

ABSTRAK

Oksigen adalah salah satu kebutuhan primer bagi manusia untuk bertahan hidup. Hal yang dapat mencemari oksigen yang kita hirup adalah polusi udara. Data tentang kualitas udara sangat penting untuk manusia, terutama pada polusi udara di masa depan. Penelitian bertujuan untuk manusia agar bisa mengambil keputusan yang bijak dan sadar setelah melihat data hasil prediksi kualitas udara. Pada penelitian ini, polusi yang akan dilihat datanya yaitu partikulat 10 mikron (PM_{10}), sulfur dioksida (SO_2), karbon monoksida (CO), ozon (O_3), dan nitrogen dioksida (NO_2).

Data yang akan diteliti adalah data kualitas udara di 5 daerah di DKI Jakarta pada tahun 2017-2021, yaitu DKI1 (Bunderan HI), DKI2 (Kelapa Gading), DKI3 (Jagakarsa), DKI4 (Lubang Buaya), dan DKI5 (Kebon Jeruk). Data kualitas udara yang digunakan yaitu data Indeks Standar Pencemaran Udara (ISPU) yang didapat dari portal *website* Jakarta *Open Data*. Dikarenakan ada beberapa data yang hilang, maka digunakan metode *K-Nearest Neighbor* (KNN) Imputer untuk melakukan pengisian data yang hilang. Metode prediksi yang digunakan yaitu *Gaussian Process Regression* (GPR).

Metode *gaussian process regression* digunakan karena memiliki banyak *kernel* yang bisa diuji untuk setiap data. Prediksi yang telah dilakukan akan diimplementasikan ke dalam *website* sebagai visualisasi dengan menggunakan kerangka *streamlit*. Hasil pengujian di tiap daerah dan zat yang paling baik menurut ukuran partisi data uji dan *kernel* yang digunakan yaitu DKI1 memiliki rentang R^2 0.566 sampai 0.887, DKI2 memiliki rentang R^2 0.322 sampai 0.893, DKI3 memiliki rentang R^2 0.457 sampai 0.914, DKI4 memiliki rentang R^2 0.534 sampai 0.929, dan DKI5 memiliki rentang R^2 0.623 sampai 0.894.

Kata Kunci: *Polusi Udara, Prediksi Data, K-Nearest Neighbor Imputer, Gaussian Process Regression, Pengisian Data yang Hilang, Web*