

ABSTRAK

Perguruan tinggi kerap menghadapi tantangan dalam pengelolaan program studi, seperti penambahan kelas untuk mahasiswa mengulang, pembimbingan tugas akhir yang melebihi kapasitas, serta ketidakseimbangan rasio dosen dan mahasiswa. Masalah ini umumnya disebabkan oleh mahasiswa yang tidak lulus tepat waktu (≤ 4 tahun). Oleh karena itu, diperlukan analisis prediksi kelulusan tepat waktu untuk menjaga keseimbangan antara jumlah mahasiswa baru dan lulusan. Penelitian ini bertujuan membangun model prediksi kelulusan mahasiswa Universitas Siliwangi (Unsil) angkatan 2020 yang lulus hingga Agustus 2024, menggunakan algoritma XGBoost, serta mengidentifikasi faktor-faktor yang memengaruhinya. Sebanyak 1.535 data mahasiswa yang mencakup aspek ekonomi orang tua dan capaian akademik digunakan dalam model, dengan fitur seperti penghasilan dan pendidikan orang tua, jumlah tanggungan, jalur masuk, status KIP, dan IPK. Proses analisis meliputi pembersihan, encoding, pembagian data, dan pelatihan model. Evaluasi performa model dilakukan menggunakan confusion matrix dengan hasil akurasi 82%, precision 84%, recall 96%, dan F1-score 90%. Analisis SHAP menunjukkan bahwa mahasiswa dengan penghasilan ayah dibawah Rp1.500.000 dan jalur mandiri cenderung berturut-turut memiliki peluang lebih besar dan lebih kecil untuk lulus tepat waktu. Sementara itu, analisis feature importance menunjukkan bahwa KIP kuliah, jumlah tanggungan, IPK, dan penghasilan ayah merupakan faktor paling berpengaruh dalam prediksi kelulusan.

Kata Kunci: CRISP-DM, Feature importance, Kelulusan, Perguruan tinggi, SHAP.

ABSTRACT

Universities often face challenges in managing study programs, such as adding classes for repeating students, supervising final assignments that exceed capacity, and an imbalance in the ratio of lecturers and students. This problem is generally caused by students who do not graduate on time (≤ 4 years). Therefore, an analysis of timely graduation prediction is needed to maintain a balance between the number of new students and graduates. This study aims to build a graduation prediction model for students of Siliwangi University (Unsil) class of 2020 who graduate by August 2024, using the XGBoost algorithm, and identify the factors that influence it. A total of 1,535 student data covering aspects of parents' economy and academic achievements are used in the model, with features such as parents' income and education, number of dependents, admission path, KIP status, and GPA. The analysis process includes cleaning, encoding, data division, and model training. Model performance evaluation was carried out using a confusion matrix with results of 82% accuracy, 84% precision, 96% recall, and 90% F1-score. SHAP analysis shows that students with father's income below Rp1,500,000 and independent pathways tend to have a greater and lesser chance of graduating on time, respectively. Meanwhile, feature importance analysis shows that KIP college, number of dependents, GPA, and father's income are the most influential factors in predicting graduation.

Keywords: CRISP-DM, Feature importance, Graduation, Higher education, SHAP.