

ABSTRAK

Penambangan data (*data mining*) adalah proses pencarian informasi yang bernilai di basis data yang besar, gudang data, atau *data mart*. Dalam penulisan tugas akhir ini, algoritma C4.5 diimplementasikan untuk pencarian pola klasifikasi mahasiswa yang tidak memenuhi sisip program. Sisip program adalah evaluasi hasil studi mahasiswa selama empat semester pertama untuk menentukan apakah mahasiswa dapat melanjutkan studi atau harus meninggalkan program studi yang bersangkutan. Data yang digunakan merupakan data penerimaan mahasiswa baru jalur tes dan latar belakang mahasiswa program studi Teknik Informatika Universitas Sanata Dharma dari tahun 2000 sampai dengan 2004. Data yang digunakan berisi informasi tentang jenis kelamin, nilai penalaran verbal, nilai kemampuan numerik, nilai penalaran mekanik, nilai hubungan ruang, nilai bahasa inggris, pilihan, gelombang, kabupaten SMA, jurusan SMA, keterangan SMA, dan status sisip program.

Program diuji dengan menggunakan teknik *5 fold cross validation* dengan sampel data sebanyak 355 data. Hasil akurasi program adalah 66,19 %.

ABSTRACT

Data mining is a process of finding valuable information in large databases, data warehouses, or data marts. In this thesis, C4.5 algorithm was implemented to find the classification patterns of students that do not fill “sisip program”. “Sisip program” is evaluation of students over four semesters of their studies to determine whether students can continue their studies or having to leave their study program. The data used is the regular student admissions data along with student academic data of the students of the Informatics Engineering Sanata Dharma University from year 2000 to 2004. Data used contain information about sex, the score of verbal reasoning test, numerical ability test, mechanical reasoning test, space relations test, English language test, chosen study program, registration periods, district high school, high school majors, school information, and status.

The program was verified using 5 fold cross validation technique towards 355 records. The accuracy of the program is 66.19 %.