# **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). Oleh karena itu, adanya prediksi kualitas udara dapat memberikan peringatan dini yang berguna bagi masyarakat untuk mengambil tindakan yang tepat. Namun, tantangan muncul dari tingginya korelasi data spatiotemporal antar stasiun dan faktor-faktor seperti kelembapan dan kecepatan angin yang mempengaruhi prediksi kualitas udara (Dun et al., 2022).

Penelitian mengenai prediksi kualitas udara telah dilakukan sebelumnya dengan berbagai metode, salah satunya adalah metode *deep learning* yang memiliki kemampuan pengembangan fitur yang lebih kuat dan unggul dibandingkan metode *machine learning* atau statistik (Dun et al., 2022). Metode *Recurrent Neural Network* (RNN) digunakan untuk melakukan prediksi polusi udara di Tokyo (Ong et al., 2016), di mana metode ini berfokus pada korelasi temporal, sehingga tidak efisien dalam menangani informasi spasial. Metode *Convolutional Neural Nework* (CNN) banyak digunakan oleh beberapa peneliti karena mampu melakukan ekstraksi fitur-fitur penting pada berbagai tingkatan abstraksi, 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 *Long-Short Term Memory* (LSTM), salah satu arsitektur RNN (Portal-Porras et al., 2023), tetapi semakin tinggi tingkat dimensi data dapat menyebabkan hilangnya beberapa karakteristik akibat kompleksitas representasi yang dilakukan. Selain itu, metode LSTM hanya dapat mengolah data dari stasiun tunggal karena bersifat *one-dimensional*.

Penggunaan arsitektur pada CNN juga pernah dibahas seperti AirRes (Zhang et al., 2020), tetapi belum mampu melakukan prediksi secara akurat pada data yang tidak stabil dan terus berubah seiring waktu. Arsitektur *Residual Network* (ResNet) (Cheng et al., 2022) untuk prediksi kualitas udara memiliki hasil yang cukup bagus, di mana hasil akurasi yang didapatkan di atas 80% pada data validasi dan pengujian. Sayangnya, metode ini (Kalajdjieski et al., 2020) masih sulit membedakan data yang serupa dan memberikan hasil prediksi yang berbeda karena adanya ketidakseimbangan kelas pada data *spatiotemporal* yang digunakan. Data *spatiotemporal* cenderung mengalami *overfitting* akibat *over-sampling* karena keterbatasan variasi data udara dan minimnya stasiun pemantauan. Oleh karena itu, arsitektur *Spatiotemporal Residual Network* (ST-ResNet) dapat dijadikan solusi untuk menyelesaikan permasalahan ketidakseimbangan kelas pada data *spatiotemporal*. ST-ResNet memanfaatkan koneksi residual untuk mengintegrasikan dimensi spasial dan temporal sesuai dengan data yang digunakan, yaitu data AQI, meteorologi dan geospasial.

Berdasarkan landasan pengetahuan yang didapatkan, penelitian ini akan memanfaatkan metode CNN dengan arsitektur ST-ResNet untuk mengatasi ketidakseimbangan kelas pada data *spatiotemporal*. Diharapkan penggunaan arsitektur ST-ResNet pada metode CNN dapat menangani permasalahan *overfitting* akibat *oversampling* saat melakukan prediksi kualitas udara*.*

# **Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah mengatasi ketidakseimbangan kelas pada data *spatiotemporal* menggunakan aristektur ST-ResNet dengan metode CNN. Ketidakseimbangan kelas dapat menyebabkan *overfitting* akibat *oversampling* saat melakukan prediksi kualitas udara.

# **Batasan Masalah**

Batasan masalah pada penelitian ini, yaitu:

1. Penelitian ini mengambil tiga jenis data historis, yaitu *Air Quality Index* (AQI), data meteorologi, dan data geospasial DKI Jakarta, Indonesia.
2. Sumber data merupakan data sekunder yang diambil dari Open Data Jakarta untuk data AQI, Badan Pusat Statistik DKI Jakarta untuk data meteorologi dan data geospasial.
3. Data AQI meliputi data pengukuran zat PM10, SO2, CO, O3, dan NO2 harian dari tahun 2019 – 2021.
4. Data meteorologi meliputi data pengukuran suhu dan kelembapan bulanan dari tahun 2019 – 2021.
5. Data geospasial meliputi data jumlah penduduk dan penggunaan kendaraan tahunan dari tahun 2019 – 2021.

# **Tujuan Penelitian**

Tujuan dari penelitian ini, yaitu:

1. Menerapkan arsitektur ST-ResNet yang dapat mengatasi *overfitting* akibat *over-sampling* pada data *imbalanced spatiotemporal* dengan metode CNN.
2. Mendapatkan hasil prediksi kualitas udara yang akurat setelah CNN dioptimalisasi dengan arsitektur ST-ResNet.
3. Mendapatkan cara menyeimbangkan data yang tidak seimbang menggunakan arsitektur ST-ResNet pada metode CNN.

# **Manfaat Penelitian**

Hasil penelitian ini dapat dijadikan acuan dan bahan pertimbangan dalam membuat model *machine learning* untuk memprediksi kualitas udara menggunakan arsitektur ST-ResNet.

# **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 sekunder yang bersifat publik dan diambil dari beberapa sumber dengan total 1.220 data. Data tersebut merupakan data historis Indeks Kualitas Udara (AQI), data meteorologi, dan data geospasial DKI Jakarta dari tahun 2019 – 2021.

1. Pra-pemrosesan Data

Tahapan selanjutnya adalah melakukan pra-pemrosesan data yang sudah diperoleh berupa pemilahan dan pembersihan menggunakan metode *over-sampling* untuk meningkatkan keseimbangan kelas pada data minoritas (sifat data bulanan dan tahunan).

1. Pembangunan Model

Setelah didapatkan data yang sesuai, proses selanjutnya adalah melakukan konfigurasi arsitektur model, pemilihan parameter, dan pembangunan struktur menggunakan arsitektur ST-ResNet dengan metode CNN.

1. Pelatihan Model

Evaluasi performa model dilakukan untuk melakukan prediksi Indeks Standar Pencemar Udara (ISPU) di DKI Jakarta selama 6 jam ke depan. *Mean Squarred Error* (MSE) akan diterapkan untuk mengukur seberapa akurat prediksi ISPU yang dilakukan oleh model.

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 kualitas udara menggunakan arsitektur ST-ResNet dengan metode CNN.

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 dianalisa 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 metode *Convolutional Neural Network* dan arsitektur ST-ResNet 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 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. |