Abstract

Instances of fatalities along with diseases due to air pollution turn out to be growing every year. WHO's 2018 report states that around six to seven million people die each and every year as a result of diseases attributable to air contamination. Thus it is extremely important to precisely monitor and anticipate the grades of contaminants in the surrounding air. In this project, we attempt to find the AQI limits by utilizing ML strategies like, Support Vector Regression (SVR), Decision Tree Regression (DTR), Multiple Linear Regression (MLR) and Random Forest Regression (RFR). Then we compare these ML models by judging them based on error metrics such as, Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Root Mean Square Logarithmic Error (RMSLE). The preliminary outcomes showed that the performance shown by SVR was bad. MLR and DTR both performed acceptably well. RFR performed the best among every other regression examples.

Performances

- Results on training set:

models	R^2	RMSE	MAE	RMSLE
MLR	0.9965	5.9334	3.2952	0.0595
Decision Tree	1.0000	0.0000	0.0000	0.0000
Random Forest	0.9996	2.0237	0.7106	0.0195
SVR	0.9494	22.628	16.076	0.1423

- Results on testing set:

Models	R^2	RMSE	MAE	RMSLE
MLR	0.9965	5.4973	3.4796	0.0517
Decision Tree	0.9955	6.2370	2.354	0.0563
Random Forest	0.9982	3.8577	1.7016	0.0422
SVR	0.9164	27.0025	19.0722	0.1686

Prediction results