

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kualitas udara menjadi salah satu indikator penting dalam kesehatan lingkungan dan masyarakat. Polusi udara, khususnya dari polutan berbahaya, yaitu PM2.5, PM10, NO2, SO2, CO, dan senyawa organik volatil (VOC), telah terbukti memiliki dampak negatif pada kesehatan manusia. Paparan jangka panjang terhadap polutan udara dapat menyebabkan berbagai masalah kesehatan, yaitu gangguan pernapasan, penyakit jantung, stroke, dan kanker paru-paru (Anandari dkk., 2024; Umah dan Gusmira, 2024)

Di perkotaan dengan kepadatan penduduk tinggi dan aktivitas industri yang besar, polusi udara telah mencapai tingkat yang mengkhawatirkan, seperti yang terjadi di Pekanbaru, di mana konsentrasi PM2.5 sering melebihi batas aman yang ditetapkan WHO (Umah dan Gusmira dkk., 2024). Studi menunjukkan bahwa polusi udara dapat meningkatkan risiko depresi dan kecemasan akibat inflamasi di otak, yang berdampak pada kesehatan mental dan produktivitas masyarakat (Anandari dkk., 2024).

Pengukuran kualitas udara secara *real-time* sangat penting untuk mengetahui tingkat polusi di suatu area. Namun, pengukuran ini tidak cukup jika tidak disertai kemampuan untuk memprediksi kondisi udara di masa depan. Misalnya, jika ada potensi lonjakan polusi, tindakan pencegahan, yaitu pengurangan aktivitas luar ruangan dapat diambil lebih awal (Anandari dkk., 2024).

Teknologi Internet of Things (IoT) telah memberikan solusi dalam memantau kualitas udara secara *real-time* dengan penggunaan sensor yang terhubung ke jaringan. Sensor-sensor ini dapat mendeteksi berbagai polutan dan faktor lingkungan, yaitu suhu dan kelembapan. Namun, data yang dihasilkan sensor perlu dianalisis untuk mendapatkan informasi yang bermanfaat, terutama untuk prediksi kondisi udara di masa depan (Waworundeng, 2024).

Machine Learning, khususnya algoritma *Random Forest*, menjadi relevan dalam prediksi kualitas udara. Algoritma ini dapat dilatih menggunakan data historis untuk mengenali pola dan membuat estimasi akurat mengenai perubahan

kualitas udara. Dibandingkan dengan algoritma lain, yaitu *Decision Tree Regression*, *Gradient Boosting Regression*, dan *Multi-Layer Perceptron (MLP)*, *Random Forest* menunjukkan kinerja terbaik dalam berbagai aspek, yaitu akurasi prediksi dan efisiensi pemrosesan (Ameer et al., 2019). *Decision Tree Regression* memiliki keunggulan dalam kecepatan komputasi, tetapi cenderung *overfitting* dan kurang akurat dalam menangani data kompleks. Sementara itu, *Gradient Boosting Regression* sering kali membutuhkan *tuning* parameter yang lebih kompleks dan memiliki waktu pemrosesan yang lebih lama. MLP mampu memberikan hasil yang baik, tetapi memerlukan jumlah data pelatihan yang besar dan memiliki risiko *overfitting* jika tidak diatur dengan baik.

Selain itu, model *Linear Regression* juga digunakan dalam penelitian terkait prediksi kualitas udara karena kesederhanaan dan interpretabilitasnya. Model ini bekerja dengan mengasumsikan hubungan linear antara variabel polutan dan indeks kualitas udara (AQI), serta memiliki komputasi yang lebih cepat dibandingkan model non-linear (Kamil, 2024). Namun, *Linear Regression* memiliki keterbatasan dalam menangkap pola yang lebih kompleks, terutama karena polusi udara dipengaruhi oleh banyak faktor non-linear, yaitu kondisi cuaca, kepadatan lalu lintas, dan aktivitas industri. Selain itu, model ini rentan terhadap *outlier* dan kurang mampu menangani interaksi antarvariabel, yang dapat menyebabkan akurasi prediksi lebih rendah jika dibandingkan dengan metode berbasis *ensemble learning*, yaitu *Random Forest* (Kamil, 2024).

Random Forest dipilih karena mampu menangani data dalam jumlah besar, mengurangi *overfitting* melalui *ensemble learning*, dan memberikan hasil yang lebih stabil dengan kesalahan prediksi (RMSE dan MAE) yang lebih rendah dibandingkan algoritma lainnya. Dengan demikian, penggunaan *Random Forest* dalam sistem prediksi kualitas udara berbasis IoT dapat meningkatkan ketepatan estimasi dan mendukung pengambilan keputusan berbasis data secara lebih efektif.

Dalam sistem prediksi kualitas udara berbasis IoT, *Message Queue Telemetry Transport (MQTT)* digunakan sebagai protokol komunikasi utama untuk mengirim data sensor ke sistem pemrosesan. MQTT adalah protokol ringan berbasis *publish-subscribe* yang dirancang untuk perangkat IoT dengan keterbatasan daya dan bandwidth rendah. Keunggulan utama MQTT adalah efisiensi dalam konsumsi

daya dan keandalan dalam jaringan yang tidak stabil, menjadikannya pilihan ideal untuk pengiriman data sensor secara real-time. Namun, meskipun MQTT menawarkan efisiensi komunikasi, protokol ini memiliki beberapa keterbatasan keamanan, yaitu kurangnya enkripsi bawaan, rentan terhadap serangan *Man-in-the-Middle* (MITM) dan *Denial of Service* (DoS) (Dinculeană & Cheng, 2019). Oleh karena itu, implementasi mekanisme keamanan tambahan, yaitu *Transport Layer Security* (TLS) sangat disarankan untuk memastikan keamanan data yang dikirimkan dalam sistem IoT.

Oleh karena itu, sistem prediksi kualitas udara berbasis IoT dan Random Forest sangat penting dalam memberikan solusi atas masalah ini. Sistem ini menggabungkan data *real-time* dari sensor yang tersebar di berbagai lokasi dan memprosesnya dengan algoritma prediksi berbasis *Random Forest* untuk memberikan estimasi kualitas udara di masa depan. Informasi ini dapat digunakan untuk peringatan dini.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka dapat dirumuskan masalah berikut ini:

- 1) Bagaimana cara memantau kualitas udara untuk menangani pencemaran udara?
- 2) Bagaimana cara mengintegrasikan sensor IoT dengan ESP32 untuk memantau kualitas udara secara *real-time*?
- 3) Bagaimana menyimpan data dari sensor ke dalam *database* dan menampilkan data kualitas udara pada *website*?
- 4) Bagaimana algoritma *Random Forest* dapat diterapkan untuk memprediksi kualitas udara di masa depan berdasarkan data dari sensor IoT?

1.3 Batasan Masalah

Agar ruang lingkup penelitian tidak terlalu luas, batasan masalah yang diterapkan adalah sebagai berikut:

- 1) Sistem hanya berfokus pada polutan PM2.5, PM10, NO2, suhu, dan kelembapan sebagai variabel utama.

- 2) Data yang digunakan untuk pelatihan model prediksi adalah data dari kota Jakarta yang diambil dari Kaggle.
- 3) Pengujian sistem dilakukan pada area terbatas, yaitu lingkungan perumahan dan lingkungan Politeknik Caltex Riau yang telah dilengkapi sensor SDS011, MQ-135, dan DHT22.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Tujuan dari penelitian ini adalah merancang dan mengimplementasikan sistem prediksi kualitas udara berbasis IoT dengan menggunakan algoritma *Random Forest*. Sistem ini akan memanfaatkan data *real-time* dari sensor untuk melakukan prediksi kualitas udara di wilayah Pekanbaru dengan akurasi yang tinggi.

1.4.2 Manfaat

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- 1) Memperluas cakupan pemantauan kualitas udara dengan biaya yang lebih rendah dibandingkan stasiun pemantau.
- 2) Memberikan informasi dan bermanfaat dalam pengendalian polusi udara.