



INTISARI

Karbon organik tanah merupakan indikator utama kualitas tanah yang berperan dalam ketersediaan hara, aktivitas biologi, stabilitas agregat, dan penyerapan karbon untuk mitigasi perubahan iklim. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Random Forest* (RF), *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB) dalam memprediksi kandungan karbon organik tanah, mengidentifikasi variabel lingkungan dominan, dan memetakan pola sebaran spasial kandungan karbon organik tanah. Sampel tanah diambil pada kedalaman 0–20 cm dan 20–40 cm dan pengukuran laboratorium dilakukan pada sifat-sifat tanah seperti pH, tekstur, berat volume, kadar lengas, Kapasitas Pertukaran Kation, dan karbon organik tanah sebagai dataset. Variabel lingkungan seperti NDVI, curah hujan, suhu, dan atribut topografi diperoleh dari pemrosesan citra Sentinel-2A dan LiDAR DEM. Model dibangun menggunakan dataset pelatihan dan pengujian (70:30) yang dioptimalkan dengan penyesuaian tuning hyperparameter dan validasi silang. Evaluasi dilakukan menggunakan RMSE, R^2 , MAE, dan RPD. Hasil menunjukkan bahwa model ensemble adalah yang paling akurat ($R^2 = 0,996$), diikuti oleh RF (0,995), XGB (0,992), dan SVM (0,889). Variabel yang paling berpengaruh adalah kedalaman, pH, tekstur lempung, elevasi, dan densitas volume. Polanya distribusi acak ditemukan pada lapisan 0–20 cm, sementara lapisan 20–40 cm pola distribusi secara berkelompok dengan kandungan tinggi di bagian selatan Sub-DAS, yang didominasi oleh agroforestri. Integrasi GIS dan pembelajaran ensemble terbukti efektif untuk memprediksi karbon organik tanah dan mendukung strategi pertanian berkelanjutan.

Kata kunci: ensemble learning; GIS; karbon organik tanah; pertanian berkelanjutan.



ABSTRACT

Soil organic carbon is a key indicator of soil quality that plays a role in nutrient availability, biological activity, aggregate stability, and carbon sequestration for climate change mitigation. This study aims to compare the performance of Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB) algorithms in predicting organic carbon content, identifying dominant environmental variables, and mapping the spatial distribution patterns of soil organic carbon content. Soil samples were taken at depths of 0–20 cm and 20–40 cm, and laboratory measurements were performed on soil properties such as pH, texture, bulk density, water content, cation exchange capacity, and soil organic carbon as a dataset. Environmental variables such as NDVI, rainfall, temperature, and topographic attributes were obtained from Sentinel-2A and LiDAR DEM image processing. Models were built using training and testing datasets (70:30) optimized with hyperparameter tuning and cross-validation. Evaluation was performed using RMSE, R^2 , MAE, and RPD. The results show that the ensemble model is the most accurate ($R^2 = 0,996$), followed by RF (0,995), XGB (0,992), and SVM (0,889). The most influential variables are depth, pH, clay texture, elevation, and bulk density. A random distribution pattern was found in the 0–20 cm layer, while the 20–40 cm layer showed a clustered distribution pattern with high content in the southern part of the sub-watershed, which is dominated by agroforestry. The integration of GIS and ensemble learning proved to be effective for predicting soil organic carbon and supporting sustainable agricultural strategies.

Keywords: ensemble learning; GIS; soil organic carbon; sustainable agriculture