

ABSTRACT

Macroalgae is an essential raw material for many industries, producing high economic value of various derived products. Each macroalga has a unique composition that might provide specific physical and chemical information that can be used as markers for authentication. Their compositions may differ depending on different factors, including geographical regions. The application of unsupervised exploratory techniques, namely principal component analysis (PCA), hierarchical cluster analysis (HCA), and nonparametric supervised techniques such as support vector machine (SVM) and random forest (RF) to the Vis-NIR spectroscopic data was performed to standardize the quality of macroalgae based on three regional zones in Indonesia (West, Center, East). A total of 35 macroalgae samples from six islands in Indonesia were analyzed. The PCA and HCA results present a tendency for the sample to be distributed and clustered according to the type of their species. Meanwhile, SVM successfully classified samples based on their regional zones, while in combination with 5-Fold Cross-Validation, achieved an accuracy of 82%. The RF model algorithm obtained an accuracy of 100%, 80%, and 82% for the training, test, and 5-Fold Cross-Validation, respectively.

Keywords: seaweed, geographical origin, exploratory study, Support Vector Machine, Random Forest