

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi efektivitas metode Naïve Bayes dalam memprediksi status kelulusan mahasiswa, yakni antara kategori dropout dan graduate. Data yang digunakan berasal dari dataset historis mahasiswa dengan jumlah fitur yang beragam, mencakup aspek akademik, demografis, dan ekonomi. Penelitian menguji performa model klasifikasi dengan tiga skenario pembagian data uji coba yakni 60:40, 70:30, dan 80:20. Evaluasi dilakukan dengan menggunakan metrik akurasi, precision, recall, F1-score, serta analisis confusion matrix sebagai indikator kelas prediksi benar dan salah untuk tiap kategori. Hasil pengujian memperlihatkan 60:40 dengan hasil akurasi mencapai 84.92%, 70:30 mencapai 85.22%, dan 80:20 akurasinya mencapai 85.67% dengan demikian model Naïve Bayes mampu memberikan prediksi yang cukup akurat pada ketiga skenario, dengan performa terbaik dicapai pada skenario 60:40 yang menunjukkan keseimbangan optimal antara data pelatihan dan pengujian. Meskipun demikian, model masih menunjukkan sejumlah kesalahan klasifikasi yang mayoritas berasal pada bagian kelas dropout, sehingga diperlukan perhatian lebih dalam pengolahan data dan teknik balancing. Penelitian ini memberikan kontribusi dalam pengembangan sistem pendukung keputusan akademik berbasis data yang dapat membantu institusi pendidikan tinggi dalam melakukan identifikasi dini terhadap mahasiswa berisiko dropout, sehingga intervensi dan kebijakan dapat dilakukan secara lebih efektif.

Kata Kunci : Naïve Bayes, klasifikasi, kelulusan mahasiswa, dropout, evaluasi performa, confusion matrix.

ABSTRACT

This study aims to evaluate the effectiveness of the Naïve Bayes method in predicting student graduation status, specifically distinguishing between dropout and graduate categories. The data used comes from historical student datasets with various features, including academic, demographic, and economic aspects. The research assesses the classification model's performance under three data-splitting scenarios: 60:40, 70:30, and 80:20. Evaluation is conducted using accuracy, precision, recall, F1-score metrics, and a confusion matrix analysis to indicate correct and incorrect predictions for each class category. The results show that the 60:40 split yields an accuracy of 84.92%, 70:30 achieves 85.22%, and 80:20 reaches 85.67%. Thus, the Naïve Bayes model is capable of providing reasonably accurate predictions across all three scenarios, with the best performance observed in the 60:40 scenario, indicating an optimal balance between training and testing data. However, the model still exhibits some misclassifications, mainly within the dropout class, highlighting the need for improved data preprocessing and balancing techniques. This research contributes to the development of data-driven academic decision support systems, enabling higher education institutions to perform early identification of students at risk of dropping out, thereby facilitating more effective interventions and policy implementation.

Keywords: Naïve Bayes, classification, student graduation, dropout, performance evaluation, confusion matrix.