

A Comparative Analysis of Machine Learning Techniques for PM10 Forecasting in Vilnius

Authors : Mina Adel Shokry Fahim, Jūratė Sužiedelytė Visockienė

Abstract : With the growing concern over air pollution (AP), it is clear that this has gained more prominence than ever before. The level of consciousness has increased and a sense of knowledge now has to be forwarded as a duty by those enlightened enough to disseminate it to others. This realisation often comes after an understanding of how poor air quality indices (AQI) damage human health. The study focuses on assessing air pollution prediction models specifically for Lithuania, addressing a substantial need for empirical research within the region. Concentrating on Vilnius, it specifically examines particulate matter concentrations 10 micrometers or less in diameter (PM10). Utilizing Gaussian Process Regression (GPR) and Regression Tree Ensemble, and Regression Tree methodologies, predictive forecasting models are validated and tested using hourly data from January 2020 to December 2022. The study explores the classification of AP data into anthropogenic and natural sources, the impact of AP on human health, and its connection to cardiovascular diseases. The study revealed varying levels of accuracy among the models, with GPR achieving the highest accuracy, indicated by an RMSE of 4.14 in validation and 3.89 in testing.

Keywords : air pollution, anthropogenic and natural sources, machine learning, Gaussian process regression, tree ensemble, forecasting models, particulate matter

Conference Title : ICEMSR 2024 : International Conference on Environmental Monitoring, Simulation and Remediation

Conference Location : Prague, Czechia

Conference Dates : July 04-05, 2024