

## ABSTRACT

Air, an essential natural resource, has been compromised in terms of quality by economic activities. The Environmental Protection Agency tracks the air pollution with commonly known criteria pollutants, i.e., ground-level ozone, SO<sub>2</sub>, particulates matter like PM<sub>10</sub> and PM<sub>2.5</sub>, CO, CO<sub>2</sub>, and Nitrogen dioxide. These substances are in compositions of a common index, called the Air Quality Index. As the AQI increases, a higher percentage of the population is exposed. We forecast the air quality of India by using machine learning to predict the AQI of a given area. AQI of India is a standard measure used to indicate the pollutant like SO<sub>2</sub>, NO<sub>2</sub>.etc levels over a period. We developed a model to predict the AQI based on historical data of previous years and predicting over a particular upcoming year as a Gradient decent boosted multivariable regression problem. We improve the efficiency of the model by applying cost estimation for our predictive problem. Our model will be capable for successfully predicting AQI of a total county provided with the historical data of pollutant concentration. In our model by implementing the proposed parameter-reducing formulations, we achieved better performance than the standard regression models. Our model has 96% accuracy on predicting the current available dataset on predicting the AQI of whole India by using XGBoost Algorithm. Thus, the essential purpose of our system is to warn the decision making authorities and to prevent the citizens from dangerous impact of air pollution beforehand.