# Prediction of Different Air Pollution Parameters using Machine Learning Approach

Paragkanti Chattopadhyay[1], Sumit Banerjee[2], Chandan Kumar Chanda[3], Sourav Bhattacharya[4]
1Department of Computer Science and Engineering, Dr. B. C. Roy Engineering College, Durgapur
2 Department of Electrical Engineering, Dr. B. C. Roy Engineering College, Durgapur
3Department of Electrical Engineering, IIEST Shibpur
4Department of Basic Science and Humanities, Dr. B. C. Roy Engineering College, Durgapur

**Abstract:** This paper analyses air pollution data focusing on pollutants such as PM2.5,PM10, Sulpher dioxide(SO2 ) and Nitrogen dioxide (NO2) levels using a CSV dataset. Initially, the data is pre-processed by cleaning it to handle missing values and outliers, ensuring consistency and accuracy. The dataset was then divided into training and testing subsets. For modelling, polynomial regression will be employed, which allowed to capture the nonlinear relationships between PM2.5, PM10, Sulpher dioxide(SO2 ) and Nitrogen dioxide (NO2) and other relevant variables. To balance model accuracy and complexity and avoid over fitting, the degree of the polynomial will be optimized. The model was trained on the processed data and evaluated using metrics such as the R2 score and Root Mean Square Error (RMSE) to assess pollutants label of PM2.5, PM10, Sulpher dioxide(SO2 ) and Nitrogen dioxide (NO2)of Kolkata with some predefined threshold values based on past experience. Finally, the above analysis will be interpreted to understand the patterns and factors influencing PM2.5, PM10, Sulpher di-oxide (SO2) and Nitrogen dioxide (NO2) pollution levels and the results are found to be in very good agreement.

Keywords: Air Pollution, PM2.5,PM10,Sulpher dioxide(SO2 ),Nitrogen dioxide (NO2), Machine Learning, Polynomial Regression.

## 1. Introduction

Air pollution is a pressing environmental and public health concern, particularly in rapidly urbanizing regions of the world. In India, the effects of industrial expansion, increasing vehicular emissions, and population growth have led to a significant decline in air quality across major cities. Kolkata, the capital of West Bengal, has consistently ranked among the country's most polluted urban centers. The city faces elevated levels of key air pollutants such as Particulate Matter (PM2.5 and PM10), Sulfur Dioxide ($SO_2$), and Nitrogen Dioxide ($NO_2$), all of which pose serious health risks including respiratory and cardiovascular illnesses, as well as contributing to broader ecological damage.

Given the long-term implications of sustained air pollution, accurately forecasting future pollutant levels is vital for informed decision-making and policy development. This study focuses on predicting the concentrations of PM2.5, PM10, $SO_2$, and $NO_2$ in Kolkata for the year 2027, based on historical air quality data from 2017 to 2024. A machine learning approach has been adopted, utilizing a linear regression model to identify patterns and trends