

Kualitas udara yang buruk di wilayah perkotaan seperti Kota Yogyakarta telah menjadi perhatian serius karena berdampak langsung terhadap kesehatan masyarakat dan lingkungan. Untuk mengurangi dampak tersebut, diperlukan regulasi dan kebijakan pengendalian polusi udara yang didukung oleh sebuah sistem pendukung keputusan berbasis data dengan pendekatan prediktif. Sayangnya, penelitian-penelitian yang telah dilakukan sebelumnya di Kota Yogyakarta belum berhasil memberikan model prediksi kualitas udara yang cukup akurat.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sebuah model peramalan konsentrasi *particulate matter* dengan ukuran lebih kecil dari $2,5 \mu\text{m}$ ($\text{PM}_{2,5}$) yang merupakan indikator polusi udara utama. Model peramalan dikembangkan dengan menggunakan pendekatan *deep learning* Long Short-Term Memory (LSTM) yang merupakan salah satu model peramalan terbaik saat ini. Penelitian ini juga bertujuan untuk menemukan data eksternal sebagai fitur tambahan yang dapat meningkatkan lebih jauh performa model yang dikembangkan.

Data yang digunakan merupakan data deret waktu dengan resolusi harian yang dikumpulkan dari berbagai sumber tepercaya seperti BMKG dan DLH Kota Yogyakarta. Sebelum digunakan dalam pelatihan model, data melalui tahap pengolahan awal yang meliputi deteksi dan penanganan nilai anomali, imputasi *missing values*, serta normalisasi untuk meningkatkan stabilitas proses pelatihan model. Model LSTM dikembangkan dan dioptimalkan untuk mendapat performa terbaik, bersama dengan model CNN dan CNN-LSTM sebagai *baseline* model *deep learning*. Model LSTM hasil pengembangan kemudian diuji dengan fitur tambahan dari data eksternal untuk melihat pengaruh dari tiap data pada akurasi model.

Model dievaluasi menggunakan metrik Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), dan koefisien determinasi (R^2). Hasil eksperimen menunjukkan bahwa model LSTM dengan nilai RMSE sebesar 6,1832; skor MAPE sebesar 22,71%; dan skor R^2 sebesar 0,5007 berhasil mengungguli model CNN dan CNN-LSTM, serta model ANFIS dan Prophet dari penelitian terdahulu. Sementara itu, penggunaan data eksternal berupa *day of week* berhasil meningkatkan performa model LSTM hingga mencapai skor RMSE sebesar 6,0997; skor MAPE sebesar 23,11%; dan nilai R^2 sebesar 0,5141.

Penelitian di Kota Yogyakarta ini diharapkan dapat berkontribusi menjadi referensi pada pengembangan metode peramalan kualitas udara berbasis *deep learning*, terutama LSTM. Selain itu, model yang dikembangkan dapat diadaptasi lebih lanjut untuk implementasi *real-time* dan perluasan ke parameter kualitas udara lainnya di masa depan. Penelitian ini juga berhasil memberikan wawasan yang lebih dalam untuk penelitian serupa selanjutnya, termasuk penelitian lain di lokasi selain Kota Yogyakarta.

Kata kunci : kualitas udara, $\text{PM}_{2,5}$, peramalan, LSTM, fitur eksternal

Poor air quality in urban areas such as Yogyakarta City has become a serious concern as it has a direct impact on public health and the environment. To reduce these impacts, air pollution control regulations and policies are needed, supported by a data-based decision support system with a predictive approach. Unfortunately, previous studies conducted in Yogyakarta City have not succeeded in providing a sufficiently accurate air quality prediction model. .

Therefore, this study aims to develop a forecasting model for the concentration of particulate matter smaller than $2.5 \mu\text{m}$ ($\text{PM}_{2.5}$) which is the main air pollution indicator. The forecasting model was developed using the deep learning Long Short-Term Memory (LSTM) approach, which is one of the best forecasting model right now. This research also aims to find external data as additional features that can improve the performance of the developed model further.

The data used are time series data with daily resolution collected from various reliable sources such as Yogyakarta City meteorology and environment agency. Before being used in model training, the data goes through a pre-processing stage that includes anomaly detection and handling, imputation, and normalization to improve the stability of the model training process. The LSTM model will be developed and optimized to get the best performance, along with CNN and CNN-LSTM models as the deep learning baseline. The developed LSTM model is then tested with additional features from external data to see the influence of each data on the accuracy of the model.

The models were evaluated using the Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2) metrics. The experimental results show that the LSTM model with an RMSE score of 6.1832, a MAPE score of 22.71%, and an R^2 score of 0.5007 successfully outperformed the CNN and CNN-LSTM models, as well as ANFIS and Prophet models from previous studies. Meanwhile, the use of external data in the form of day of week managed to improve the performance of the LSTM model to achieve an RMSE score of 6.0997, a MAPE score of 23.11%, and an R^2 score of 0.5141.

This research in Yogyakarta City is expected to contribute as a reference to the development of deep learning-based air quality forecasting methods, especially LSTM. In addition, the developed model can be further adapted for real-time implementation and extension to other air quality parameters in the future. This research also succeeded in providing deeper insights for further similar research, including other research in locations other than Yogyakarta City.

Keywords : air quality, $\text{PM}_{2.5}$, forecasting, CNN-LSTM, external features