

ABSTRACT

Detecting confusion has been considered as a critical issue in online education platforms. Several studies have been done to develop confusion-state classifier models. The best previous model was developed using the Bidirectional LSTM (Bi-LSTM) algorithm and has an average accuracy of 75%. Unfortunately, the model still has several drawbacks. The first one would be the characteristics of the Bi-LSTM which are prone to the over-fitting problem and has no specific mechanism to deal with sparsity. The second one is the lack of any hyper-parameter technique in the model's learning process. Those drawbacks could adversely affect the model's accuracy.

Several solutions to those problems are proposed through this research. The first one is to replace the machine learning algorithm with XGBoost. XGBoost is sparsity-aware and contains a regularization function to avoid the over-fitting problem. The second solution is to apply a sophisticated hyper-parameter optimization technique called the Tree-structured Parzen Estimator (TPE) in the model's learning process. This research aims to improve the accuracy of the previous confusion-state classifier model using those solutions.

The proposed model, XGB-RFE-TPE-6, had outperformed the previous ones by achieving an average accuracy of 87%. This study had presented the current confusion-state classifier model. This study also brought out the most optimal configuration of features and hyper-parameters to build such a model.

Keywords : confusion, classifier, XGBoost, Tree-Structured Parzen Estimator, hyper-parameter optimization

INTISARI

Mendeteksi kebingungan merupakan isu penting yang banyak dibahas dalam dunia pendidikan, khususnya pada sistem pembelajaran daring. Sejumlah peneliti telah mengembangkan model *Machine Learning* (ML) pendeteksi kebingungan. Model terbaik dari penelitian terdahulu dikembangkan menggunakan algoritme ML Bidirectional LSTM (Bi-LSTM) dan mampu meraih nilai rata-rata akurasi sebesar 75%. Sayangnya, masih terdapat sejumlah kelemahan pada model tersebut, yaitu sifat Bi-LSTM yang rentan terhadap permasalahan *over-fitting* dan tidak memiliki mekanisme khusus untuk menangani permasalahan *sparsity*, serta tidak diterapkannya teknik optimasi *hyper-parameter* dalam proses pengembangan model. Semua kelemahan tersebut dapat mengakibatkan nilai akurasi model menjadi kurang optimal.

Penelitian ini menawarkan solusi-solusi dari kelemahan-kelemahan tersebut. Solusi yang pertama adalah mengganti algoritme ML menjadi XGBoost. XGBoost memiliki fungsi regularisasi untuk menekan risiko *over-fitting* dan memiliki algoritme pembelajaran yang *sparsity-aware*. Solusi yang kedua adalah menambahkan teknik optimasi *hyper-parameter* yang mutakhir, yaitu Tree-Structured Parzen Estimator (TPE). Penelitian ini bertujuan untuk meningkatkan nilai rata-rata akurasi dari model-model terdahulu menggunakan solusi-solusi tersebut.

Model yang diusulkan, XGB-RFE-TPE-6, telah melampaui model-model terdahulu dengan nilai rata-rata akurasi sebesar 87%. Penelitian ini telah menghasilkan model pendeteksi kebingungan yang paling mutakhir. Penelitian ini juga telah menemukan konfigurasi fitur dan *hyper-parameter* yang paling optimal untuk model pendeteksi kebingungan.

Kata kunci – kebingungan, klasifikasi, XGBoost, Tree-Structured Parzen Estimator, optimasi *hyper-parameter*