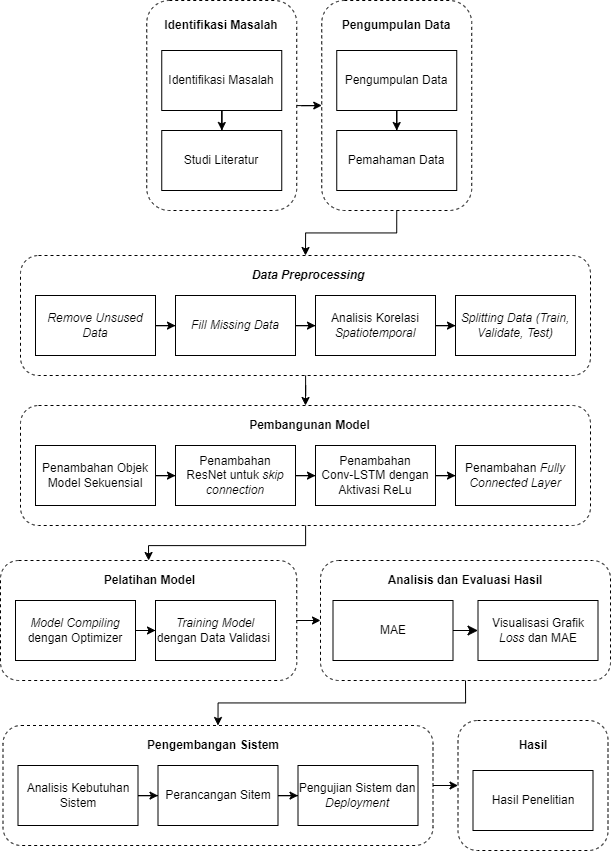
# **BAB III**

# **METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM**

# **Metodologi Penelitian**

Pada bagian ini akan dibahas mengenai metodologi yang akan dilakukan dalam penelitian ini, yaitu mengatasi *vanishing gradient* menggunakan ResNet Conv-LSTM pada prediksi konsentrasi polutan dalam kualitas udara. Penelitian ini menerapkan metode kuantitatif menggunakan data numerik dan statistik untuk menafsirkan informasi yang digunakan dalam pengujian hipotesis dengan teori-teori yang sudah ada. Adapun data yang digunakan dalam penelitian ini adalah data historis zat polutan dan meteorologi Provinsi DKI Jakarta. Tahapan penelitian ditunjukkan pada **Gambar 3.1**.



**Gambar 3.1.** Tahapan Penelitian

Tahapan penelitian di mulai dari identifikasi masalah, pengumpulan data, *data preprocessing*, pembangunan model, pelatihan model, analisis dan evaluasi hasil, pengembangan sistem, dan diakhir dengan hasil penelitian.

# **Identifikasi Masalah**

1. Identifikasi Masalah

Penelitian ini diawali dengan mengidentifikasi masalah yang akan diangkat dan diselesaikan dalam penelitian. Masalah yang diangkat adalah masalah *vanishing gradient* akibat meningkatnya kompleksitas jaringan pada prediksi data *spatiotemporal* yang menyebabkan akurasi menjadi tidak optimal.

1. Studi Literatur

Studi literatur dilakukan untuk memperoleh informasi yang relevan yang akan digunakan sebagai dasar atau acuan pada penelitian ini. Tahapan ini dilakukan dengan menelusuri beberapa sumber seperti buku, jurnal, dan penelitian terdahulu mengenai *vanishing gradient* dan prediksi konsentrasi polutan dalam kualitas udara. Informasi yang didapatkan digunakan sebagai dasar untuk menyelesaikan masalah dan mencapai tujuan penelitian. Informasi dari penelitian-penelitian terdahulu dapat dilihat dalam *State of the Art* pada **Tabel 1**.

# **Pengumpulan Data**

1. Pengumpulan Data

Penelitian ini menggunakan kombinasi data primer-sekunder, yaitu data historis konsentrasi polutan dari lima Stasiun Pemantauan Kualitas Udara dan data meteorologi dari lima Stasiun Meteorologi di DKI Jakarta. Masing-masing data diperoleh dari *website* Satu Data Jakarta untuk data konsentrasi polutan dan Data Online Pusat Database BMKG untuk data meteorologi dalam bentuk *excel*. Data ini dikumpulkan setiap 1 hari dari 1 Januari 2019 – 31 Desember 2021 sebanyak 10.960 data.

Pada penelitian ini, dipilih lima konsentrasi polutan, yaitu PM10, SO2, CO, O3, dan NO2 serta empat faktor meteorologi, yaitu suhu rata-rata, kelembaban rata-rata, kecepatan angin rata-rata, dan arah angin saat kecepatan maksimum yang diambil dari Jakarta Pusat, Jakarta Barat, Jakarta Utara, Jakarta Selatan, dan Jakarta Timur.

1. Pemahaman Data

Data konsentrasi polutan dan meteorologi yang diperoleh memiliki beberapa parameter yang digunakan untuk melakukan prediksi menggunakan metode ResNet Conv-LSTM. Detail dari setiap parameter dapat dilihat pada **Tabel 3.1** untuk data konsentrasi polutan dan **Tabel 3.2** untuk data meteorologi.

**Tabel 3.1.** Detail Parameter Data Konsentrasi Polutan

| **No** | **Variabel** | **Deskripsi** | **Keterangan** |
| --- | --- | --- | --- |
| 1 | tanggal | Tanggal pengukuran konsentrasi polutan | Format penanggalan mm/dd/yyyy |
| 2 | stasiun | Nomor/nama Stasiun Pemantauan Kualitas Udara (SPKU) di DKI Jakarta | Sebaran SPKU:   1. DKI 1 – Bundaran HI 2. DKI 2 – Kelapa Gading 3. DKI 3 – Jagakarsa 4. DKI 4 – Lubang Buaya 5. DKI 5 – Kebon Jeruk |
| 3 | pm10 | Rata-rata konsentrasi zat PM10 dalam periode waktu 1 hari pengukuran | Data yang tidak ada dikosongkan atau diberi tanda --- |
| 4 | so2 | Rata-rata konsentrasi zat SO2 dalam periode waktu 1 hari pengukuran | Data yang tidak ada dikosongkan atau diberi tanda --- |
| 5 | co | Rata-rata konsentrasi zat CO dalam periode waktu 1 hari pengukuran | Data yang tidak ada dikosongkan atau diberi tanda --- |
| 6 | no2 | Rata-rata konsentrasi zat NO2 dalam periode waktu 1 hari pengukuran | Data yang tidak ada dikosongkan atau diberi tanda --- |
| 7 | max | Nilai konsentrasi polutan maksimum dalam 1 hari | Data yang tidak ada diberi nilai 0 |
| 8 | critical | Zat polutan dengan nilai konsentrasi maksimum dalam 1 hari | Data yang tidak ada dikosongkan |
| 9 | categori | Kategori Indeks Standar Pencemar Udara (ISPU) dalam 1 hari | Kategori ISPU:   1. BAIK 2. SEDANG 3. TIDAK SEHAT 4. SANGAT TIDAK SEHAT 5. TIDAK ADA DATA |

**Tabel 3.2.** Detail Parameter Data Meteorologi

| **No** | **Variabel** | **Deskripsi** | **Keterangan** |
| --- | --- | --- | --- |
| 1 | Tanggal | Tanggal pengukuran faktor meterologi | Format penanggalan dd-mm-yyy |
| 2 | Tavg | Nilai suhu rata-rata dalam waktu 1 hari pengukuran | Data yang tidak terukur diberi nilai 8888, sedangkan data yang tidak ada diberi nilai 9999 atau dikosongkan. |
| 3 | RH\_avg | Nilai kelembaban rata-rata dalam waktu 1 hari pengukuran | Data yang tidak terukur diberi nilai 8888, sedangkan data yang tidak ada diberi nilai 9999 atau dikosongkan. |
| 4 | RR | Nilai curah hujan dalam waktu 1 hari pengukuran | Data yang tidak terukur diberi nilai 8888, sedangkan data yang tidak ada diberi nilai 9999 atau dikosongkan. |
| 5 | ddd\_x | Nilai arah angin dalam kecepatan maksimum dalam waktu 1 hari pengukuran | Data yang tidak terukur diberi nilai 8888, sedangkan data yang tidak ada diberi nilai 9999 atau dikosongkan. |
| 6 | ff\_avg | Nilai kecepatan angin rata-rata dalam waktu 1 hari pengukuran | Data yang tidak terukur diberi nilai 8888, sedangkan data yang tidak ada diberi nilai 9999 atau dikosongkan. |

# ***Data Preprocessing***

Data yang telah didapatkan kemudian diproses agar menjadi lebih terstruktur sehingga dapat digunakan dalam pembuatan dan pelatihan model. Beberapa proses yang dilakukan pada tahapan ini, yaitu menghapus data-data yang tidak diperlukan, mengisi nilai yang hilang, menganalisis analisis korelasi *spatiotemporal*, dan membagi data.

1. *Remove Unused Data*

Penghapusan data dilakukan karena tidak semua variabel pada *dataset* akan digunakan dalam pembuatan model. Variabel yang dihapus meliputi max, critical, dan categori pada *dataset* konsentrasi polutan.

1. *Fill Missing Value*

Setelah variabel yang tidak digunakan telah dihapus, selanjutnya adalah mengisi nilai yang hilang pada *dataset* konsentrasi polutan dan meteorologi. Pengisian nilai yang hilang ini menggunakan metode interpolasi linear. Misalnya, pada *dataset* konsentrasi polutan terdapat nilai yang hilang pada tanggal 27 Januari 2019 seperti pada **Tabel 3.3**.

**Tabel 3.3**. Contoh Missing Value pada Data Konsentrasi Polutan

|  |  |
| --- | --- |
| **Index** | **O3** |
| 1 | 118 |
| 2 | NaN |
| 3 | 64 |

Di, mana dan , maka penerapan metode interpolasi linear menggunakan rumus (2.17) adalah sebagai berikut:

Sehingga, diperoleh hasil yang merupakan nilai O3 pada tanggal 27 Januari 2019 sebesar 91.

1. Analisis Korelasi *Spatiotemporal*

Korelasi aspek spasial dan temporal sangat penting untuk melakukan prediksi konsentrasi polutan dalam kualitas udara. Berdasarkan data historis ditemukan bahwa perubahan tren pada polutan dan faktor meteorologi secara umum konsisten, yang juga mencerminkan hubungan terkait antara keduanya. **Gambar 3.2** dan **Gambar 3.3** menunjukkan perubahan numerik dalam periode waktu 3 tahun setiap konsentrasi polutan dari 5 Stasiun Pemantauan Kualitas Udara, di mana (a) PM10, (b) SO2, (c) CO, (d) O3, dan (e) NO2. Setelah dilakukan perhitungan statistik, ditemukan bahwa antara tahun 2019 – 2021 sekitar 62% waktu konsentrasi PM10 lebih dari yang ditetapkan WHO, yaitu 50 µg/m3. Konsentrasi PM10 mencapai rata-rata harian tertinggi yaitu 179 µg/m3 terjadi pada 7 Desember 2021 di stasiun DKI 4 Lubang Buaya. Oleh karena itu, korelasi antara PM10 dengan polutan lainnya harus dipertimbangkan untuk memperoleh prediksi konsentrasi yang akurat dan mencegah dampak yang lebih buruk terhadap kesehatan masyarakat sejak dini.

|  |
| --- |
|  |
| (a) |
|  |
| (b) |
|  |
| (c) |

**Gambar 3.2.** Grafik Perubahan Konsentrasi Polutan Tahun 2019 – 2021 dari 5 Stasiun Pemantauan Kualitas Udara, (a) PM10, (b) SO2, dan (c) CO

|  |
| --- |
|  |
| (d) |
|  |
| (e) |
|  |
| (f) |

**Gambar 3,3.** Lanjutan Grafik Perubahan Konsentrasi Polutan Tahun 2019 – 2021 dari 5 Stasiun Pemantauan Kualitas Udara, (d) O3, (e) NO2, dan (f) ISPU

**Gambar 3.4** dan **Gambar 3.5** menunjukkan perubahan numerik dalam periode waktu 3 tahun setiap faktor meteorologi dari 5 Stasiun Meteorologi, di mana (a) suhu rata-rata, (b) kelembaban rata-rata, (c) arah angin, dan (d) kecepatan angin. Tipe numerik dan interval dari masing-masing faktor meteorologi sangat berbeda, tetapi tren perubahannya sangat mirip, yang menandakan bahwa mungkin terdapat pengaruh timbal balik antara faktor tersebut. Sebagai contoh, **Gambar 3.4**(a) menunjukkan perubahan suhu yang cukup fluktuatif dibandingkan kelembaban pada **Gambar 3.4**(b). Pada rentang bulan Juli – September, suhu dan kelembaban rata-rata cenderung mengalami penurunan. Sementara, pada rentang bulan Januari – Maret suhu rata-rata cenderung mengalami penurunan, sedangkan kelembaban rata-rata cenderung mengalami kenaikan. Faktor meteorologi juga konsisten dengan perubahan konsentrasi PM10, yang menandakan terdapat korelasi antara polutan dan faktor meteorologi. Sebagai contoh, Pada rentang bulan Januari – Maret, konsentrasi PM10 cenderung mengalami penurunan, sedangkan Juli – September cenderung mengalami kenaikan. Hal ini berbanding terbalik dengan kelembaban rata-rata. Berdasarkan hasil analisis temporal yang telah dilakukan, faktor meteorologi digunakan sebagai salah satu *input* pada model untuk mengekstrak fitur tersembunyi antara polutan dan faktor meteorologi pada prediksi konsentrasi polutan dalam kualitas udara.

|  |
| --- |
|  |
| (a) |
|  |
| (b) |

**Gambar 3.4.** Grafik Perubahan Faktor Meteorologi Tahun 2019 – 2021 dari 5 Stasiun Meteorologi, (a) Suhu Rata-rata dan (b) Kelembaban Rata-rata

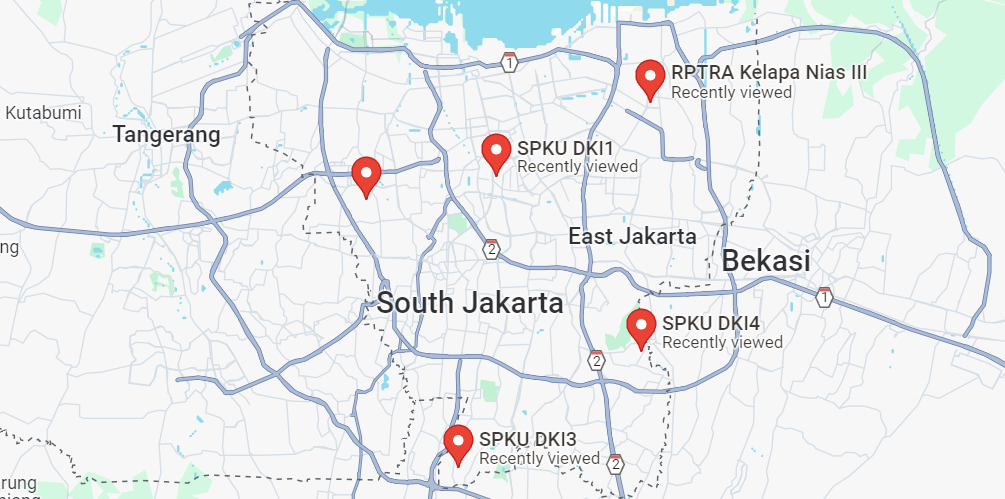
|  |
| --- |
|  |
| (c) |
|  |
| (d) |

**Gambar 3.5.** Lanjutan Grafik Perubahan Faktor Meteorologi Tahun 2019 – 2021 dari 5 Stasiun Meteorologi, (c) Arah Angin, (d) Kecepatan Angin

Polutan PM10 mungkin juga memiliki beberapa karakteristik spasial, misalnya stasiun target adalah Stasiun DKI 1 Bundaran HI. Karakteristik spasial ini dipengaruhi oleh jarak antara Stasiun 1 dengan stasiun lainnya. Lokasi dari seluruh Stasiun Pemantauan Kualitas Udara DKI Jakarta ditunjukkan pada **Gambar 3.6** dan jarak Stasiun 1 dengan stasiun lain ditunjukkan pada **Tabel 3.4**.

**Tabel 3.4.** Jarak Stasiun Target dan Stasiun Lainnya

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Stasiun 2** | **Stasiun 3** | **Stasiun 4** | **Stasiun 5** |
| **Stasiun 1** | 10.52 km | 18.05 km | 13.86 km | 8.12 km |



**Gambar 3.6.** Lokasi Stasiun Pemantauan Kualitas Udara DKI Jakarta

Dilakukan perhitungan *Pearson correlation coefficient* untuk menganalisis korelasi spasial antara Stasiun 1 dengan stasiun lain di sekitarnya. Data konsentrasi polutan PM10 sebagian ditunjukkan pada **Tabel 3.5** dan dihitung menggunakan rumus (2.19), di mana x1 untuk Stasiun 1 dan y2 untuk Stasiun 2 sebagai berikut:

**Tabel 3.5.** Konsentrasi PM10

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Stasiun 1** | **Stasiun 2** | **Stasiun 3** | **Stasiun 4** | **Stasiun 5** |
| 0 | 29 | 14 | 21 | 16 | 10 |
| 1 | 24 | 20 | 20 | 17 | 7 |
| 2 | 16 | 16 | 23 | 15 | 9 |
| 3 | 38 | 33 | 42 | 32 | 39 |
| 4 | 37 | 36 | 54 | 50 | 33 |
| : | : | : | : | : | : |
| 1091 | 52 | 50 | 75 | 68 | 54 |
| 1092 | 51 | 53 | 51 | 59 | 44 |
| 1093 | 31 | 28 | 47 | 61 | 34 |
| 1094 | 55 | 47 | 60 | 60 | 53 |
| 1095 | 62 | 61 | 64 | 58 | 60 |

Hasil perhitungan korelasi spasial PM10 antara Stasiun 1 dengan stasiun lainnya menggunakan *Pearson correlation coefficient* seperti yang ditunjukkan pada **Tabel 3.5**.

**Tabel 3.6.** Pearson Correlation Coefficient Polutan antara Stasiun 1 dengan Stasiun Lainnya

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | PM10 | SO2 | CO | O3 | NO2 | ISPU |
| Stasiun 1 & 2 | **0.872** | 0.613 | **0.620** | 0.711 | **0.741** | **0.830** |
| Stasiun 1 & 3 | 0.681 | **0.703** | 0.454 | **0.747** | 0.717 | 0.632 |
| Stasiun 1 & 4 | 0.712 | 0.528 | 0.194 | 0.590 | 0.242 | 0.705 |
| Stasiun 1 & 5 | **0.847** | 0.478 | 0.201 | 0.655 | 0.209 | **0.810** |

Kombinasi antara **Gambar 3.6** dan **Tabel 3.6** menunjukkan bahwa semakin pendek jarak stasiun dengan Stasiun 1, maka semakin tinggi tingkat korelasi polutan, seperti yang ditunjukkan pada tabel dengan huruf tebal. Selain itu, korelasi koefisien polutan PM10 secara umum lebih tinggi dibandingkan polutan lainnya. Sedangkan, semakin bertambahnya jarak stasiun dari Stasiun 1, korelasi koefisien polutan antara Stasiun 1 dengan stasiun lain akan semakin menurun. Pengaruh jarak menunjukkan bahwa di setiap wilayah stasiun memiliki relevansi spasial polutan udara. Oleh karena itu, perlu dilakukan pencegahan polutan lokal untuk mengurangi dampak buruk yang ditimbulkan dari zat polutan.

1. *Data Splitting*

Tahapan ini diakhiri dengan melakukan *data splitting* untuk membagi data menjadi tiga bagian, yaitu *data train, validate,* dan *test*. Pembagian data yaitu 70% untuk *data* *train*, 15% untuk *data* *validate*, dan 15% untuk *data test* sesuai penelitian Bo Zhang (2022).

# **Pembangunan Model**

# **Pelatihan Model**

# **Analisis dan Evaluasi Hasil**

# **Pengembangan Sistem**

# **Hasil**

# **Pengembangan Sistem**