

Perbandingan Pemilihan Variabel *Stepwise Regression* dan Boruta untuk Estimasi Stok Karbon Vegetasi Tegakan Menggunakan Citra Hiperspektral PRISMA: Studi Kasus pada Sebagian Kawasan Taman Nasional Gunung Merbabu

Pegi Melati
18/429720/GE/08905

ABSTRAK

Ekosistem hutan memainkan peran penting dalam siklus karbon yang mempengaruhi keseimbangan karbon. Penurunan fungsi hutan sebagai penyimpan cadangan karbon akibat deforestasi dan degradasi hutan dapat menyebabkan pelepasan karbon ke atmosfer (emisi karbon) sehingga mempengaruhi kondisi iklim global. Pengukuran nilai biomassa hingga stok karbon diperlukan sebagai upaya untuk memahami kondisi ekosistem hutan dan kaitannya dengan perubahan iklim. Kombinasi sistem penginderaan jauh sensor hiperspektral dan pembelajaran mesin dapat digunakan untuk mengestimasi stok karbon vegetasi tegakan yang menjadi salah satu simpanan karbon terbesar. Namun, diperlukan optimalisasi variabel melalui pemilihan variabel guna mengoptimalkan data masukan sekaligus mencari kombinasi variabel prediktor terbaik. Penelitian ini bertujuan untuk: 1) Mengetahui variabel paling optimal dari Citra Hiperspektral PRISMA untuk memodelkan stok karbon vegetasi tegakan; 2) Menguji akurasi model estimasi stok karbon vegetasi tegakan yang dibangun menggunakan variabel optimal hasil pemilihan variabel *Stepwise Regression* dan Boruta; 3) Melakukan estimasi dan pemetaan stok karbon vegetasi tegakan di sebagian Kawasan Taman Nasional Gunung Merbabu. Metode pemilihan variabel menggunakan pendekatan statistik dengan *Stepwise Regression* serta metode *Wrapper* dengan algoritma Boruta. Sebanyak 6 kombinasi saluran/band tunggal Citra PRISMA hasil pemilihan variabel digunakan untuk memodelkan stok karbon menggunakan algoritma *Random Forest*. Hasilnya diperoleh, model *Random Forest*-Boruta memiliki akurasi lebih tinggi dibandingkan model *Random Forest*-*Stepwise Regression*. Model *Random Forest*-Boruta I ($\alpha = 0,01$) menunjukkan konsistensi nilai yang paling baik dengan nilai RMSE (2,25 ton/piksel), nRMSE (22,77%), SEE (2,6 ton/piksel), maksimum akurasi (65,52%), serta nilai Bias (0,23). Nilai *R-Square* dan *Adjusted R-Square* model tersebut sebesar 0,6963 dan 0,6457. Sementara itu, akurasi terbaik dari pemilihan variabel *Stepwise Regression* dimiliki Model *Random Forest*-*Stepwise Regression* II. Nilai RMSE, nRMSE, SEE, maksimum akurasi, serta nilai bias berturut-turut adalah 3 ton/piksel, 30,37%, 3,47 ton/piksel, 54,01%, dan 0,20. Sementara nilai *R-Square* dan *Adjusted R-Square* sebesar 0,3205 dan 0,2073.

Kata Kunci: *Above Ground Carbon, Variable Selection, Machine Learning, Stepwise Regression, Boruta*

Comparison of Variable Selection Stepwise Regression and Boruta for Estimating Trees Vegetation Carbon Stock using PRISMA Hyperspectral Imagery: Study Case in a Part of Mount Merbabu National Park

Pegi Melati
18/429720/GE/08905

ABSTRACT

Forest ecosystems play an important role in the carbon cycle which affects carbon balance. The decline in the function of forests as a store of carbon stocks due to deforestation and forest degradation can cause the release of carbon into the atmosphere (carbon emissions) thereby affecting global climate conditions. Measurement of biomass and carbon stock is needed as an effort to understand the condition of forest ecosystems and their relation to climate change. The combination of hyperspectral remote sensing systems and machine learning can be used to estimate the carbon stock of trees vegetation which is one of the largest carbon stores. However, it is necessary to optimize the variables by selecting variables in order to optimize the input data as well as find the best combination of predictor variables. This study aims to: 1) Determine the most optimal variables from PRISMA Hyperspectral Imagery for modeling carbon stock of trees vegetation; 2) Test the accuracy model of the trees vegetation carbon stock built using the optimal variable from Stepwise Regression and Boruta; 3) Estimating and mapping the trees vegetation carbon stock in part of Mount Merbabu National Park. Variable selection method used a statistical approach with Stepwise Regression and Wrapper method with Boruta algorithm. A total of 6 single band combinations of PRISMA imagery as a result of selecting variables were used to build model carbon stock with Random Forest algorithm. The result is that the Random Forest-Boruta model has higher accuracy than the Random Forest-Stepwise Regression model. Model Random Forest-Boruta I ($\alpha = 0.01$) shows the best consistency of values with RMSE (2.25 tonnes/pixel), nRMSE (22.77%), SEE (2.6 tonnes/pixel), maximum accuracy (65.52%), and Bias (0.23). R-Square and Adjusted R-Square values of the model are 0.6963 and 0.6457. Meanwhile, the best accuracy of Stepwise Regression variable selection belongs to Random Forest-Stepwise Regression II. RMSE, nRMSE, SEE, maximum accuracy, and bias values were 3 tonnes/pixel, 30.37%, 3.47 tonnes/pixel, 54.01%, and 0.20, respectively. While the R-Square and Adjusted R-Square values are 0.3205 and 0.2073.

Keyword: Above Ground Carbon, Variable Selection, Machine Learning, Stepwise Regression, Boruta