

ABSTRAK

Perbandingan Kinerja Algoritma C.45 DAN *Naive Bayes* Dalam Menentukan Resiko Kesehatan Ibu Hamil

Oleh

Reandi Setiwan Syah

[reandysetiawan@darmajaya.ac.id](mailto:reandysetiawan@ darmajaya.ac.id)

Seiring dengan perkembangan teknologi khususnya perkembangan penyimpanan data yang semakin besar. Salah satu organisasi yang memiliki penyimpanan data yang besar adalah Rumah Sakit . Organisasi Rumah sakit menggunakan data untuk mendapatkan informasi, terutama informasi tentang pasien. Data pasien memiliki banyak atribut sehingga kita dapat membuat prediksi seperti prediksi menentukan resiko kesehatan ibu hamil,. 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 77,72 %, *precision* 92,86% dan *recall* 78.08% serta nilai *AUC* 0.823. Sedangkan *Naive Bayes* memiliki tingkat akurasi 78,57%, *precision* 93.29% dan *recall* 98.38% serta nilai *AUC* 0.838.

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

ABSTRACT

Comparison of C.45 AND Naive Bayes Algorithm Performance in Determining Health Risk of Pregnant Women

By

Reandi Setiawan Syah
reandysetiawan@darmajava.ac.id

Along with the development of technology, especially the development of increasingly large data storage. One organization that has a large data storage is the Hospital. Hospital organizations use data to obtain information, especially information about patients. Patient data has many attributes so that we can make predictions such as predictions determine the health risks of pregnant women. Data mining methods in education are classified into five dimensions, one of which is prediction such as predicting output values based on input data. From the results of research conducted from the initial stage to the testing stage, the application of the C4.5 algorithm gets higher accuracy results than naïve bayes because in the classification stage C4.5 processes one by one attribute data Unlike the case with naïve bayes which are 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 77.72%, precision of 92.86% and recall of 78.08% and an AUC value of 0.823. While Naive Bayes has an accuracy rate of 78.57%, precision 93.29% and recall 98.38% and an AUC value of 0.838

Keywords: pregnancy, prediction, data mining, C4.5, Naive Bayes