

ABSTRACT

Student performance is intensively explored in the context of Educational Data Mining to assist in dealing with issues such as low academic performance, increasing dropout rates, and delayed graduation. Machine learning classification is one method that may be utilized in this regard to develop a predictive model of student performance that can be used as a reference in taking corrective steps for students who are at risk of failure or developing an early warning system. There are at least two important factors contribute to this: optimal model performance and interpretable models that allows stakeholders to understand how the model produces predictions. A white-box model has the advantage of being easier to interpret and analyze but it does not outperform a black-model in most cases. The correct setting of hyper-parameters can affect the performance of a model. Manual testing of hyper-parameter settings is ineffective due to several factors, including the vast number of hyper-parameters, the complexity of machine learning models, the time required and the non-linear relationship between hyper-parameters.

In this study, an automated technique for hyper-parameter setting is presented using a combination of random search and grid search methods with a coarse-to-fine approach to exploit the possibility in finding more optimal configuration close to the hyper-parameter values obtained from the initial search space. From the results of trials carried out at the evaluation stage, the average improvement in model performance was: 1.82% for the Accuracy metric; 4.29% for the F1-Score metric; and 6.25% for the MCC metric with an average computation time of 87.42 minutes and an average total iteration of 3,641 iterations.

Keywords : educational data mining, student performance prediction, automated hyper-parameter setting

INTISARI

Dalam konteks *Educational Data Mining*, kinerja murid diteliti secara luas untuk membantu dalam menangani beberapa permasalahan seperti capaian kinerja akademis yang rendah, meningkatnya angka putus sekolah hingga kelulusan yang tertunda. Salah satu metode yang dapat digunakan terkait hal tersebut adalah metode klasifikasi *machine learning* untuk mengembangkan sebuah model prediksi kinerja murid yang dapat digunakan sebagai acuan dalam mengambil langkah-langkah perbaikan terhadap murid yang berisiko mengalami kegagalan hingga mengembangkan sistem peringatan dini. Setidaknya terdapat dua faktor penting dalam mendukung hal tersebut, yaitu kinerja model yang optimal dan model yang dapat diinterpretasikan sehingga para pemangku kepentingan memahami bagaimana model tersebut melakukan prediksi. Model *white-box* cenderung lebih mudah untuk diinterpretasikan dan dijelaskan namun secara umum memiliki kinerja yang tidak lebih baik daripada model *black-box*. Penentuan *hyper-parameter* yang tepat dapat berpengaruh pada kinerja sebuah model. Pengujian secara manual terhadap pengaturan *hyper-parameter* kurang efektif untuk ditempuh dikarenakan beberapa faktor antara lain banyaknya jumlah *hyper-parameter*, kompleksitas model *machine learning*, waktu yang diperlukan serta hubungan non-linier antar *hyper-parameter*.

Dalam penelitian ini disajikan teknik otomasi dalam penentuan *hyper-parameter* menggunakan kombinasi metode *random search* dan *grid search* dengan pendekatan *coarse-to-fine* untuk mengeksplorasi kemungkinan konfigurasi yang lebih optimal berada di sekitar dari nilai-nilai *hyper-parameter* yang didapatkan dari *search space* awal. Dari hasil uji coba yang dilakukan pada tahapan evaluasi, diperoleh rata-rata peningkatan kinerja model sebesar: 1,82% untuk metrik *Accuracy*; 4,29% untuk metrik *F1-Score*; dan 6,25% untuk metrik MCC dengan rata-rata waktu komputasi mencapai 87,42 menit serta rata-rata total iterasi sebanyak 3.641 iterasi.

Kata kunci – *educational data mining*, prediksi kinerja murid, *penentuan hyper-parameter* otomatis