

## ABSTRAK

Perubahan iklim yang ditandai dengan peningkatan suhu rata-rata tahunan dipengaruhi oleh berbagai faktor, salah satunya emisi karbon dioksida (CO<sub>2</sub>) dari sektor agri-pangan yang melibatkan banyak jenis emisi dan memiliki ketergantungan temporal. Penelitian ini bertujuan membandingkan kinerja *Random Forest* dan *Catboost* dalam memprediksi peningkatan suhu rata-rata tahunan suatu negara berdasarkan emisi CO<sub>2</sub> sektor agri-pangan dari berbagai negara pada periode 1990-2020. Penelitian dilakukan melalui tiga skenario pengujian, yaitu tanpa *feature engineering*, penambahan *basic features*, serta kombinasi *basic features* dan *time-aware features*, dengan pembagian data menggunakan *time-based split* menjadi data *training* (1990–2012), *validation* (2013–2017), dan *test* (2018–2020) untuk menghindari *temporal leakage*. Pemilihan *hyperparameter* dilakukan menggunakan *GridSearchCV* dengan skema *PredefinedSplit*. Hasil pengujian menunjukkan bahwa penambahan *time-aware features* memberikan peningkatan performa yang paling signifikan dibandingkan dua skenario lainnya. Model *Catboost* dengan parameter optimal *iterations* = 700, *depth* = 6, dan *learning\_rate* = 0.05 menghasilkan performa terbaik dengan nilai R<sup>2</sup> Test sebesar 0.9794, MAE Test sebesar 0.0309°C, dan RMSE Test sebesar 0.0796°C, serta secara konsisten mengungguli *Random Forest*. Hasil penelitian ini menunjukkan bahwa penerapan *Catboost* dan *Random Forest* dengan *feature engineering* berbasis waktu mampu secara efektif mempelajari pola peningkatan suhu rata-rata tahunan suatu negara.

Kata kunci: emisi, prediksi, *feature engineering*, *Random Forest*, *Catboost*

## ABSTRACT

Climate change, as indicated by the continuous increase in annual average temperature, is influenced by various factors, one of which is carbon dioxide (CO<sub>2</sub>) emissions from the agri-food sector, which involve multiple emission sources and exhibit temporal dependency. This study aims to compare the performance of Random Forest and Catboost in predicting the annual average temperature increase of a country based on agri-food CO<sub>2</sub> emissions from multiple countries over the period 1990–2020. The experiments were conducted using three testing scenarios: without feature engineering, with the addition of basic features, and with a combination of basic features and time-aware features. The dataset was split using a time-based approach into training data (1990–2012), validation data (2013–2017), and test data (2018–2020) to prevent temporal leakage. Hyperparameter selection was performed using GridSearchCV with a PredefinedSplit scheme. The results indicate that the inclusion of time-aware features yields the most significant performance improvement compared to the other scenarios. The Catboost model with optimal parameters (iterations = 700, depth = 6, and learning\_rate = 0.05) achieved the best performance, with an R<sup>2</sup> Test value of 0.9794, MAE Test of 0.0309°C, and RMSE Test of 0.0796°C, consistently outperforming Random Forest. Overall, this study demonstrates that the application of Catboost and Random Forest with time-based feature engineering is able to effectively learn patterns of annual average temperature increase at the country level.

Keywords: emissions, prediction, feature engineering, Random Forest, Catboost