

ABSTRACT

Academic performance prediction is one of the main topics in Educational Data Mining that can assist educational institutions in identifying students academic achievement patterns at an early stage. This study aims to implement CatBoost with SMOTEN for multi-class classification of students' academic performance and to analyze feature contributions using an Explainable Artificial Intelligence approach. The dataset used in this study is a public student academic performance dataset consisting of 872 records, 30 categorical predictor features, and four target classes: Below Average, Average, Good, and Excellent. The research stages include data cleaning, feature transformation using manual ordinal mapping and native categorical handling, data balancing using SMOTEN on the training fold, CatBoost model development, evaluation using Stratified 10-Fold Cross-Validation with an Out-of-Fold scheme, and model interpretation using SHAP and LIME. The evaluation results show that the CatBoost baseline achieved an accuracy of 71.10%, while CatBoost with SMOTEN achieved the highest accuracy of 72.82%, with a macro precision of 74.03%, macro recall of 69.92%, and macro F1-score of 71.65%. The global interpretation using SHAP indicates that class attendance, accommodation, study time, midterm preparation, and exam preparation are the most influential features in the model prediction. Meanwhile, the local interpretation using LIME shows that the basis of prediction may differ for each student depending on their feature characteristics. This study indicates that CatBoost with SMOTEN can be used for academic performance prediction while providing both global and local interpretability.

Keywords: CatBoost, SMOTEN, Explainable Artificial Intelligence, SHAP, LIME, student academic performance.

ABSTRAK

Prediksi performa akademik mahasiswa merupakan salah satu kajian dalam *Educational Data Mining* yang dapat membantu institusi pendidikan mengidentifikasi pola capaian akademik mahasiswa secara lebih dini. Penelitian ini bertujuan untuk mengimplementasikan CatBoost dengan SMOTEN pada klasifikasi multikelas performa akademik mahasiswa serta menganalisis kontribusi fitur terhadap hasil prediksi menggunakan pendekatan *Explainable Artificial Intelligence*. Dataset yang digunakan merupakan dataset publik performa akademik mahasiswa yang terdiri atas 872 data dengan 30 fitur prediktor kategorikal dan empat kelas target, yaitu *Below Average*, *Average*, *Good*, dan *Excellent*. Tahapan penelitian meliputi *data cleaning*, transformasi fitur melalui *manual ordinal mapping* dan *native categorical handling*, *data balancing* menggunakan SMOTEN pada *training fold*, pengembangan model CatBoost, evaluasi menggunakan *Stratified 10-Fold Cross-Validation* dengan skema *Out-of-Fold*, serta interpretasi model menggunakan SHAP dan LIME. Hasil evaluasi menunjukkan bahwa CatBoost *baseline* memperoleh akurasi sebesar 71,10%, sedangkan CatBoost dengan SMOTEN memperoleh akurasi tertinggi sebesar 72,82%, *precision macro* sebesar 74,03%, *recall macro* sebesar 69,92%, dan *F1-score macro* sebesar 71,65%. Hasil interpretasi global menggunakan SHAP menunjukkan bahwa fitur kehadiran di kelas, tempat tinggal mahasiswa, waktu belajar, persiapan ujian tengah semester, dan pola persiapan ujian menjadi fitur yang paling berkontribusi terhadap prediksi model. Sementara itu, interpretasi lokal menggunakan LIME menunjukkan bahwa dasar prediksi dapat berbeda pada setiap mahasiswa sesuai karakteristik fitur yang dimiliki. Penelitian ini menunjukkan bahwa CatBoost dengan SMOTEN dapat digunakan sebagai pendekatan prediksi performa akademik mahasiswa dengan dukungan interpretabilitas global dan lokal.

Kata kunci: CatBoost, SMOTEN, *Explainable Artificial Intelligence*, SHAP, LIME, performa akademik mahasiswa.