

ABSTRAK

Evaluasi dan deteksi dini terhadap mahasiswa yang berisiko mengalami sisip program (ketidaklayakan melanjutkan studi) sangat penting dilakukan agar Dosen Pembimbing Akademik (DPA) dapat memberikan peringatan secara tepat waktu kepada mahasiswa. Namun, upaya prediksi berbasis data sering terkendala oleh masalah ketidakseimbangan data (imbalanced data) karena jumlah kasus mahasiswa sisip program jauh lebih sedikit dibandingkan mahasiswa yang tidak mengalami sisip program, sehingga performa algoritma klasifikasi cenderung bias pada kelas mayoritas. Penelitian ini bertujuan untuk membangun sistem prediksi berbasis web menggunakan *framework Streamlit* dengan algoritma C4.5, serta mengevaluasi kinerja metode SMOTE-ENN dalam mengatasi ketidakseimbangan data. Data yang digunakan mencakup atribut akademik dan latar belakang mahasiswa semester 1 hingga semester 3, dengan pelabelan status sisip program yang ditentukan hingga semester 4. Hasil pengujian fitur dan fungsionalitas sistem menggunakan metode *black-box testing* menunjukkan bahwa seluruh fitur pada sistem prediksi berjalan sesuai rancangan. Pada tahapan validasi menggunakan *5-Fold Cross Validation* dengan jumlah minimum sampel pada daun sebesar 2, pemodelan dengan SMOTE-ENN berhasil menghasilkan rata-rata F1-Skor tertinggi sebesar 93,78%, mengungguli model tanpa penanganan SMOTE-ENN yang memperoleh nilai 91,18%. Aturan optimal yang diperoleh kemudian digunakan untuk membangun model final. Hasil pengujian model final pada data uji asli menghasilkan tingkat akurasi sebesar 97,16%, presisi 85,71%, recall sebesar 100%, dan F1-Skor sebesar 92,31%. Selain itu, hasil pengujian menggunakan kuesioner *Technology Acceptance Model* (TAM) terhadap DPA menghasilkan persentase rata-rata sebesar 84,2%, sehingga sistem dikategorikan sangat layak untuk diimplementasikan sebagai sistem prediksi mahasiswa berisiko sisip program.

Kata kunci : C4.5, Klasifikasi, SMOTE-ENN, data tidak seimbang, sisip program.

ABSTRACT

Early evaluation and detection of students at risk of Sisip Program (ineligibility to continue their studies) are important to enable Academic Advisors (DPA) to provide timely warnings. However, data-driven prediction is often hindered by imbalanced data, as the number of Sisip Program cases is much smaller than the number of students who do not experience Sisip Program, causing classification algorithms to be biased toward the majority class. This study aims to develop a web-based prediction system using the Streamlit framework and the C4.5 algorithm, and to evaluate the performance of the SMOTE-ENN method in handling data imbalance. The dataset consists of students' academic and background attributes from semesters 1 to 3, while Sisip Program status is determined up to semester 4. Black-box testing showed that all system features functioned as designed. During validation using 5-Fold Cross Validation with a minimum of 2 samples per leaf, the SMOTE-ENN model achieved the highest average F1-Score of 93.78%, outperforming the model without SMOTE-ENN treatment (91.18%). The resulting rules were used to build the final model. Evaluation on the original test data yielded an accuracy of 97.16%, precision of 85.71%, recall of 100%, and an F1-Score of 93.72%. A Technology Acceptance Model (TAM) questionnaire completed by Academic Advisors produced an average score of 84.2%, indicating that the system is highly feasible for implementation as a prediction system for students at risk of Sisip Program.

Keywords: C4.5, Classification, SMOTE-ENN, Imbalanced Data, Sisip Program.