ABSTRAK

Putus sekolah memerlukan deteksi dini yang akurat dan dapat dijelaskan agar hasil model dapat mendukung pengambilan keputusan akademik secara bertanggung jawab dalam sistem peringatan dini. Penelitian ini membandingkan regresi logistik, Random Forest, dan extreme gradient boosting untuk prediksi putus sekolah, serta mengkaji bagaimana kecerdasan buatan yang dapat dijelaskan menguraikan faktor-faktor prediktif yang berpengaruh melalui SHapley Additive exPlanations (SHAP). Dataset "Predict Students' Dropout and Academic Success" dari Machine Learning Repository berisi 4.424 pengamatan dan 36 fitur. Kelas "Enrolled" dikecualikan karena bukan merupakan hasil akhir, sehingga menghasilkan dataset biner dengan 3.630 pengamatan yang terdiri dari 1.421 catatan "Dropout" dan 2.209 catatan "Graduate". Data dibagi menjadi 60% untuk set pelatihan, 20% untuk set validasi, dan 20% untuk set pengujian. Hiperparameter dan ambang batas klasifikasi dipilih berdasarkan data validasi, sedangkan set pengujian disisihkan untuk evaluasi akhir. Regresi logistik mencapai area di bawah kurva presisi-recall tertinggi sebesar 0,9728 dan recall putus sekolah sebesar 0,9754. Random Forest menghasilkan skor F1 tertinggi sebesar 0,8346 dan akurasi seimbang sebesar 0,8710. Extreme Gradient Boosting mencapai area di bawah kurva presisi-recall sebesar 0,9705, recall sebesar 0,9613, presisi sebesar 0,7319, skor F1 sebesar 0,8311, dan akurasi seimbang sebesar 0,8675. Penjelasan aditif Shapley menempatkan program studi, satuan kurikulum yang disetujui pada semester kedua, status pembayaran uang kuliah, dan prestasi akademik awal sebagai fitur-fitur dominan. Extreme gradient boosting kompetitif namun tidak secara universal unggul. Validasi lokal, evaluasi temporal, penilaian keadilan, dan pengujian intervensi diperlukan sebelum penerapan operasional.

Keywords: Early Warning System; Explainable Artificial Intelligence; SHapley Additive exPlanations (SHAP); Student Dropout Prediction; XGBoost.