

ABSTRAK

Parameter Tuning merupakan sebuah tahapan optimasi pada proses pembentukan model machine learning. Probst et al. (2019) menjelaskan bahwa proses penggunaan package dari R berdampak kecil terhadap model *Random Forest*. Wu et al. (2019) menggunakan Algoritma *Grid Search* yang mana menghasil peningkatan *accuracy* tapi memakan waktu yang cukup lama selama prosesnya. Yang & Shami (2020) menerapkan beberapa algoritma optimasi parameter yang meningkatkan akurasi hanya sebesar 3 dan diperlukan berbagai macam *package* external dari python untuk melakukan pengujian dan juga memakan waktu yang cukup lama untuk proses pemasangan dan pengujian *package*. Tsai et al. (2018) menggunakan *Brute Force* untuk mencari Parameter dengan nilai *accuracy* tertinggi, tapi *accuracy* model kurang bagus.

Random Forest dipilih karena memiliki nilai *accuracy* yang besar, bagus dalam menangani *outliers* dan *noise* serta mampu mengestimasi *error* dan *feature importance* (Brovelli, 2020) dan juga mampu menangani data yang berjumlah besar, memiliki banyak *features* dan *parameters* serta tidak sensitif terhadap *overfitting* (Triscowati et al., 2020). Parameter pada *Random Forest Classifier* dapat diatur dan dapat secara langsung memengaruhi kemampuan model *machine learning*, sehingga optimasi *parameter* diperlukan untuk mencapai tingkat kinerja terbaik dari sebuah algoritma (Shaharum, 2020).

Penelitian ini bertujuan untuk menjelaskan dampak perubahan kombinasi *features* dan nilai *parameter* pada model *Random Forest Classifier* untuk pemetaan lahan vegetasi di Kabupaten Sleman. Pada penelitian ini data Sentinel 2 dan Dynamic World yang digunakan diunduh dari Katalog Data milik Google Earth Engine. Indeks Vegetasi seperti *NDVI* dan *EVI* ada sebagai *features* tambahan.

Pengujian dilakukan 4 kali dengan kombinasi *features* yang berbeda-beda yang bertujuan untuk mencari *accuracy* terbaik. Pada masing-masing pengujian juga dilakukan parameter tuning pada model *Random Forest Classifier* yang dibuat. Dari hasil percobaan, kombinasi semua *features* yang ada (B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, *NDVI*, dan *EVI*) memiliki *accuracy* yang lebih tinggi dibandingkan kombinasi *feature* yang lain. Sedangkan *accuracy* tertinggi pada kombinasi semua *features* dicapai dengan menggunakan parameter *maxNodes* = 50 dan *numberOfTrees* = 100. *Accuracy* yang dicapai adalah 97.4% dan *feature* yang paling berpengaruh pada model ini adalah B1 (Aerosol).

Kata kunci: Lahan Vegetasi, *Random Forest Classifier*, *Normalized Difference Vegetation Index*, *Enhanced Vegetation Index*

ABSTRACT

Parameter tuning is an optimization step in the process of building a machine learning model. Probst et al. (2019) explained that the process of using packages from R has little impact on the Random Forest model. Wu et al. (2019) use the Grid Search Algorithm which results in increased accuracy but takes a long time during the process. Yang & Shami (2020) applied several parameter optimization algorithms which increased accuracy by only 3 and required various external packages from python to test and also took a long time to install and test packages. Tsai et al. (2018) used Brute Force to find the parameter with the highest accuracy value, but the accuracy of the model was not good.

Random Forest was chosen because it has a large accuracy value, is good at handling outliers and noise and is able to estimate error and feature importance (Brovelli, 2020) and is also able to handle large amounts of data, has many features and parameters and is not sensitive to overfitting (Triscowati et al., 2020). The parameters of the Random Forest Classifier can be adjusted and can directly affect the capabilities of the machine learning model, so parameter optimization is needed to achieve the best performance level of an algorithm (Shaharum, 2020).

This research aims to explain the impact of changes in the combination of features and parameter values on the Random Forest Classifier model for vegetation land mapping in Sleman Regency. In this research, the Sentinel 2 and Dynamic World data used were downloaded from Google Earth Engine's Data Catalog. Vegetation indices such as NDVI and EVI are available as additional features.

Tests were conducted 4 times with different combinations of features to find the best accuracy. In each test, parameter tuning was also performed on the Random Forest Classifier model created. From the experimental results, the combination of all existing features (B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, NDVI, and EVI) has a higher accuracy than other feature combinations. While the highest accuracy in the combination of all features is achieved by using the parameters $maxNodes = 50$ and $numberOfTrees = 100$. The accuracy achieved is 97.4% and the most influential feature in this model is B1 (Aerosol).

Keyword: Vegetation Land, Random Forest Classifier, Normalized Difference Vegetation Index, Enhanced Vegetation Index