

ANALISIS FAKTOR PENGHAMBAT KELULUSAN MAHASISWA MENGUNAKAN MODEL C5.0 TEROPTIMASI GRIDSEARCHCV DAN INTERPRETASI SHAP

Oleh:

Irvan Falasifa Hasan

1197050056

ABSTRAK

Tingginya angka keterlambatan kelulusan di Program Studi Teknik Informatika UIN Sunan Gunung Djati Bandung mengancam kualitas lulusan dan status akreditasi 'Unggul'. Penelitian ini bertujuan mengoptimasi model klasifikasi *Decision Tree C5.0* menggunakan *GridSearchCV* dan memvisualisasikan faktor penghambat kelulusan menggunakan *Explainable AI (SHAP)*. Mengikuti kerangka kerja CRISP-DM, data kuesioner dari 198 mahasiswa diproses. Karena ketiadaan label target, *K-Means Clustering* diterapkan sebagai *pseudo-labeling* untuk membentuk kelas 'Risiko Tinggi' dan 'Risiko Rendah'. Ketidakseimbangan data latih ditangani dengan SMOTE, dilanjutkan seleksi fitur *SelectKBest*. Model *C5.0* kemudian dilatih dan dioptimasi secara otomatis melalui *GridSearchCV* terintegrasi *5-Fold Cross Validation*. Hasil penelitian membuktikan skenario pembagian data 80:20 sebagai model terbaik dengan parameter: kriteria *entropy*, batas kedalaman 6, minimum sampel percabangan 10, dan 8 fitur utama. Model ini menghasilkan performa klasifikasi sangat tinggi dengan tingkat Akurasi 92,50%, *Precision* 92,26%, *Recall* 90,46%, dan *F1-Score* 91,27%. Analisis *SHAP* membuktikan bahwa fasilitas pendukung belajar yang kurang memadai merupakan penentu utama tingginya risiko keterlambatan studi. Keterbatasan infrastruktur ini secara nyata menghambat kelancaran proses akademik, memberikan dampak yang jauh lebih besar dibandingkan masalah pribadi ataupun lingkungan sosial.

Kata Kunci: *C5.0, GridSearchCV, SHAP, K-Means, SMOTE.*

***ANALYSIS OF FACTORS HINDERING STUDENT GRADUATION USING
THE C5.0 MODEL OPTIMISED BY GRIDSEARCHCV AND SHAP
INTERPRETATION***

By:

Irvan Falasifa Hasan

1197050056

ABSTARCT

The high rate of delayed graduation in the Computer Science programme at UIN Sunan Gunung Djati Bandung threatens the quality of graduates and the programme's 'Excellent' accreditation status. This study aims to optimise the C5.0 Decision Tree classification model using GridSearchCV and to visualise the factors hindering graduation using Explainable AI (SHAP). Following the CRISP-DM framework, questionnaire data from 198 students were processed. Due to the absence of target labels, K-Means Clustering was applied as pseudo-labelling to form 'High Risk' and 'Low Risk' classes. The imbalance in the training data was addressed using SMOTE, followed by feature selection via SelectKBest. The C5.0 model was then trained and automatically optimised via GridSearchCV integrated with 5-fold cross-validation. The research results confirmed that an 80:20 data split yielded the best model with the following parameters: entropy criterion, depth limit of 6, minimum branching sample of 10, and 8 key features. This model achieved very high classification performance with an Accuracy of 92.50%, Precision of 92.26%, Recall of 90.46%, and an F1-Score of 91.27%. SHAP analysis confirmed that inadequate learning support facilities are the primary determinant of a high risk of academic delay. These infrastructure limitations significantly hinder the smooth running of academic processes, having a far greater impact than personal or social environmental issues.

Keywords: C5.0, GridSearchCV, SHAP, K-Means, SMOTE.