

ABSTRACT

The growth of ITS accelerated by the current availability and simplicity of its internet access. ITS is an artificial intelligence technology used in the education field which plays an important role in determining the student's knowledge state. However, problems related to student's knowledge state estimation still exist. Therefore to overcome those problems it is needed to build a suitable model.

BKT as a student model with a probabilistic approach has advantage in terms of interpretability, thus student's knowledge state change can be traced. Based on previous research, the resulting AUC value is in the range of 71.8% – 83.94%. Hence the BKT performance can still be improved. The lack of pre-processing, including feature selection, leads to poor BKT performance. Several previous studies in the field of Educational Data Mining have proven that feature selection can significantly improve model performance. Therefore, this study aims to improve student modeling based on BKT using feature selection. The feature selection methods used are PCC, F-regression, LASSO and MI. Meanwhile, the dataset used in this research is ASSISTments 2009-2010, besides the Assistments Math 2006-2007 and Bridge to Algebra 2006-2007 datasets are also used to support model performance testing.

The experimental results show that feature selection can improve BKT performance. PCC with ASSISTments 2009-2010 dataset has the highest AUC value of 79.87% from the four feature selection methods used. The AUC value obtained is better than the three previous studies. Meanwhile, the feature subset generated by PCC consists of 24 features.

Keywords: student modeling, Bayesian Knowledge Tracing, feature selection, student's knowledge state

INTISARI

Ketersediaan dan kemudahan akses internet yang ada saat ini memicu perkembangan ITS. ITS merupakan salah satu penerapan *Artificial Intelligence* di bidang pendidikan yang berperan penting dalam menentukan *student's knowledge state*. Meskipun demikian masih ditemukan permasalahan terkait estimasi *student's knowledge state* sehingga diperlukan model yang mampu mengatasi permasalahan tersebut.

BKT sebagai *student model* dengan pendekatan probabilistik memiliki kelebihan dari segi interpretabilitas sehingga perubahan *student's knowledge state* dapat ditelusuri. Namun berdasarkan penelitian sebelumnya nilai AUC yang dihasilkan berada pada rentang 71,8% – 83,94% sehingga kinerja BKT masih tergolong kurang dan dapat ditingkatkan. Kinerja BKT yang belum optimal tersebut disebabkan oleh tidak dilakukannya serangkaian tahap *pre-processing* termasuk seleksi fitur. Beberapa penelitian pada bidang *Educational Data Mining* yang telah dilakukan sebelumnya membuktikan bahwa seleksi fitur dapat meningkatkan kinerja model secara signifikan. Oleh karena itu penelitian ini bertujuan untuk meningkatkan kinerja BKT menggunakan seleksi fitur. Metode seleksi fitur yang digunakan adalah PCC, *F-regression*, LASSO dan MI. Sedangkan *dataset* yang digunakan pada penelitian ini adalah ASSISTments 2009-2010, selain itu *dataset* Assistments Math 2006-2007 dan Bridge to Algebra 2006-2007 juga digunakan untuk mendukung pengujian performa model. Kemudian metrik untuk mengevaluasi kinerja model adalah AUC dan RMSE.

Hasil eksperimen yang dilakukan menunjukkan bahwa seleksi fitur dapat meningkatkan kinerja BKT. Dari keempat metode seleksi fitur yang digunakan PCC dengan *dataset* ASSISTments 2009-2010 memiliki nilai AUC tertinggi yaitu sebesar 79,87%. Nilai AUC tersebut lebih unggul dibandingkan dengan tiga penelitian sebelumnya. Sementara itu *subset* fitur yang dihasilkan oleh PCC terdiri dari 24 fitur.

Kata kunci: pemodelan siswa, *Bayesian Knowledge Tracing*, seleksi fitur, *student's knowledge state*