

ABSTRAK

Komparasi Algoritma C4.5 Dan Metode *Naive Bayes* Untuk Memprediksi Kelulusan Tepat Waktu Mahasiswa Universitas Muhammadiyah Pringsewu Lampung

Oleh

Sadimin

sadimin@umpri.ac.id

Seiring dengan perkembangan teknologi khususnya perkembangan penyimpanan data yang semakin besar. Salah satu organisasi yang memiliki penyimpanan data yang besar adalah organisasi pendidikan. Organisasi pendidikan menggunakan data untuk mendapatkan informasi, terutama informasi tentang mahasiswa. Data mahasiswa memiliki banyak atribut sehingga kita dapat membuat prediksi seperti prediksi kinerja mahasiswa, prediksi penerima beasiswa dan prediksi kelulusan mahasiswa. Metode *data mining* dalam pendidikan diklasifikasikan menjadi lima dimensi, salah satunya adalah prediksi seperti memprediksi nilai keluaran berdasarkan data masukan. Dari hasil penelitian yang dilakukan dari tahap awal hingga tahap pengujian penerapan Algoritma C4.5 mendapatkan hasil akurasi lebih tinggi dari *naive bayes* karena dalam tahapan pengklasifikasiannya C4.5 memproses satu persatu data atribut. Berbeda halnya dengan *naive bayes* yang dipengaruhi oleh banyaknya data yang digunakan, perbandingan jumlah data *training* dan *testing*. Kelayakan model yang didapatkan didukung dengan tingkat *accuracy*, *precision*, *recall* serta *AUC* yang diperoleh dari kedua algoritma yang telah diuji. Algoritma C4.5 memiliki tingkat akurasi 79,91 %, *precision* 89,06% dan *recall* 81.38% serta nilai *AUC* 0.823. Sedangkan *Naive Bayes* memiliki tingkat akurasi 76,95%, *precision* 75.95% dan *recall* 98.38% serta nilai *AUC* 0.838.

Kata Kunci: Kelulusan, Prediksi, Penambangan Data, C4.5, *Naive Bayes*

ABSTRACT

Comparison of the C4.5 Algorithm and the Naive Bayes Method for Predicting On-time Graduation of Muhammadiyah University Pringsewu Lampung Students

By

Sadimin
sadimin@umpri.ac.id

Along with technological developments, especially the development of increasingly large data storage. One organization that has large data storage is an educational organization. Educational organizations use data to obtain information, especially information about students. Student data has many attributes so that we can make predictions such as predictions of student performance, predictions of scholarship recipients and predictions of student death. Data mining methods in education are classified into five dimensions, one of which is prediction, such as predicting the output value based on the data entered. From the results of the research conducted from the initial stage to the testing stage of the application of the C4.5 Algorithm, the accuracy results are higher than Naive Bayes because in the classification stage C4.5 processes one data attribute that is processed. Unlike the case with naive Bayes which is influenced by the amount of data used, the comparison of the amount of training and testing data. The feasibility of the model obtained is supported by the level of accuracy, precision, recall and AUC obtained from the two algorithms that have been tested. The C4.5 algorithm has an accuracy rate of 79.91%, 89.06% precision and 81.38% recall and an AUC value of 0.823. Meanwhile, Naive Bayes has an accuracy rate of 76.95%, precision of 75.95% and recall of 98.38% and an AUC value of 0.838.

Keywords: Graduation, Prediction, Data Mining, C4.5, Naive Bayes