

BAB I

PENDAHULUAN

1.1. Latar Belakang

Polusi udara merupakan salah satu masalah lingkungan yang semakin mengkhawatirkan seiring dengan pesatnya urbanisasi dan industrialisasi. Kota-kota besar menjadi pusat konsentrasi zat polutan akibat tingginya aktivitas transportasi dan industri. Di Indonesia, jumlah penggunaan barang elektronik, bahan bakar, dan kendaraan bermotor terus meningkat, yang berkontribusi terhadap penyebaran emisi gas berbahaya. DKI Jakarta, sebagai salah satu kota besar di Indonesia, mengalami penyebaran konsentrasi zat polutan udara, terutama di area dengan kepadatan lalu lintas tinggi dan aktivitas industri [1]. Penyebaran zat polutan yang membuat turunnya kualitas udara dapat menyebabkan masalah kesehatan pada masyarakat seperti penyakit pernapasan [1]. Oleh karena itu, diperlukan sebuah sistem pemantauan dan prediksi polusi udara yang lebih akurat untuk membantu mitigasi risiko kesehatan dan perumusan kebijakan lingkungan.

Dalam upaya mitigasi dampak polusi udara, model prediksi berbasis kecerdasan buatan menjadi solusi yang dapat diandalkan. Model prediksi yang akurat dapat membantu pemerintah dan masyarakat dalam mengantisipasi lonjakan polusi udara, mengatur kebijakan lalu lintas, dan merencanakan tindakan preventif seperti pembatasan kendaraan saat kondisi polusi tinggi [2]. Model berbasis pembelajaran mesin telah banyak diterapkan dalam prediksi konsentrasi polutan, dengan pendekatan berbasis jaringan saraf tiruan dan metode statistik tradisional seperti regresi [3]. Namun, banyak dari model ini masih memiliki keterbatasan dalam menangkap pola *non-linear* yang kompleks dalam data polusi udara, yang sangat dipengaruhi oleh faktor-faktor seperti kondisi cuaca, emisi kendaraan, dan aktivitas industri. Oleh karena itu, diperlukan pendekatan yang solutif untuk meningkatkan akurasi prediksi dan memahami karakteristik data polusi dengan lebih baik.

Metode yang memiliki potensi besar dalam prediksi polusi udara adalah *Long Short-Term Memory* (LSTM). Model ini merupakan pengembangan dari RNN yang dirancang untuk mengatasi masalah *vanishing gradient* [4]. Keunggulan utama LSTM adalah kemampuannya dalam menangani ketergantungan jangka panjang dan menangkap pola *non-linear* dalam data polusi udara [3]. Sebuah penelitian telah menunjukkan bahwa LSTM memiliki keunggulan dibandingkan metode lain seperti

ANN dan SVM dalam memprediksi konsentrasi polutan udara [5]. Namun, efektivitas model LSTM sangat bergantung pada pemilihan *hyperparameter* yang optimal, sehingga diperlukan metode optimasi yang efisien untuk meningkatkan performa prediksi.

Salah satu metode yang sering digunakan untuk optimasi *hyperparameter* adalah *Bayesian Optimization*. *Bayesian Optimization* adalah teknik optimasi yang efisien dan sangat cocok untuk menemukan kombinasi parameter optimal. Metode ini menggunakan model aproksimasi untuk memperkirakan hasil dari berbagai kombinasi parameter, sehingga dapat memilih kombinasi parameter yang paling memungkinkan untuk meningkatkan hasil pada setiap iterasi [6]. Dalam sebuah penelitian, penggunaan metode *Bayesian Optimization* terbukti dapat meningkatkan performa prediksi model dibandingkan dengan pemilihan *hyperparameter* secara manual [7].

Untuk meningkatkan performa model LSTM, data *time series* polusi udara diolah dengan teknik dekomposisi sinyal seperti *Complete Ensemble Empirical Mode Decomposition with Adaptive Noise* (CEEMDAN). Teknik ini memungkinkan pemisahan data polusi menjadi beberapa komponen *Intrinsic Mode Functions* (IMF) yang mewakili frekuensi berbeda, sehingga memudahkan model dalam menangkap pola data yang tersembunyi [8]. Dengan mendekomposisi data sebelum digunakan dalam model LSTM, informasi yang diekstraksi menjadi lebih terstruktur, mengurangi *noise*, dan membantu model dalam memahami tren polusi udara secara lebih akurat [9]. Penerapan dekomposisi CEEMDAN telah terbukti meningkatkan akurasi prediksi data *time series* seperti pada penelitian yang dilakukan oleh Cui, dkk [10].

Berdasarkan berbagai penelitian dan analisis yang telah dilakukan, *Long Short-Term Memory* (LSTM) terbukti efektif dalam memprediksi polutan udara pada data deret waktu. Selain itu, CEEMDAN dapat digunakan untuk mendekomposisi data menjadi komponen yang lebih informatif, sementara *Bayesian Optimization* dapat mengoptimalkan *hyperparameter* model guna meningkatkan akurasi prediksi. Oleh karena itu, penelitian ini berfokus pada "PENERAPAN MODEL CEEMDAN-LSTM DENGAN OPTIMASI BAYESIAN DALAM PREDIKSI INDEKS STANDAR PENCEMARAN UDARA DI DKI JAKARTA". Data yang digunakan dalam penelitian ini berasal dari SPKU DKI Jakarta, yang menyediakan variabel ISPU. Data ini akan diolah menggunakan CEEMDAN untuk mengekstraksi fitur yang lebih

representatif, yang kemudian akan digunakan sebagai input dalam model LSTM. Proses optimasi *hyperparameter* model dilakukan dengan *Bayesian Optimization* untuk meningkatkan performa prediksi. Hasil prediksi akan dievaluasi menggunakan metrik evaluasi *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan *Root Mean Squared Error* (RMSE). Diharapkan metode ini mampu meningkatkan akurasi prediksi indeks standar pencemaran udara dengan lebih adaptif terhadap perubahan pola yang dinamis dan sulit diprediksi.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, perumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa model CEEMDAN-LSTM dalam prediksi indeks standar pencemaran udara?
2. Bagaimana performa model CEEMDAN-LSTM dengan optimasi Bayesian dalam prediksi indeks standar pencemaran udara?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan, tujuan dalam penelitian ini adalah sebagai berikut:

1. Mengetahui performa model CEEMDAN-LSTM terhadap prediksi indeks standar pencemaran udara.
2. Mengetahui performa model CEEMDAN-LSTM dengan optimasi Bayesian dalam prediksi indeks standar pencemaran udara.

1.4. Manfaat Penelitian

Adapun manfaat yang diharapkan dari dilakukannya penelitian ini adalah sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode prediksi polusi udara dengan pendekatan dekomposisi dan *machine learning*.
2. Penelitian ini diharapkan dapat menambah wawasan tentang pengaruh dekomposisi CEEMDAN dan optimasi Bayesian terhadap performa model LSTM dalam analisis data polusi udara.

1.5. Batasan Penelitian

Batasan penelitian digunakan untuk memperjelas ruang lingkup dan fokus penelitian agar tidak melebar ke hal-hal yang tidak relevan. Berikut merupakan batasan dalam penelitian ini:

1. Penelitian ini dibatasi pada penggunaan data harian ISPU periode 2010 hingga 2025 yang diambil dari data.go.id.
2. Variabel yang diprediksi adalah nilai ISPU.
3. Metode yang digunakan dalam penelitian ini adalah CEEMDAN-LSTM dengan optimasi Bayesian.