

Abstract

Air and water pollution is the substantial health threat globally and particularly in India as it plays major role in rendering the death and disease. Besides adversely affecting the health and reducing the lifespan, air and water pollution also mitigates the economic productivity of the country. It has also been observed that the adverse effect of air and water pollution on health is increasing in India and several studies have been conducted in this regard. Indeed, prediction of the mortality rate can be considered as the problem related to large data and this prediction is of the varying nature because it depends on the factors that influence it such as air and water pollution. Researchers have been implemented many machine learning techniques such as Random Forest and Extreme Gradient Boosting (XGBoost) for predicting the mortality rate due to air pollution. Different machine learning techniques have also been used for predicting the water quality like artificial neural network (ANN), group method of data handling (GMDH) and support vector machine (SVM). In this report, I will develop and implement an ensemble model of Random Forest and Linear Regression for this pursuit. This model will predict the mortality rate collectively due to air and water pollution.