ABSTRACT

Oxygen is one of the primary needs for humans to survive. The Thing that can contain the oxygen we breathe is air pollution. Data on air quality is very important for humans, especially on air pollution in the future. The research aims for humans to be able to make wise and unconscious decisions after seeing data from air quality predictions. In this study, the pollution data that will be examined are 10 micron particulates, sulfur dioxide, carbon monoxide, ozone, and nitrogen dioxide.

The data to be studied is air quality data in five areas in DKI Jakarta in 2017-2021, namely DKI1 (Bunderan HI), DKI2 (Kelapa Gading), DKI3 (Jagakarsa), DKI4 (Lubang Buaya), and DKI5 (Kebon Jeruk). The air quality data used is the Air Pollution Standard Index (ISPU) data obtained from the Jakarta Open Data website portal. Due to some missing data, the K-Nearest Neighbor (KNN) Imputer method is used to fill in the missing data. The prediction method used is Gaussian Process Regression (GPR).

Gaussian process regression method is used because it has many kernels that can be tested for each data. The predictions that have been made will be implemented into the website as a visualization using a streamlit framework. The test results in each region and the best substances according to the size of the test data partition and the kernel used are DKI1 having an R2 range of 0.566 to 0.887, DKI2 having an R2 range of 0.322 to 0.893, DKI3 having an R2 range of 0.457 to 0.914, DKI4 having an R2 range of 0.534 to 0.929, and DKI5 has an R2 range of 0.623 to 0.894.

Keywords: *Air Pollution, Data Prediction, K-Nearest Neighbor Imputer, Gaussian Process Regression, Filling missing data, Web*