



APD-BayTM: Jakarta Air Quality Index Prediction using Bayesian Optimized LSTM

Raey Faldo^{1,2}, Satria Mandala^{1,2}, Mohd Shahrizal Sunar³

¹ Human Centric (HUMIC) Engineering, Telkom University, Bandung, Indonesia

² School of Computing, Telkom University Bandung, 40257, Indonesia²

³ MaGIC-X (Media and Games Innovation Centre of Excellence), Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

ARTICLE INFORMATION

Received: February 00, 00
 Revised: March 00, 00
 Accepted: March 00, 00
 Available online: April 00, 00

KEYWORDS

Deep learning, LSTM, air pollution, air quality index (AQI)

CORRESPONDENCE

Email:satriamandala@telkomuniversity.ac.id

A B S T R A C T

The Air Quality Index (AQI) is a metric for evaluating air quality in a region. Jakarta holds the fifth position globally in terms of air pollution. Several studies have been performed to forecast pollution levels in Jakarta. However, existing studies exhibit limitations such as outdated datasets, lack of data normalization, absence of machine learning parameter setting, neglect of k-fold cross-validation, and a failure to incorporate deep learning algorithms for pollution detection. This study introduces an air quality detection system called APD-BayTM to address these issues. This proposed system leverages Long Short-Term Memory (LSTM) and uses Bayesian Optimization (BO) to enhance the performance of air pollution detection. The methodology used in this research involves four key steps: data preprocessing, LSTM model development, hyperparameter tuning through BO, and performance assessment using 5-fold cross-validation. APD-BayTM exhibits robust performance, comparable to previous research outcomes. The LSTM model in APD-BayTM on the training dataset achieved average precision, recall, F1 score, and accuracy values of 93.29%, 91.41%, 91.89%, and 95.90%, respectively. These metrics improved on the test dataset, reaching 97.44%, 99.71%, 98.52%, and 99.34%, respectively. These findings show the robustness of APD-BayTM across datasets of varying sizes, encompassing both large and small datasets.