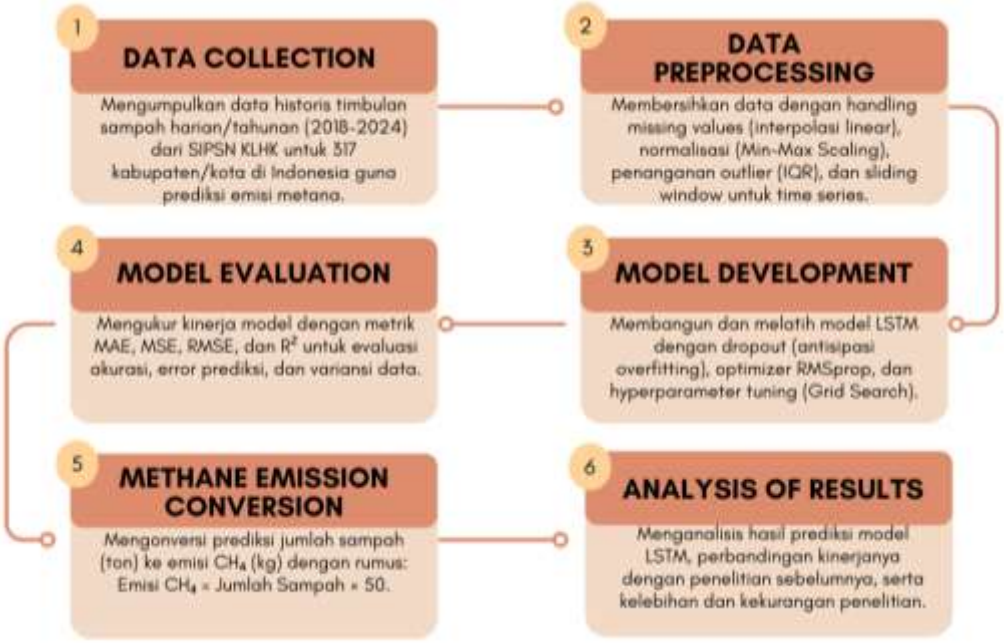
BAB III

METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif dengan model prediksi berbasis *deep learning*, yaitu *Long Short-Term Memory* (LSTM), untuk memprediksi jumlah sampah dan mengestimasi emisi gas metana di masa mendatang. Desain penelitian ini mencakup beberapa tahap, yaitu pengumpulan data, pra-pemrosesan data, pembuatan model, evaluasi model, konversi hasil prediksi jumlah sampah ke estimasi emisi gas metana, serta analisis hasil yang dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Diagram Alur Penelitian

3.1 Pengumpulan Data (*Data Collection*)

Data yang digunakan dalam penelitian ini merupakan data historis jumlah timbulan sampah harian dan tahunan dalam satuan ton dari 317 kabupaten/kota se-

Indonesia tahun 2018 hingga 2024 dalam format CSV, yang diperoleh dari Sistem Informasi Pengelolaan Sampah Nasional (SIPSN) Kementerian Lingkungan Hidup dan Kehutanan (SIPSN KLHK, 2024).

Dataset terdiri atas 1.699 entri yang memuat informasi tahun, provinsi, kabupaten/kota, serta jumlah timbulan sampah harian dan tahunan. Data ini telah melalui tahap pra-pemrosesan berupa pembersihan, normalisasi, dan pembentukan format deret waktu untuk meningkatkan kualitas *input* model. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Model *Long Short-Term Memory* (LSTM) dilatih menggunakan data tersebut untuk memprediksi jumlah sampah tahunan di masa depan. Evaluasi dilakukan dengan membandingkan hasil prediksi dan data aktual menggunakan metrik MAE, MSE, RMSE, dan *R-Squared* (R^2).

3.2 Pra-pemrosesan Data (*Data Preprocessing*)

Preprocessing data dilakukan untuk meningkatkan kualitas dan reliabilitas data sebelum digunakan dalam proses pelatihan model prediksi, sehingga model mampu mempelajari pola secara lebih akurat dan efektif. Langkah-langkah *preprocessing* meliputi beberapa proses berikut:

a. *Handling Missing Values*

   Proses ini mengisi data hilang yang diperkirakan menggunakan interpolasi linear berdasarkan tren data sekitar untuk mempertahankan kontinuitas *time series*.

b. Normalisasi Data

   Transformasi data dilakukan dengan menggunakan metode *Min-Max Scaling* untuk mereduksi skala data ke dalam rentang [0.1], bertujuan mempercepat konvergensi model dan mencegah dominasi fitur tertentu.

c. Deteksi *Outlier*

Distribusi data dianalisis untuk mendeteksi adanya nilai pencilan (*outlier*) dengan menggunakan metode *Interquartile Range* (IQR). Data pencilan yang melebihi batas *upper* dan *lower quartile* akan dihapus atau disesuaikan untuk mencegah bias dalam pelatihan model.

d. *Sliding window*

Teknik *sliding window* digunakan untuk membagi data *time series* menjadi urutan (*sequence*) tertentu, memungkinkan model menangkap ketergantungan temporal dan pola berulang dalam data historis.

3.3 Pembuatan Model LSTM (*Model Development*)

Model *Long Short-Term Memory* (LSTM) dikembangkan untuk memproses data deret waktu (*time series*) jumlah sampah dan menghasilkan prediksi untuk periode berikutnya. LSTM dipilih karena mampu menangkap pola musiman, tren, dan dependensi jangka panjang pada data timbulan sampah melalui mekanisme *cell state* dan tiga *gate* yang efektif mengelola informasi serta mencegah *vanishing gradient*. Model ini juga terbukti lebih akurat dibandingkan metode statistik tradisional dalam prediksi *time series*.

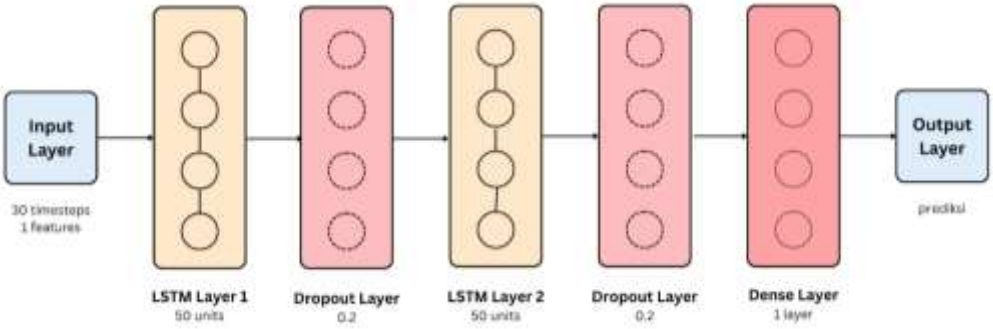
Sebelum menentukan konfigurasi akhir model, dilakukan proses *Grid Search* untuk menemukan kombinasi *hyperparameter* optimal. *Grid Search* dilakukan untuk menemukan kombinasi *hyperparameter* optimal secara sistematis. Proses ini menguji seluruh kemungkinan kombinasi dari parameter-parameter pada Tabel 3.1.

Tabel 3. 1 Parameter yang Diuji dengan *Grid Sarch*

Parameter	**Nilai yang Diuji**	**Alasan Pengujian**
Jumlah Unit LSTM	32, 50, 64	Menguji kapasitas model dari rendah (32) hingga tinggi (64). Unit lebih banyak memberi kapasitas lebih besar tetapi meningkatkan risiko *overfitting*.
Dropout Rate	0.1, 0.2, 0.3	Menguji tingkat regularisasi. *Dropout* terlalu rendah (0.1) mungkin kurang mencegah *overfitting*, terlalu tinggi (0.3) dapat menghilangkan terlalu banyak informasi.
Learning Rate	0.0001, 0.001, 0.01	Menguji kecepatan pembelajaran. LR terlalu kecil memperlambat konvergensi, terlalu besar dapat membuat model tidak stabil.
Batch Size	16, 32, 64	Menguji *trade-off* antara stabilitas gradien dan kecepatan *training*. *Batch* kecil (16) memberi gradien lebih *noisy*, *batch* besar (64) lebih stabil tetapi butuh lebih banyak memori.
Optimizer	Adam, RAdam, RMSprop	Membandingkan berbagai *optimizer* adaptif untuk menemukan yang paling sesuai dengan karakteristik data *time series* sampah.

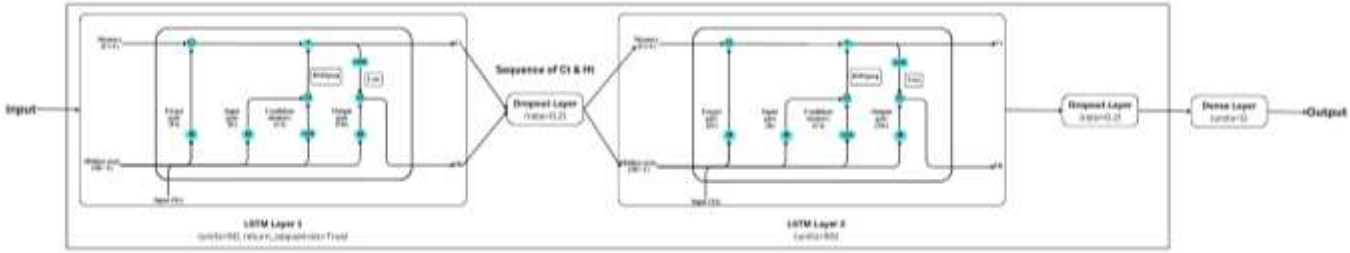
Proses evaluasi *Grid Search* melibatkan 243 kombinasi hasil dari variasi lima *hyperparameter*, dengan setiap kombinasi dilatih hingga 100 *epoch* menggunakan *early stopping* (*patience* = 10). Evaluasi dilakukan menggunakan empat metrik, yaitu MAE, MSE, RMSE, dan R^2, sementara 20% data dialokasikan sebagai data validasi.

Pemilihan RMSProp sebagai *optimizer* utama dilakukan berdasarkan karakteristik adaptifnya yang menyesuaikan *learning rate* menggunakan *moving average squared gradients*. Mekanisme ini membuat RMSProp lebih stabil dalam menghadapi fluktuasi gradien pada data *time series* dan tidak seagresif Adam dalam memperbarui bobot, sehingga lebih sesuai untuk data dengan *noise*. Temuan ini sejalan dengan (Mukkamala & Hein, 2017) yang menegaskan efektivitas RMSProp pada gradien tidak stabil serta (Q. Zhang dkk., 2020) yang menunjukkan jaminan konvergensinya pada optimasi *non-convex*.



Gambar 3. 2 *Layer* Model LSTM

Arsitektur model LSTM pada Gambar 3.2 menggunakan *input* berbentuk (30, 1) untuk menangkap pola temporal dari 30 *timestep* data historis. Model terdiri dari dua *layer* LSTM masing-masing dengan 50 unit, *layer* pertama menggunakan `return_sequences=True` untuk menghasilkan *hidden state* setiap *timestep*, sedangkan *layer* kedua menggunakan `return_sequences=False` untuk mempelajari representasi yang lebih abstrak. Setiap *layer* diikuti *dropout* 0.2 sebagai regularisasi. Tahap akhir terdiri atas *dense layer* dan *output layer* yang menghasilkan nilai prediksi jumlah sampah pada periode berikutnya. Konfigurasi ini dipilih karena mampu menangkap pola kompleks secara efektif sambil menjaga generalisasi dan mencegah *overfitting*.



Gambar 3. 3 Arsitektur Internal Model LSTM

Arsitektur pada Gambar 3.3 menampilkan struktur internal dari sebuah LSTM *cell* yang terdiri atas tiga mekanisme gerbang utama, yaitu *Forget Gate* yang

menentukan informasi mana yang harus dilupakan, *Input Gate* yang mengatur informasi baru yang akan disimpan, serta *Output Gate* yang mengontrol informasi yang dikeluarkan sebagai *output*. Pada mekanisme tersebut, *cell state* (Ct) berfungsi sebagai memori jangka panjang, sedangkan *hidden state* (Ht) menjadi *output* yang diteruskan ke tahap berikutnya. Kedua gambar ini saling berhubungan karena setiap "LSTM *Layer*" pada Gambar 3.2 tersusun atas banyak LSTM *cell* yang bekerja paralel dengan struktur internal seperti pada Gambar 3.3. Setiap *layer* berisi 50 unit, sehingga terdapat 50 LSTM *cell* yang secara bersamaan mengekstraksi berbagai pola temporal dari data *input*.

3.4 Evaluasi Model (*Model Evaluation*)

Evaluasi model dilakukan secara komprehensif menggunakan beberapa metrik evaluasi untuk mengukur performa dan keakuratan prediksi. Metrik-metrik tersebut meliputi:

a. *Mean Absolute Error* (MAE) digunakan untuk menghitung rata-rata kesalahan absolut antara nilai aktual dan prediksi dalam satuan yang sama dengan data asli, sehingga MAE yang rendah menunjukkan prediksi yang lebih akurat.
b. *Mean Squared Error* (MSE) untuk mengkuadratkan selisih antara nilai aktual dan prediksi sehingga lebih sensitif terhadap *error* besar.
c. *Root Mean Squared Error* (RMSE) untuk mengukur akar kuadrat dari MSE yang mengembalikan nilai *error* ke satuan asli data sekaligus tetap memberikan penalti lebih besar pada *error* yang besar.

d. *R-Squared* (R^2) digunakan untuk mengukur seberapa besar variasi data yang dapat dijelaskan oleh model, dengan nilai mendekati 1 menunjukkan kinerja yang sangat baik.

Proses evaluasi dilakukan dengan melatih model pada 80% data dan menguji performanya pada 20% data yang tidak pernah dilihat selama *training*. Keempat metrik tersebut dihitung untuk memberikan gambaran performa dari berbagai perspektif, mulai dari kesalahan absolut hingga kemampuan model menjelaskan variasi data. Hasil evaluasi kemudian dibandingkan dengan penelitian sebelumnya untuk memastikan validitas dan keunggulan model yang dikembangkan.

3.5 Konversi ke Estimasi Gas Metana (*Methane Emission Conversion*)

Hasil prediksi jumlah sampah tahunan yang dihasilkan oleh model LSTM selanjutnya dikonversi menjadi estimasi emisi gas metana menggunakan formula berdasarkan faktor emisi IPCC yang telah divalidasi oleh (Rarastry, 2016), sebagaimana dijelaskan pada subbab 2.9, dengan Persamaan 2.13.

Formula ini mengasumsikan bahwa setiap 1 ton sampah menghasilkan 50 kg gas metana. Konversi ini bertujuan untuk menganalisis dampak lingkungan dari jumlah sampah yang diprediksi, sehingga tidak hanya memberikan informasi kuantitatif terkait volume sampah, tetapi juga mengaitkannya dengan potensi kontribusi terhadap emisi gas rumah kaca.

Melalui pendekatan ini, hasil penelitian tidak hanya memproyeksikan tren jumlah sampah tahunan, tetapi juga mengevaluasi implikasi ekologisnya dalam konteks pemanasan global, mengingat metana memiliki potensi pemanasan global 28 kali lebih tinggi dibandingkan CO_2 dalam periode 100 tahun. Berdasarkan hal

tersebut, model ini dapat menjadi alat strategis dalam mendukung pengambilan keputusan terkait pengelolaan sampah dan mitigasi perubahan iklim di tingkat nasional maupun regional.

3.6 Analisis Hasil (*Analysis of Results*)

Tahapan ini akan membahas secara komprehensif hasil pemodelan LSTM yang telah dibangun, mencakup evaluasi kinerja model, kesesuaian hasil prediksi dengan data aktual, serta implikasi prediksi terhadap estimasi emisi gas metana. Subbab ini menyajikan perbandingan kinerja model dengan penelitian-penelitian terdahulu untuk menilai posisi dan kontribusi penelitian ini. Pembahasan dilengkapi dengan analisis kelebihan dan kekurangan penelitian, sehingga memberikan gambaran menyeluruh mengenai efektivitas model sebagai alat prediksi sekaligus dukungan dalam perencanaan pengelolaan sampah nasional yang berkelanjutan.