

ABSTRACT

Educational Data Mining has gained great attention in the last few years. Data Mining generates new knowledge that helps institutions to improve their teaching methods and learning process. Educational Data Mining researches continues to increase, but some studies show an increase in performance methods that have not been optimal enough. This study aims to improve the performance of the model to be more optimal by predicting students' academic performance.

This research proposes a combination of modeling methods in improving its performance, Ensemble method and feature selection method. Ensemble method that applied in this research consist of Bagging and Boosting. Various studies have proven that the Ensemble method is able to improve the performance of classification modeling. And the feature selection method used Wrapper, where this method uses predictive model to compute the feature set. The Wrapper method trains new models for each subset, which usually provides the best performing feature set for a particular model type. This research proposed the implementation of Ensemble method and feature selection method, which is divided into four modeling scenarios using Decision Tree C4.5 algorithm. Evaluation of modeling performance using four indicators; accuracy, precision, recall, and F Measure.

The experimental results show that by using a combination of Ensemble methods and feature selection can improve the performance of the classification modeling. The test that applied the combination methods was able to increase the average value to 74.97% for accuracy and recall, 71,43% for precision value, 72,35% for F Measure value by applying a combination of the Wrapper and Bagging method. The results show that the proposed method was able to improve the model performance significantly.

Keywords : *Ensemble Methods, Feature Selection Methods, Decision Tree, Classification, Educational Data Mining.*

INTISARI

Educational Data Mining telah mendapat perhatian besar dalam beberapa tahun terakhir. *Educational Data Mining* menghasilkan pengetahuan baru yang memungkinkan membantu institusi dalam meningkatkan metode pengajaran dan proses pembelajaran. Penelitian terkait *Educational Data Mining* terus meningkat, namun beberapa penelitian menunjukkan kenaikan kinerja metode yang belum cukup baik. Penelitian ini bertujuan untuk meningkatkan kinerja model menjadi lebih baik dengan memaksimalkan kinerjanya dalam memprediksi prestasi akademik mahasiswa.

Penelitian ini mengusulkan kombinasi metode pemodelan dalam meningkatkan kinerjanya, yaitu dengan metode Ensemble dan metode seleksi fitur. Metode Ensemble yang diterapkan pada penelitian ini terdiri dari Bagging dan Boosting. Berbagai penelitian telah membuktikan bahwa metode Ensemble mampu meningkatkan kinerja pemodelan klasifikasi. Dan untuk metode seleksi fitur menggunakan Wrapper, dimana metode ini menggunakan model prediktif untuk menghitung himpunan bagian fitur. Metode Wrapper melatih model baru untuk setiap *subset*, yang biasanya menyediakan set fitur berkinerja terbaik untuk jenis model tertentu. Pada penelitian ini diusulkan penerapan metode seleksi fitur dan metode Ensemble yang dibagi menjadi 4 skenario pemodelan dengan menggunakan algoritme Decision Tree C4.5. Evaluasi kinerja pemodelannya menggunakan empat indikator, yaitu akurasi, presisi, *recall*, dan *F Measure*.

Hasil percobaan menunjukkan bahwa dengan menggunakan kombinasi metode Ensemble dan seleksi fitur mampu meningkatkan kinerja pemodelan klasifikasi. Pengujian yang menerapkan kombinasi metode mampu meningkatkan nilai rata-ratanya menjadi 74,97% untuk nilai akurasi dan *recall*, 71,43% untuk nilai presisi, 72,35% untuk nilai *F Measure* dengan menerapkan kombinasi metode Wrapper dan Bagging. Hasil menunjukkan bahwa metode usulan tersebut memberikan peningkatan kinerja model yang signifikan.

Kata kunci – Metode Ensemble, Metode Seleksi Fitur, Decision Tree, Klasifikasi, *Educational Data Mining*.