelajari cara terbaik untuk mengkombinasikan prediksi dari berbagai *base model* secara adaptif berdasarkan data [15], [16]. Penelitian oleh Tian dkk. (2024) dan Morshed-Bozorgdel dkk. (2022) menunjukkan bahwa pendekatan meta-ensemble mampu meningkatkan akurasi prediksi secara signifikan dibandingkan model tunggal [15], [16].

Berdasarkan uraian tersebut, penelitian ini mengusulkan pengembangan model prediksi meta-ensemble *stacking* dengan arsitektur dua tingkat. Pada Level-0, *Multiple Linear Regression* (MLR) digunakan sebagai *base model* untuk menangkap hubungan linear utama. Pada Level-1, prediksi dari MLR digunakan sebagai fitur input untuk melatih *Random Forest Regressor* (RFR) sebagai *meta-model*, yang bertugas mengoreksi residu (*error*) non-linear yang tidak dapat dijelaskan oleh MLR. Kombinasi ini diharapkan mampu menghasilkan model prediksi yang lebih akurat dan stabil dibandingkan model

tunggal, karena mampu menangkap karakteristik data AQI Kota Medan yang mengandung pola linear sekaligus non-linear.

Sebagai bentuk implementasi dan pembuktian konsep (*proof of concept*) dari model yang dikembangkan, penelitian ini juga mengintegrasikan model prediksi ke dalam sebuah prototipe *website*. *Website* berfungsi sebagai media visualisasi hasil prediksi yang memudahkan masyarakat dan pemangku kebijakan dalam mengakses dan menginterpretasikan informasi kualitas udara. Pemilihan *website* sebagai media implementasi didasarkan pada aksesibilitasnya yang tinggi melalui berbagai perangkat tanpa instalasi aplikasi khusus, serta kemampuannya menyajikan informasi dengan visualisasi kode warna yang intuitif sesuai standar US EPA/KLHK [9], [10], [11]. Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan model prediksi yang akurat, tetapi juga pada penyediaan sistem informasi yang dapat menjembatani hasil penelitian dengan kebutuhan masyarakat.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, terdapat dua permasalahan utama yang perlu diselesaikan dalam penelitian ini:

1. Bagaimana mengembangkan model prediksi meta-ensemble *stacking* yang menggabungkan *Multiple Linear Regression* (MLR) sebagai *base model* dan *Random Forest Regressor* (RFR) sebagai *meta-model* untuk memprediksi Indeks Kualitas Udara (AQI) di Kota Medan dengan rentang waktu 1 hingga 7 hari ke depan melalui pendekatan prediksi rekursif (*recursive multi-step forecasting*), serta mengevaluasi apakah model tersebut mampu menghasilkan prediksi yang lebih akurat dibandingkan model tunggal (MLR dan RFR)?
2. Bagaimana mengintegrasikan model prediksi meta-ensemble *stacking* terbaik ke dalam sebuah prototipe *website* yang berfungsi sebagai media visualisasi hasil prediksi, sehingga masyarakat dan pemangku kebijakan di Kota Medan dapat mengakses informasi prediksi AQI secara mudah, informatif, dan dilengkapi dengan kategori kualitas udara, kode warna sesuai standar, serta rekomendasi kesehatan?

1.3 Tujuan

1. Mengembangkan dan mengevaluasi model prediksi meta-ensemble *stacking* dua tingkat yang menggabungkan *Multiple Linear Regression* (MLR) sebagai *base model* dan *Random Forest Regressor* (RFR) sebagai *meta-model* untuk memprediksi

nilai Indeks Kualitas Udara (AQI) di Kota Medan dengan rentang waktu 1 hingga 7 hari ke depan melalui pendekatan prediksi rekursif (*recursive multi-step forecasting*), serta membandingkan performanya dengan model tunggal MLR dan RFR menggunakan metrik evaluasi MAE, RMSE, dan R^2 .

2. Mengimplementasikan model prediksi meta-ensemble *stacking* terbaik ke dalam sebuah prototipe website sebagai media visualisasi hasil prediksi, yang menyajikan informasi AQI dalam bentuk nilai numerik, kategori kualitas udara dengan kode warna sesuai standar US EPA/KLHK, status kesehatan, serta rekomendasi kesehatan bagi masyarakat dan pemangku kebijakan di Kota Medan.

1.4 Manfaat

Berdasarkan tujuan penelitian yang telah ditetapkan, maka manfaat yang diharapkan dari penelitian ini adalah:

1. Penelitian ini menghasilkan model prediksi meta-ensemble *stacking* yang menggabungkan *Multiple Linear Regression* (MLR) dan *Random Forest Regressor* (RFR) untuk prediksi Indeks Kualitas Udara (AQI) di Kota Medan. Model ini diharapkan dapat menjadi alternatif pendekatan yang lebih akurat dibandingkan model tunggal dalam menangani data kualitas udara yang mengandung pola linear dan non-linear secara simultan.
2. Prototipe website yang dikembangkan berfungsi sebagai bukti konsep (*proof of concept*) bahwa model prediksi meta-ensemble dapat diimplementasikan dalam sistem informasi yang mudah diakses. Website ini dapat menjadi landasan awal bagi pengembangan lebih lanjut menuju sistem peringatan dini kualitas udara yang dapat diakses publik.
3. Penelitian ini mendokumentasikan secara sistematis alur pengembangan model prediksi berbasis *stacking ensemble*, mulai dari pengumpulan data, pre-processing, pelatihan model, evaluasi, hingga integrasi ke dalam sistem informasi. Dokumentasi ini dapat menjadi panduan praktis bagi peneliti lain yang ingin mengembangkan model serupa.

1.5 Ruang Lingkup

Pada penelitian yang dilakukan dalam pembuatan ruang lingkup difokuskan terhadap beberapa hal, diantaranya:

1. Data kualitas udara Kota Medan periode 1 Januari 2023 hingga 31 Desember 2025 yang diperoleh dari platform *OpenAQ* dan BMKG, dengan pembagian data latih (*training*) untuk periode 1 Januari 2023–31 Desember 2024 dan data uji (*testing*) untuk periode 1 Januari 2025–31 Desember 2025.
2. Variabel independen terdiri atas lima polutan (PM_{2.5}, PM₁₀, O₃, NO₂, SO₂) dan tiga variabel meteorologis (suhu, kecepatan angin, kelembapan).
3. Target prediksi berupa nilai numerik AQI untuk 7 hari ke depan ($t+1$ hingga $t+7$) berdasarkan data historis hari sebelumnya (t).
4. Evaluasi performa model menggunakan MAE (*Mean Absolute Error*), RMSE (*Root Mean Square Error*), dan koefisien determinasi (R^2).
5. Implementasi sistem dibangun dalam bentuk *website* yang terdiri dari:
 - a. *Front-end*: HTML, CSS, dan *JavaScript* untuk antarmuka pengguna (halaman utama prediksi dan halaman data historis).
 - b. *Back-end*: *Python* dengan *framework Django* untuk logika *server*, eksekusi model prediksi, dan penyediaan API.
 - c. Model: Disimpan dalam format *.pkl* (*Pickle*) untuk MLR, RFR, dan *stacking*.
6. *Website* dijalankan secara lokal (*localhost*) untuk tujuan pembuktian konsep (*proof of concept*) dan demonstrasi fungsionalitas sistem, sehingga aksesnya terbatas pada lingkungan pengembangan. Oleh karena itu, manfaat yang diberikan bersifat ilustratif sebagai landasan awal bagi pengembangan lebih lanjut menuju sistem yang dapat di-deploy secara publik, bukan sebagai sistem produksi yang siap diakses oleh masyarakat luas.