# **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 zat polutan dalam kualitas udara dapat memberikan peringatan dini yang berguna bagi masyarakat untuk mengambil tindakan yang tepat. Namun, konsentrasi zat polutan ini sangat dipengaruhi oleh data meteorologi seperti kelembapan, suhu, kecepatan angin, arah angin, dan curah hujan (Liu et al., 2019)(Serlina, 2020) (Millah et al., 2022)(Hidayat et al., 2022). Oleh karena itu, data polutan dan meteorologi akan untuk melakukan prediksi konsentrasi zat polutan dalam kualitas udara.

Penelitian mengenai prediksi konsentrasi zat 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 *one-dimensional*. 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), di mana LSTM mampu mengekstraksi fitur temporal dan CNN mengekstraksi fitur spasial. Namun, semakin tinggi tingkat dimensi data dapat menyebabkan hilangnya beberapa karakteristik akibat kompleksitas representasi yang dilakukan, yang disebut *gradient vanishing*.

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* (Zhang et al., 2020), tetapi 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 (Song et al., 2020) dan (Cheng et al., 2022) memiliki hasil yang cukup bagus, di mana hasil akurasi yang didapatkan di atas 80% pada data validasi dan pengujian. Selain itu, ResNet juga digunakan untuk mengatasi masalah gradient vanishing karena tingkat dimensi data yang semakin tinggi, di mana nilai gradien menjadi sangat kecil, terutama pada jaringan dalam (Kalajdjieski et al., 2020). Oleh karena itu, ResNet dapat dijadikan solusi untuk menyelesaikan permasalahan *gradient vanishing* pada prediksi konsentrasi zat polutan dalam kualitas udara menggunakan kombinasi metode CNN dan LSTM.

Berdasarkan landasan pengetahuan yang didapatkan, penelitian ini akan memanfaatkan metode CNN dengan arsitektur ResNet untuk ekstraksi fitur spasial dan LSTM untuk ekstraksi fitur temporal pada data kualitas udara. Diharapkan penggunaan arsitektur ResNet CNN dapat mengatasi *gradient vanishing* saat melakukan prediksi konsentrasi zat polutan.

# **Rumusan Masalah ????**

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah mengatasi *gradient vanishing (tujuan)* akibat tingginya dimensi data pada prediksi konsentrasi zat polutan dalam kualitas udara menggunakan metode ResNet CNN dan LSTM.

Tujuan..

Masalah , kenapa diatasi????

Tujuan : membuat roti

Masalah ? roti yang dijual mahal

Roti yang dijual tidak sehat

Roti tidak ada yang menjual

Tujuan : mengatasi *gradient vanishing*

1. Data hilang saat training
2. Kalau hilang maka…..
3. Urgensinya , GV diatasi, harus mengelupas
4. Data cuaca dimensinya tinggi sehingga akan hilang saat training????,

# **Batasan Masalah**

Batasan masalah pada penelitian ini, yaitu:

1. Penelitian ini mengambil tiga jenis data historis, yaitu data zat 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 zat polutan dan Data Online Pusat Database BMKG untuk data meteorologi.
3. Data zat polutan meliputi data pengukuran zat PM10, SO2, CO, O3, dan NO2 harian dari tahun 2019 – 2021.
4. Data meteorologi meliputi data pengukuran temperatur rata-rata, kelembapan rata-rata, kecepatan angin rata-rata, curah hujan, dan arah angin harian dari tahun 2019 – 2021.
5. Data geospasial yang juga berpengaruh pada prediksi konsentrasi zat polutan dalam kualitas udara seperti data lalu lintas, jumlah penggunaan kendaraan dan lahan, serta data kependudukan tidak digunakan dalam penelitian ini karena ketersediaan data dari pemerintah yang mini ditampilkan di internet.

# **Tujuan Penelitian**

Tujuan dari penelitian ini, yaitu:

1. Menerapkan arsitektur ResNet yang dapat mengatasi *gradient vanishing* akibat tingginya dimensi data pada prediksi menggunakan metode CNN dan LSTM.
2. Melakukan esktraksi fitur spasial menggunakan metode CNN dan esktraksi fitur temporal menggunakan metode LSTM.
3. Mendapatkan hasil prediksi konsentrasi zat polutan dalam kualitas udara yang akurat setelah CNN dioptimalisasi dengan arsitektur ST-ResNet dan dikombinasikan dengan LSTM.

# **Manfaat Penelitian**

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

# **Tahapan Penelitian**

Tahapan-tahapan yang dijalankan pada penelitian ini memiliki target pencapaian masing-masing dan terbagi menjadi dua, yaitu metodologi penelitian dan metode pengembangan sistem.

# **Metodologi Penelitian**

Tahapan-tahapan pada penelitian yang akan dilakukan menggunakan metode kuantitatif dengan menerapkan model Stacked ResNet-LSTM (Cheng 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 18.250 data. Data tersebut merupakan data historis lima zat polutan dan lima data meteorologi dari lima stasiun pemantauan kualitas udara di DKI Jakarta dari tahun 2019 – 2021.

1. Pra-pemrosesan Data

Tahapan selanjutnya adalah melakukan pra-pemrosesan data yang sudah diperoleh berupa pengisian nilai yang hilang dengan nilai 0.

1. Pembangunan Model

Setelah didapatkan data yang sesuai, proses selanjutnya adalah melakukan konfigurasi arsitektur model, pemilihan parameter, dan pembangunan struktur menggunakan metode ResNet CNN untuk esktraksi fitur spasial dan mengatasi *gradient vanishing*. Kemudian, menggunakan metode LSTM untuk ekstraksi fitur temporal dan prediksi konsentrasi zat polutan dalam kualitas udara.

1. Pelatihan Model

Evaluasi performa model dilakukan untuk melakukan prediksi Indeks Standar Pencemar Udara (ISPU) dan konsentrasi zat polutan di lima kota/kabupaten adminstrasi DKI Jakarta selama 24 jam ke depan. *Mean Squarred Error* (MSE), *Root Mean Squarred Error* (RMSE), dan *Mean Absolute Error* (MAE) akan diterapkan untuk mengevaluasi model prediksi.

1. Analisis dan Evaluasi Hasil

Setelah seluruh tahapan telah dilakukan, langkah terakhir adalah melakukan pengujian model pada data baru. Analisis ini dilakukan untuk mendapatkan kesimpulan mengenai kemampuan model melakukan prediksi konsentrasi zat polutan pada kualitas udara menggunakan metode ResNet CNN dan LSTM.

1. *Deployment*

Setelah model berhasil dibangun dan didapatkan hasil yang diinginkan, langkah selanjutnya adalah membangun sistem dan melakukan *deployment* dalam bentuk *website*. Kemudian, sistem akan diuji menggunakan metode pengujian *black box*.

# **Metodologi Pengembangan Sistem**

Tahapan-tahapan pada pengembangan penelitian menerapkan metode *Waterfall* (Pressman, 2015). Metode *Waterfall* merupakan pengembangan perangkat lunak yang bersifat linear dan berurutan antar satu fase ke fase berikutnya. Terdapat lima tahapan dalam metode *Waterfall*, yaitu:

1. *Communication (Project Initiation & Requirement Gathering)*

Tahapan pertama adalah mengumpulkan informasi kebutuhan pengguna untuk menentukan karakteristik, tujuan, dan kendala sistem melalui observasi dan diskusi. 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.

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

Pada tahap ini, sistem sudah dapat dioperasikan oleh pengguna dan dilakukan pemeliharaan untuk memperbaiki kesalahan yang tidak terdeteksi selama 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 arsitektur *Residual Network* pada metode *Convolutional Neural Network* dan *Long-Short Term Memory* yang mendasari penelitian secara terperinci dan 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 zat 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. |