Environmental Engineering and Management Journal

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
http://www.eemj.icpm.tuiasi.ro/; http://www.eemj.eu

"Gheorghe Asachi" Technical University of Iasi, Romania

UNCOVERING THE RELATIONSHIP BETWEEN TRAFFIC AND EMISSION USING MACHINE LEARNING: A CASE STUDY OF JAKARTA

Fauzi Dwi Setiawan[1,2], Rizqon Fajar[2], Junghyun Kim[3*]

[1]*Department of Global Development and Entrepreneurship, Handong Global University, Pohang, South Korea*
[2]*Department of Research Center for Transportation Technology, National Research and Innovation Agency, South Tangerang, Indonesia*
[3]*School of Aerospace Engineering, Sejong University, Seoul, South Korea*

Abstract

This research proposes a data-driven methodology that leverages machine learning algorithms to uncover the relationship between traffic and emissions, with a focus on Jakarta, Indonesia. The proposed methodology follows five key steps. First, traffic and emission datasets are collected from the Jakarta Transportation Department and the Ministry of Environment and Forestry. Next, performing data preprocessing, including outlier detection, missing data handling, and feature selection to improve model performance. A machine learning-based regression model is developed to identify the relationship between traffic (input) and emissions (output). This is followed by the implementation of a Long Short-Term Memory (LSTM) model to forecast future traffic and emission data, which integrates with the machine learning-based regression model. Finally, the model performance is evaluated. The results indicate that the Random Forest regression model, after feature selection and hyperparameter tuning, achieved an R^2 of 0.81, effectively identifying key traffic variables such as vehicle speed, bus/truck count, and car count as significant contributors to PM10 emissions. Additionally, the LSTM time-series model demonstrated strong predictive capabilities with an R^2 of 0.78, accurately forecasting future PM10 levels based on temporal trends. These findings highlight the potential of machine learning in supporting data-driven air quality management, enabling policymakers to design more effective traffic control and pollution mitigation strategies. By combining real-time emission estimation with short-term forecasting, this methodology provides predictive insights into both current and future pollution levels. This framework can serve as a model for sustainable urban mobility planning, applicable not only in Jakarta but also in other cities facing similar air pollution challenges.

Key words: emissions, long short-term memory, machine learning, random forest, traffic congestion

Received: April, 2025; Revised final: January, 2026; Accepted: February, 2026

* Author to whom all correspondence should be addressed: E-mail: junghyun.kim@sejong.ac.kr