

**AIR QUALITY PREDICTION IN JAKARTA
USING LSTM AND GRU MODELS
FOR INFORMATION DISSEMINATION**

Written By

Ni Putu Nita Nathalia, NIM 2215101050

Department of Informatics Engineering

ABSTRACT

Recent time air quality has deteriorated due to the rise in industrial activities, transportation, and urbanization particularly in DKI Jakarta. This study aims to understand the performance of each model in predicting air quality and identify the most suitable model. The deep learning models employed in this research are Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). The methodology follows the CRISP-DM framework. However, this study is limited to the Evaluation stage. The dataset used covers a one-year period and was obtained from five selected monitoring stations: DKI1-Bundaran HI, DKI2-Kelapa Gading, DKI3-Jagakarsa, DKI4-Lubang Buaya, and DKI5-Kebon Jeruk. The training and validation processes were carried out using Time Series Cross-Validation ($k=5$). The model architecture comprises two hidden layers, two dropout layers, and one input and one output layer. A total of 10 models were trained (5 LSTM models and 5 GRU models), with the optimal timestep determined for each model individually. The initial experiments indicated that both LSTM and GRU models faced challenges in predicting certain air pollutant components at specific stations. Therefore, a second round of experiments involving re-training and re-testing was conducted. Based on the final inputs and evaluation, it was found that, for this particular case, the GRU model outperformed the LSTM model.

Keywords: air pollution, LSTM, GRU, CRISP-DM

PREDIKSI KUALITAS UDARA DI JAKARTA MENGUNAKAN MODEL LSTM DAN GRU UNTUK DISEMINASI INFORMASI

Oleh

Ni Putu Nita Nathalia, NIM 2215101050

Jurusan Teknik Informatika

ABSTRAK

Dalam beberapa waktu terakhir, kualitas udara mengalami penurunan akibat meningkatnya aktivitas industri, transportasi, dan urbanisasi, khususnya di DKI Jakarta. Penelitian ini bertujuan untuk memahami kinerja masing-masing model dalam memprediksi kualitas udara serta mengidentifikasi model yang paling sesuai. Model deep learning yang digunakan dalam penelitian ini adalah Long Short-Term Memory (LSTM) dan Gated Recurrent Unit (GRU). Metodologi penelitian mengikuti kerangka kerja CRISP-DM, namun penelitian ini dibatasi hingga tahap evaluasi. Dataset yang digunakan mencakup periode satu tahun dan diperoleh dari lima stasiun pemantauan terpilih, yaitu DKI1-Bundaran HI, DKI2-Kelapa Gading, DKI3-Jagakarsa, DKI4-Lubang Buaya, dan DKI5-Kebon Jeruk. Proses pelatihan dan validasi dilakukan menggunakan *Time Series Cross-Validation* ($k=5$). Arsitektur model terdiri atas dua *hidden layer*, dua *dropout layer*, satu *input layer*, dan satu *output layer*. Sebanyak 10 model dilatih (5 model LSTM dan 5 model GRU), dengan timestep optimal ditentukan secara individual untuk setiap model. Hasil eksperimen awal menunjukkan bahwa baik model LSTM maupun GRU menghadapi tantangan dalam memprediksi beberapa komponen polutan udara pada stasiun tertentu. Oleh karena itu, dilakukan eksperimen tahap kedua berupa pelatihan ulang dan pengujian ulang. Berdasarkan input dan hasil evaluasi akhir, ditemukan bahwa pada kasus ini model GRU menunjukkan kinerja yang lebih baik dibandingkan dengan model LSTM.

Kata kunci: kualitas udara, LSTM, GRU, CRISP-DM