

PERBANDINGAN KINERJA *RANDOM FOREST* DAN *XGBOOST* DALAM PREDIKSI ISPU BULANAN DI JAKARTA

Rhosidah Hannum, (2022-31-037)

Dibawah bimbingan Dr. Meilia Nur Indah Susanti, S.T., M.Kom

ABSTRAK

Polusi udara di Jakarta telah mencapai tingkat mengkhawatirkan dengan nilai ISPU yang sering berada pada kategori sedang hingga tidak sehat. Penelitian ini membandingkan kinerja *Random Forest Regressor* dan *XGBoost Regressor* dalam memprediksi nilai ISPU maksimum bulanan di Jakarta menggunakan pendekatan regresi. Data bulanan periode 2020–2022 dari lima stasiun pemantauan mencakup 12 variabel, yaitu enam parameter polutan (PM2.5, PM10, SO₂, CO, O₃, NO₂) dan enam variabel meteorologi (radiasi matahari, suhu, kelembapan, kecepatan angin, arah angin, curah hujan). Metodologi mengikuti kerangka CRISP-DM dengan pembagian data 70:30 berdasarkan urutan waktu. Evaluasi dilakukan menggunakan metrik RMSE, MAE, dan R². Hasil menunjukkan *XGBoost* unggul pada empat dari lima stasiun, dengan performa terbaik pada stasiun DKI1 (Bundaran HI) yang mencatat RMSE sebesar 2,148, MAE sebesar 1,927, dan R² sebesar 0,938. Sementara itu, *Random Forest* menunjukkan kinerja lebih baik pada stasiun DKI4 (Lubang Buaya) yang memiliki karakteristik data lebih fluktuatif, dengan RMSE sebesar 7,161, MAE sebesar 5,662, dan R² sebesar 0,564. Uji *Wilcoxon Signed Ranks Test* menunjukkan tidak terdapat perbedaan yang signifikan secara statistik antara kedua model ($p > 0,05$), namun *XGBoost* cenderung lebih akurat pada mayoritas stasiun. Analisis *feature importance* mengonfirmasi PM2.5 sebagai variabel paling dominan dengan skor 0,859 pada *Random Forest* dan 0,507 pada *XGBoost*, diikuti PM10 dan beberapa variabel meteorologi melalui mekanisme interaksi non-linear.

Kata kunci: ISPU, *Random Forest Regressor*, *XGBoost Regressor*, kualitas udara

PERFORMANCE COMPARISON OF *RANDOM FOREST* AND *XGBOOST* IN MONTHLY ISPU PREDICTION IN JAKARTA

Rhosidah Hannum, (2022-31-037)

Under the Guidance of Dr. Meilia Nur Indah Susanti, S.T., M.Kom

ABSTRACT

Air pollution in Jakarta has reached alarming levels, with the Air Pollution Standard Index (ISPU) frequently falling in the moderate to unhealthy categories. This study compares the performance of Random Forest Regressor and XGBoost Regressor in predicting monthly maximum ISPU values in Jakarta using a regression approach. Monthly data from 2020 to 2022 collected from five monitoring stations encompass 12 variables, comprising six pollutant parameters (PM_{2.5}, PM₁₀, SO₂, CO, O₃, NO₂) and six meteorological variables (solar radiation, temperature, humidity, wind speed, wind direction, and rainfall). The methodology follows the CRISP-DM framework with a 70:30 data split based on chronological order. Model evaluation was conducted using RMSE, MAE, and R² metrics. The results show that XGBoost outperformed Random Forest at four out of five stations, achieving its best performance at DKI1 (Bundaran HI) with an RMSE of 2.148, MAE of 1.927, and R² of 0.938. Meanwhile, Random Forest demonstrated superior performance at DKI4 (Lubang Buaya), which exhibits more fluctuating data characteristics, recording an RMSE of 7.161, MAE of 5.662, and R² of 0.564. The Wilcoxon Signed Ranks Test revealed no statistically significant difference between the two models ($p > 0.05$); however, XGBoost consistently tended to be more accurate across the majority of stations. Feature importance analysis confirmed PM_{2.5} as the most dominant variable, with an importance score of 0.859 in Random Forest and 0.507 in XGBoost, followed by PM₁₀ and several meteorological variables through non-linear interaction mechanisms.

Keywords: air quality, ISPU, *Random Forest Regressor*, *XGBoost Regressor*