

INTISARI

Pada saat ini, pemanfaatan teknologi dalam penginderaan jauh telah sangat berkembang. Pengolahan citra satelit untuk pemetaan *Land Use Land Cover* telah banyak melibatkan *machine learning*. Perbandingan jenis algoritma pembelajaran pada *machine learning* dengan tujuan klasifikasi LULC telah banyak ditemukan, salah satunya adalah untuk jenis pembelajaran secara terbimbing (*Supervised Learning*). Algoritma yang diterapkan untuk klasifikasi secara terbimbing antara lain *Support Vector Machine* (SVM), *Random Forest*, Naïve Bayes dan *Decision Tree*. Pada penelitian ini, perbandingan algoritma *machine learning* untuk klasifikasi LULC adalah algoritma SVM dan NB. Algoritma SVM dan Naïve Bayes telah banyak diterapkan dalam klasifikasi tutupan lahan. Dimana kedua algoritma ini memberikan hasil yang cenderung lebih baik dibandingkan algoritma lainnya (Qian dkk., 2014; Huang dkk., 2003). Algoritma Naïve Bayes sendiri memiliki kelebihan dalam hal klasifikasi yang dapat dilakukan secara cepat namun dengan tingkat akurasi yang baik. Sementara, SVM memberikan kelebihan dalam hasil optimal dengan dapat dilakukannya optimasi parameter. Tujuan dari penelitian ini sendiri adalah untuk mengevaluasi performa dari kedua algoritma setelah dilakukannya optimasi parameter pada *classifier* dalam proses klasifikasi dan memahami penyebab variasi akurasi dari model hasil klasifikasi.

Citra satelit yang dimanfaatkan dalam klasifikasi adalah citra sentinel-2 *MultiSpectral Instrument*(MSI) level-2A. Dengan *random sampling*, dibuat 3 dataset dengan jumlah sampel secara berturut-turut adalah 7402 piksel (120 fitur), 8018 piksel (156 fitur) dan 8722 piksel (192 fitur) dengan perbandingan *training* dan *testing sample* adalah 70% berbanding 30%. Skema klasifikasi tutupan lahan yang digunakan mengacu pada SNI 7645-1:2014 dengan modifikasi. Pengujian keterpisahan antar kelas dilakukan dengan perhitungan statistik jarak *Jeffries-Matusita*. Klasifikasi tutupan lahan dengan algoritma SVM dilakukan dengan menerapkan *hyperparameter* dengan penggunaan kernel Linear dan RBF. Pengujian akurasi hasil klasifikasi tutupan lahan dilakukan dengan menggunakan matriks konfusi, demikian pula pada evaluasi model dengan penggunaan algoritma yang berbeda. Perbandingan model dengan algoritma berbeda dilakukan dengan evaluasi nilai *precision*, *recall*, *f1-score* untuk mengetahui sensitivitas serta presisi masing-masing model dalam menangkap data setiap kelas tutupan lahan, sementara itu evaluasi model secara keseluruhan dilakukan dengan menggunakan koefisien *Kappa*.

Klasifikasi tutupan lahan yang dilakukan menghasilkan sebanyak 22 model pada penggunaan algoritma SVM dan sebanyak 2 model pada penggunaan algoritma Naïve Bayes untuk masing-masing dataset. Secara keseluruhan tingkat keterpisahan antar kelas berada dalam rentang 1,67 hingga 2. Dengan banyak kelas tutupan lahan yang di klasifikasi adalah sebanyak 7 kelas yakni Tubuh Air, Daerah Terbangun, Lahan Pertanian, Lahan Terbuka, Lahan Terbuka Lainnya, Vegetasi dan Semak Belukar. *Overall Accuracy* tertinggi dihasilkan oleh algoritma *Support Vector Machine* dengan tipe kernel Linear yakni sebesar 96.86% dengan koefisien *Kappa* berada dalam kategori *almost perfect agreement*. Hasil ini sesuai dengan hipotesis yang diambil dalam penelitian, dimana algoritma SVM memberikan performa yang lebih baik dalam klasifikasi tutupan lahan dibandingkan algoritma Naïve Bayes.

Kata Kunci: LULC, SVM, Naïve Bayes, *Machine Learning*, Sentinel 2 MSI level-2A

ABSTRACT

At this time, the use of technology in remote sensing has been developed. Machine learning has involved in satellite image processing for Land Use Land Cover (LULC) mapping. There is many comparison of learning type algorithms of machine learning with the purpose of LULC classification, for example supervised learning algorithm. Supervised learning algorithm that applied for LULC classification are among Support Vector Machine (SVM), Random Forest (RF), Decision Tree, and Naïve Bayes. In this study, machine learning algorithms for LULC classification is compared between SVM and Naïve Bayes. SVM and Naïve Bayes algorithm have been widely applied in LULC classification. Both of them tend to give better performance than other algorithm (Qian dkk., 2014; Huang dkk., 2003). The Naïve Bayes algorithm has advantage in doing classification faster with good result. Meanwhile, SVM algorithm provide advantage in tuning parameter for more optimal result. The purpose of this research is to evaluate SVM and Naïve Bayes performance and understanding the cause of variation in accuracy of the model classification result.

Sentinel-2 MultiSpectral Instrument (MSI) level-2A is used in this study. By random sampling, three set of dataset were created with number of pixel detail for each dataset is 7402 pixels (120 features), 8018 pixels (156 features), and 8722 pixels (192 features). The ratio of training and testing sample is 70%:30%, where the 70% of sample in dataset is training data. The land classification scheme that used in this study refers to SNI 7645-1:2014. The separability test between each class is done by Jeffries-Matusita distance statistic calculation. Classification using SVM algorithm is carried out by applying hyperparameters with Linear and RBF kernels. Accuracy assessment for the classification result is carried out by using confusion matrix, as well as for model with different algorithm evaluations. Comparison between models with different algorithm is carried out by evaluating the precision, recall and f1-score value to get information about model sensitivity and precision in capturing data for each land cover class, while the evaluation of the overall model is carried out using cohen kappa.

The result of this study produce 22 models using SVM algorithm and 2 models using Naïve Bayes algorithm for each dataset. The separability level between two classes overall is in the range of 1,67 to 2. There are 7 classes of LULC to be classified in this study, consist of Water Body, Built-up Area, Agriculture Land, Bareland, other Bareland, Vegetation and Grass/Shrub Land. The highest overall accuracy is generated by the SVM algorithm using linear kernel type with accuracy 96.86% and the cohen kappa of the model is within almost perfect agreement category. The obtained result clarified the hypothesis in this research is true, where the SVM algorithm has better performance in LULC classification than Naïve Bayes algorithm.

Key words: LULC, SVM, Naïve Bayes, Machine Learning, Sentinel-2 MSI Level-2A