Komparasi permodelan regresi berbasis machine learning studi kasus: sistem prediksi kadar fenolik daun bisbul (diospyros discolor willd.) berbasis citra hiperspektral VNIR = Comparison of machine learning-based regression modeling case study: phenolic level prediction system for bisbul leaf (diospyros discolor willd.) based on VNIR hyperspectral image

Naufal Alharits Sadly, author

Abstrak

Sistem prediksi kadar fenolik pada daun Bisbul (Diospyros discolor Willd.) berbasis citra hiperspektral visible and near-infrared (VNIR) terbukti mampu dibuat dan mendapatkan hasil dengan nilai yang baik. Kamera hiperspektral dengan rentang panjang gelombang 400-1000 nm digunakan dalam mengakuisisi citra VNIR pada daun Bisbul. Penelitian ini membahas mengenai komparasi dari beberapa model regresi baru dengan penelitian terdahulu yang diharapkan bisa mendapatkan hasil yang lebih baik dalam memprediksi kadar fenolik pada daun Bisbul. Digunakan tiga model regresi dalam membuat sistem prediksi ini yaitu model Partial Least Square Regression (PLSR), Random Forest, dan XGBoost Regressor. Sistem Prediksi menggunakan PLSR menghasilakan sebesar 3,62 (RMSE test), 0,81 (R2 test), nilai akurasi sebesar 91,3%, dan waktu training 0,27 detik. Sistem Prediksi menggunakan Random Forest tanpa menggunakan seleksi fitur menghasilakan sebesar 4,04 (RMSE test), 0,81 (R2 test), nilai akurasi sebesar 90,86%, dan waktu training 17,81 detik. Sistem Prediksi menggunakan Random Forest dengan seleksi fitur menghasilakan sebesar 3,84 (RMSE test), 0,79 (R2 test), nilai akurasi sebesar 91,31%, dan waktu training 19,05 detik. Sistem Prediksi menggunakan XGBoost Regressor dengan menghasilakan sebesar 3,48 (RMSE test), 0,83 (R2 test), nilai akurasi sebesar 91,1%, dan waktu training 24,9 detik. Performa terbaik dihasilkan oleh model XGBoost Regressor dengan sedikit perbedaan dengan PLSR. Model XGBoost Regressor berhasil meningkatkan performa sebesar 14% pada RMSE dan 2% pada R2 berbanding dengan PLSR.
.....Phenolic levels prediction system on Bisbul leaves (Diospyros discolor Willld.) Based on visible and near-infrared (VNIR) hyperspectral images proved to be able to be made and get results with good values. Hyperspectral camera with a wavelength range of 400-1000 nm is used in acquiring VNIR images on Bisbul leaves. This study discusses the comparison of several new regression models with previous studies that are expected to get better results in predicting phenolic levels in Bisbul leaves. Three regression models are used in making this prediction system, namely the Partial Least Square Regression (PLSR), Random Forest, and XGBoost Regressor models. The prediction system using PLSR produces 3.62 (RMSE test), 0.81 (R2 test), an accuracy of 91.3%, and a training time of 0.27 seconds. The prediction system uses Random Forest without using the selection feature with results of 4.04 (RMSE test), 0.81 (R2 test), an accuracy of 90.86%, and a training time of 17.81 seconds. The prediction system using Random Forest with feature selection resulted in 3.84 (RMSE test), 0.79 (R2 test), an accuracy of 91.31%, and a training time of 19.05 seconds. The prediction system using the XGBoost Regressor produces 3.48 (RMSE test), 0.83 (R2 test), an accuracy of 91.1%, and training time of 24.9 seconds. The best performance is produced by XGBoost Regressor with a slight difference from PLSR. The XGBoost Regressor model managed to improve performance by 14% on RMSE and 2% on R2 compared to PLSR.