BAYESIAN OPTIMIZATION IN GAUSSIAN PROCESS REGRESSION FOR ACCURATE AIR QUALITY PREDICTION.

Karlavi M.M.[1] and Chatrabgoun O.[2]

[1]Department of Data Science and Computational Intelligence, School of Computing, Mathematics and Data Science, Faculty of Engineering, Coventry University, UK

[2]School of Computing, Mathematics and Data Science Faculty of Engineering, Environment and Computing Coventry University, Gulson Road, Coventry, CV1 2JH, UK

Air pollution, especially nitrogen dioxide (NO_2), poses significant risks to public health and environmental quality. Traditional statistical models for predicting NO_2 levels often fail to capture the complex, nonlinear relationships in environmental data and typically do not provide uncertainty estimates. This study addresses these shortcomings by utilizing advanced machine learning techniques, specifically Gaussian Process Regression (GPR), to enhance the accuracy and reliability of NO_2 predictions. Using a comprehensive dataset of hourly averaged responses from chemical sensors in an urban area in Italy, we developed and evaluated GPR models with various kernels, including polynomial-like, rational quadratic, and combined kernels. These models were compared with traditional regression models, such as Lasso and Ridge regression. The results showed that GPR models, particularly those with optimized polynomial-like and rational quadratic kernels, significantly outperformed the traditional models. The polynomial-like kernel GPR model achieved a Mean Squared Error (MSE) of 0.034 and an R-squared (R^2) value of 0.959. Similarly, the rational quadratic kernel GPR model achieved an MSE of 0.035 and an R^2 of 0.959. In contrast, Lasso and Ridge regression models had higher MSEs and lower R^2 values. Additionally, the GPR models provided valuable uncertainty estimates, enhancing prediction reliability. This study demonstrates the effectiveness of GPR models in environmental monitoring and underscores the importance of kernel optimization in improving model performance, suggesting substantial potential for GPR in air quality prediction and environmental management.

Keywords: *Air Quality Prediction, Gaussian Process Regression (GPR), Lasso Regression, Predictive Modeling, Ridge Regression.*