# **BAB I**

# **PENDAHULUAN**

# **Latar Belakang**

Penurunan kualitas udara di Indonesia disebabkan karena adanya polusi udara oleh pencemaran emisi dari kendaraan bermotor, industri/pabrik, pembangkit listrik, rumah tangga, pembakaran lahan, dan sebagainya. Tingkat kandungan zat-zat polutan yang tinggi dapat mempengaruhi kesehatan manusia dan lingkungan, di mana akan berdampak sangat buruk pada seluruh aspek kehidupan di muka bumi (Amalia et al., 2022). Berdasarkan data IQAir dan AQI.in, pada bulan Desember 2023, Ibukota Negara Indonesia, DKI Jakarta menduduki peringkat ke-2 kota paling berpolusi di Indonesia dengan konsentrasi partikel polutan PM2.5 sebesar 43 mikrogram per meter kubik dan PM10 sebesar 89 mikrogram per meter kubik. Konsentrasi partikel ini 2.9 kali lebih tinggi dari ambang batas yang ditetapkan oleh WHO. PM2.5 merupakan salah satu dari tujuh zat polutan yang menjadi parameter pengukuran Indeks Standar Pencemar Udara (ISPU) oleh Kementerian Lingkungan Hidup dan Kehutanan (KLHK), yaitu PM10, PM2.5, NO2, SO2, CO, O3, dan HC yang berdampak akut dan kronis terhadap kesehatan manusia (Chaniago et al., 2020). Adanya prediksi konsentrasi polutan dalam kualitas udara dapat memberikan peringatan dini yang berguna bagi masyarakat untuk mengambil tindakan yang tepat. Namun, konsentrasi polutan ini sangat dipengaruhi oleh data meteorologi seperti kelembapan, suhu, kecepatan angin, arah angin, dan curah hujan (Hidayat & Anov, 2023; Liu et al., 2019; Millah et al., 2022; Serlina, 2020). Oleh karena itu, data polutan dan meteorologi akan untuk melakukan prediksi konsentrasi polutan dalam kualitas udara.

Penelitian mengenai prediksi konsentrasi polutan telah dilakukan sebelumnya dengan berbagai metode. Metode *deep learning* memiliki kemampuan pengembangan fitur yang lebih kuat dan unggul dibandingkan metode *machine learning* atau statistik (Dun et al., 2022). Metode *Recurrent Neural Network* (RNN), seperti *Gated Recurrent Unit* (GRU) (Duan et al., 2021) dan *Long-Short Term Memory* (LSTM) (Li et al., 2023) digunakan untuk melakukan ekstraksi fitur temporal pada data *time series*. Namun, metode ini hanya berfokus pada korelasi temporal, sehingga tidak efisien dalam menangani fitur spasial. Selain itu, metode LSTM hanya dapat mengolah data dari stasiun tunggal karena bersifat satu dimensi. Metode *Convolution Neural Network* (CNN) banyak digunakan oleh beberapa peneliti karena mampu melakukan ekstraksi fitur-fitur penting pada berbagai tingkatan abstraksi, terutama pola spasial (Kow et al., 2020), tetapi memiliki tingkat sensitivitas yang rendah pada perbedaan yang cukup kecil sehingga sulit untuk membedakan kelas yang sangat mirip pada data yang serupa (Wang et al., 2019).

Metode CNN juga dikombinasikan dengan LSTM (Portal-Porras et al., 2023; Qin et al., 2019), di mana LSTM mampu mengekstraksi fitur temporal dan CNN mengekstraksi fitur spasial. Namun, CNN-LSTM memiliki tiga permasalahan utama (Shi et al., 2015). Pertama, CNN sulit untuk mengekstrak fitur spasial dari data polutan secara mendalam, yang dapat dengan mudah menyebabkan hilangnya informasi fitur dan degradasi model. Kedua, CNN-LSTM mengekstraksi fitur temporal dan spasial polutan sebagai proses yang tidak sinkron, sehingga sulit untuk mengekstrak fitur korelasi *spatiotemporal* polutan dari berbagai kota dan data meteorologi (He et al., 2022; L. Zhang et al., 2017). Ketiga, LSTM kebanyakan mengekstrak fitur deret waktu satu dimensi, di mana tidak mungkin untuk memproses data masukan berdimensi tinggi (Lindemann et al., 2021; Q. Zhang et al., 2020). Selain itu, LSTM tidak mampu mendeskripsikan karakteristik lokal pada data spasial (B. Zhang et al., 2022). Oleh karena itu, lahirlah metode Conv-LSTM untuk menangkap karakteristik spasial dan temporal, terutama pada data berdimensi tinggi secara bersamaan.

Metode Conv-LSTM digunakan dalam prediksi curah hujan (Shi et al., 2015). Namun, Conv-LSTM tunggal memiliki kelemahan di satu sisi, dibatasi oleh fitur data masukan, yaitu dimensi *hidden state* dipengaruhi oleh dimensi fitur data masukan. Di sisi lain, seiring **bertambahnya jumlah *layer* Conv-LSTM, model akan mengalami peningkatan kompleksitas** dan biaya pelatihan yang semakin tinggi (B. Zhang et al., 2022). Semakin dalam arsitektur jaringan, semakin banyak parameter yang harus dipelajari dan semakin sulit untuk mengoptimalkan model, sehingga besar kemungkinan terjadi degradasi jaringan (Szegedy et al., 2017). Conv-LSTM juga memiliki kemungkinan mengalami *vanishing gradient* akibat peningkatan kompleksitas jaringan, di mana nilai gradien menjadi sangat kecil, hingga dikatakan menghilang sehingga nilai bobot tidak mengalami perubahan dan model memperoleh hasil yang stagnan atau tidak konvergen (Muflikhah & Cholissodin, 2022).

Arsitektur CNN juga digunakan untuk meningkatkan akurasi prediksi kualitas udara, salah satunya adalah Residual Networks (ResNet). Arsitektur ini juga dapat dikombinasikan dengan *deep learning* menjadi *Deep Residual Networks* (Q. Zhang et al., 2020), tetapi belum mampu melakukan prediksi secara akurat pada data yang tidak stabil dan terus berubah seiring waktu sehingga hanya menghasilkan prediksi sebesar 80%. Arsitektur ResNet untuk prediksi PM2.5 (Cheng et al., 2022; Song et al., 2020) memiliki hasil yang cukup bagus, di mana hasil akurasi yang didapatkan lebih dari 80% pada data validasi dan pengujian. Selain itu, teknik *skip connections* atau *residual connections* yang dimiliki ResNet dapat mengatasi masalah degradasi jaringan dan *vanishing gradient* akibat peningkatan kompleksitas jaringan (Kalajdjieski et al., 2020; Szegedy et al., 2017). Kombinasi ResNet dan Conv-LSTM tidak hanya menyelesaikan masalah korelasi fitur spasial dan temporal pada data berdimensi tinggi, tetapi juga mampu mengatasi *vanishing gradient* (B. Zhang et al., 2022; L. Zhang et al., 2017). Oleh karena itu, ResNet dan Conv-LSTM dapat dijadikan solusi untuk menyelesaikan permasalahan *vanishing gradient* pada prediksi konsentrasi polutan dalam kualitas udara.

Berdasarkan landasan pengetahuan yang didapatkan, penelitian ini akan memanfaatkan ResNet untuk ekstraksi fitur spasial dan Conv-LSTM untuk ekstraksi fitur *spatiotemporal* pada data polutan dan meteorologi. Diharapkan penggunaan arsitektur ResNet Conv-LSTM dapat mengatasi *vanishing gradient* saat melakukan prediksi konsentrasi polutan dalam kualitas udara.

# **Rumusan Masalah**

Rumusan masalah pada penelitian ini, yaitu:

1. **Hilangnya nilai gradien pada tiap layer selama proses pelatihan diakibatkan oleh meningkatnya kompleksitas jaringan** pada prediksi konsentrasi polutan dalam kualitas udara, mengakibatkan akurasi menjadi tidak optimal.
2. Apakah ResNet Conv-LSTM dapat mengatasi masalah *vanishing gradient* sehingga akurasi menjadi lebih optimal?
3. Apakah ResNet Conv-LSTM dapat meningkatkan akurasi prediksi konsentrasi polutan dalam kualitas udara?

# **Batasan Masalah**

Batasan masalah pada penelitian ini, yaitu:

1. Penelitian ini mengambil dua jenis data historis, yaitu data konsentrasi polutan dan data meteorologi dari lima stasiun pemantauan kualitas udara pada lima kota/kabupaten administrasi di DKI Jakarta, Indonesia.
2. Sumber data merupakan kombinasi data primer-sekunder yang diambil dari Satu Data Jakarta untuk data konsentrasi polutan dan Data Online Pusat Database BMKG untuk data meteorologi.
3. Data konsentrasi polutan meliputi data pengukuran zat PM10, SO2, CO, O3, dan NO2 harian dari tahun 2019 – 2021.
4. Data meteorologi meliputi data pengukuran suhu rata-rata, kelembaban rata-rata, kecepatan angin rata-rata, dan arah angin saat kecepatan maksimum harian dari tahun 2019 – 2021.
5. Data pendukung seperti data lalu lintas dan penggunaan lahan tidak digunakan dalam penelitian ini.

# **Tujuan Penelitian**

Tujuan dari penelitian ini, yaitu:

1. Menerapkan **ResNet Conv-LSTM untuk mengatasi *vanishing gradient* akibat meningkatnya kompleksitas jaringan** pada prediksi.
2. Mendapatkan hasil prediksi konsentrasi polutan dalam kualitas udara yang akurat menggunakan ResNet Conv-LSTM.

# **Manfaat Penelitian**

Hasil penelitian ini dapat dijadikan acuan dan bahan pertimbangan dalam membuat model *machine learning* untuk memprediksi konsentrasi polutan dalam kualitas udara menggunakan metode ResNet dan Conv-LSTM.

# **Tahapan Penelitian**

Tahapan-tahapan pada penelitian yang akan dilakukan menggunakan metode kuantitatif dengan menerapkan model ResNet-Conv-LSTM (B. Zhang et al., 2022) untuk melakukan pengukuran evaluasi, sebagai berikut:

1. Identifikasi Masalah

Tahapan pertama dari penelitian ini adalah mengidentifikasi masalah yang bertujuan untuk mendapatkan informasi sebagai langkah awal penelitian. Masalah yang diidentifikasi diperoleh melalui studi literatur dari jurnal, buku, dan penelitian terdahulu yang digunakan sebagai landasan dari penelitian yang dikerjakan.

1. Pengumpulan Data

Tahapan selanjutnya adalah pengumpulan data. Data yang digunakan pada penelitian adalah data primer-sekunder yang bersifat publik dan diambil dari beberapa sumber dengan total 27.400 data. Data tersebut merupakan data historis lima zat polutan dari lima stasiun pemantauan kualitas udara dan empat data meteorologi dari lima stasiun meteorologi di DKI Jakarta dari tahun 2019 – 2021.

1. *Data Preprocessing*

Tahapan selanjutnya adalah melakukan *data preprocessing* pada data yang telah diperoleh, yaitu dengan menghapus data yang tidak digunakan dan mengisi data yang hilang menggunakan interpolasi linear. Kemudian, menerapkan *correlation coefficient Pearson* untuk menganalisis korelasi *spatiotemporal* antara data polutan dan meteorologi. Terakhir, data akan dibagi menjadi data latih, validasi, dan tes.

1. Pembangunan Model

Setelah didapatkan data yang sesuai, proses selanjutnya adalah melakukan konfigurasi arsitektur model, pemilihan parameter, dan pembangunan struktur menggunakan metode ResNet yang mampu mengatasi *vanishing gradient* untuk mengekstrak fitur spasial. Kemudian, menambahkan metode Conv-LSTM untuk mengekstrak fitur *spatiotemporal*. Terakhir, menambahkan *fully-connected layer* untuk melakukan prediksi konsentrasi polutan dalam kualitas udara.

1. Pelatihan Model

Model yang telah dibangun, kemudian di-*compile* menggunakan Adam *optimizer* dan dilatih dengan *dataset* yang telah disiapkan untuk validasi dan tes model.

1. Analisis dan Evaluasi Hasil

Setelah seluruh tahapan telah dilakukan, langkah terakhir adalah mengevaluasi model menggunakan MAE dan memvisualisasikan performa model untuk dianalisis apakah hasil yang didapat mampu menjawab masalah yang telah dirumuskan sebelumnya. Kemudian, akan dilakukan pula pengujian secara *real-time* menggunakan data baru. Hasil analisis dan evaluasi yang didapatkan pada tahapan ini akan digunakan sebagai acuan dalam penarikan kesimpulan mengenai kemampuan model melakukan prediksi konsentrasi polutan pada kualitas udara menggunakan metode ResNet Conv-LSTM.

1. Pengembangan Sistem (Metode *Waterfall* (Pressman, 2015))
2. *Communication (Project Initiation & Requirement Gathering)*

Tahapan pertama adalah mengumpulkan informasi mengenai kebutuhan sistem perangkat lunak melalui observasi. Informasi kemudian diolah dan dianalisis untuk mendapatkan data yang lengkap mengenai spesifikasi sistem perangkat lunak yang akan dikembangkan.

1. *Planning (Estimating, Scheduling, Tracking)*

Tahapan selanjutnya adalah melakukan perancangan arsitektur sistem berdasarkan kebutuhan yang telah diidentifikasi. Rancangan ini dilakukan untuk memberikan gambaran lengkap mengenai apa yang harus dikerjakan, menyiapkan kebutuhan perangkat keras, dan mendefinisikan arsitektur perangkat lunak secara keseluruhan.

1. *Modeling (Analysis & Design)*

Pada tahap ini. sistem dikembangkan melalui pembagian menjadi unit-unit kecil dengan menggunakan bahasa pemrograman. Setiap unit akan dilakukan tahap pengujian dan pemeriksaan fungsionalitas.

1. *Construction (Code & Test)*

Setiap unit yang dikembangkan akan diintegrasikan menjadi satu kesatuan dan dilakukan pemeriksaan serta pengujian sistem secara keseluruhan untuk memastikan sistem telah memenuhi kriteria dan mengidentifikasi adanya kegagalan. Sistem diuji menggunakan metode pengujian *black box*.

1. *Deployement (Delivery, Support, & Feedback)*

Pada tahap ini, sistem sudah dapat dioperasikan oleh pengguna dan dilakukan pemeliharaan untuk memperbaiki kesalahan yang tidak terdeteksi dalam tahap pembuatan. Selain itu, juga dilakukan pengembangan unit sistem, serta peningkatan dan penyesuaian sistem dengan kebutuhan pengguna berdasarkan *feedback* yang diberikan.

# **Sistematika Penulisan**

Sistematika penulisan dalam penyusunan laporan penelitian ini terbagi menjadi lima bab, yaitu:

|  |  |
| --- | --- |
| **BAB I** | **PENDAHULUAN**  Menjelaskan latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, tahapan penelitian, dan sistematika penulisan. |
| **BAB II** | **TINJAUAN PUSTAKA**  Menjelaskan studi pustaka mengenai data polutan dan meteorologi, metode ResNet dan Conv-LSTM, interpolasi linear untuk mengisi data yang hilang, *coefficient correlation Pearson* untuk menentukan korelasi data spatiotemporal, dan metode evaluasi model MAE yang mendasari penelitian secara terperinci serta memuat landasan teori yang akan dibahas pada penelitian, hasil penelitian sebelumnya, dan *gap research* yang akan dilakukan. |
| **BAB III** | **METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM**  Menjelaskan metode yang digunakan dalam penelitian untuk menyelesaikan permasalahan yang diangkat terdiri dari identifikasi masalah hingga pengujian. |
| **BAB IV** | **HASIL DAN PEMBAHASAN**  Menjelaskan analisis dan pembahasan dari hasil yang didapat pada sistem yang telah dibangun. Hasil penelitian mencakup tampilan program dan hasil prediksi konsentrasi polutan pada kualitas udara yang dibuat berdasarkan rancangan pada metodologi penelitian. |
| **BAB V** | **KESIMPULAN DAN SARAN**  Menjelaskan kesimpulan yang didapat dari hasil penelitian ini dan memberikan saran yang dapat digunakan sebagai acuan untuk pengembangan penelitian selanjutnya. |