

Perancangan sistem prediksi kandungan polutan PM2.5 di Jakarta dengan metode RNN-LSTM dan ARIMA = Design of PM2.5 content prediction system in Jakarta with RNN-LSTM and ARIMA

Achmad Faiz Siraj, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20516230&lokasi=lokal>

Abstrak

PM2.5 merupakan salah satu penyebab tingginya angka polusi di Jakarta. Skripsi ini akan membahas penerapan Recurrent Neural Network jenis Long Short-Term Memory (RNN-LSTM) dan Autoregressive Integrated Moving Average (ARIMA), dua metode yang dapat digunakan untuk melakukan prediksi pada dataset jenis time series, sebagai algoritma untuk melakukan prediksi pada kandungan polutan PM2.5 di Jakarta. Terdapat dua jenis preprocessing yang diujicoba pada pengujian ini, yaitu dengan imputation menggunakan mean dan linear interpolation. Saat pembuatan model pada ARIMA, dilakukan pengaturan order untuk mencari model terbaik yang dapat melakukan prediksi dengan akurasi tertinggi. Sementara untuk RNN-LSTM, pencarian model terbaik dilakukan dengan melakukan serangkaian ujicoba dengan perubahan pada beberapa parameter seperti ukuran dari rolling window, batch size, dan optimizer. Berdasarkan hasil akurasi, didapatkan model dengan ARIMA order (2,0,1) sebagai model paling baik ketika dilakukan ujicoba dengan imputation jenis mean dengan RMSE sebesar 17,84. Lebih baik dari hasil yang didapatkan RNN-LSTM pada metode imputation tersebut yang hanya mendapat RMSE 18,00. Namun RNN-LSTM memiliki hasil akurasi yang lebih baik ketika dilakukan ujicoba dengan metode imputation dengan linear interpolation dimana RMSE yang didapatkan sebesar 17,47. Lebih baik dari ARIMA yang hanya mendapat RMSE sebesar 17,66.

.....PM2.5 is one of the causes of Jakarta's high pollution level. This thesis will discuss the implementation of Recurrent Neural Network type Long Short-Term Memory (RNN-LSTM) and Autoregressive Integrated Moving Average (ARIMA), two algorithm that are able to predict a time series dataset, as two algorithms used to do a forecasting in PM2.5 pollutant level in Jakarta. There are two preprocessing used in this test, mean imputation and linear interpolation. In ARIMA, tweaking to find model with best accuracy was done by altering its order. While in RNN-LSTM, the search for the best model was done by tweaking several parameters such as the size of its rolling window, batch size, and optimizer. Based on its accuracy, an ARIMA model with order of (2,0,1) was found as the best model during the test with mean imputation with RMSE of 17,84 compared to RNN-LSTM's 18,00. But RNNLSTM has better accuracy when tested with linear interpolation, where it got RMSE of 17,47. Where ARIMA only has RMSE of 17,66.