

## ABSTRAK

Penambangan data merupakan suatu teknik untuk mengestraksi pola yang penting atau informasi yang menarik dari basis data yang berukuran besar. Penambangan data memberikan informasi yang digunakan sebagai penunjang dalam melakukan pengambilan keputusan. Teknik yang digunakan dalam melakukan penambangan data diantaranya adalah klasifikasi, yaitu teknik yang menentukan kelas-kelas ataupun kategori dari suatu objek berdasarkan sifat/atribut yang dimilikinya. Salah satu metode klasifikasi adalah algoritma pohon keputusan.

Dalam penulisan tugas akhir ini, algoritma pohon keputusan diimplementasikan untuk mengenali karakteristik calon mahasiswa yang tidak melakukan daftar ulang. Data-data yang digunakan merupakan data penerimaan mahasiswa baru Universitas Sanata Dharma dari tahun 2005 sampai dengan 2006. Seluruh data yang digunakan berisi informasi tentang gelombang masuk, prioritas pilihan program studi di mana calon mahasiswa tersebut diterima, jenis\_kelamin, jenis sma, program studi tempat calon mahasiswa diterima dan status daftar ulang yang dilakukan serta nilai akhirnya. Data-data ini digunakan sebagai himpunan data pelatihan dan himpunan data tes dalam proses penambangan data. Pengujian dilakukan terhadap himpunan data pelatihan dan himpunan data tes. Hasil pengujian menggunakan *fivefold cross-validation* dengan sampel data sebanyak 1400 *record* adalah 61.64 %.

## ABSTRACT

Data mining is a way to extract important pattern or interesting information from large databases. Data mining gives information that is used to make a decision. Classification is one of technique that can be applied on data mining. The technique will define classes or categories of an object based on its characteristics. One of classification technique is decision tree algorithm.

On this undergraduate thesis, the decision tree algorithm is applied to recognize characteristics of new student candidates of Sanata Dharma University that do not perform her-registration. The data that is used is new student candidates of 2005 to 2006 academic year, which consists of registration periods, the priority of applied study program, sex, the type of their high school, the study program where the new student are accepted, her-registration status and the final score of acceptance test. The data's are used as training set and test set on data mining process. The experiment is done to test training set and test set. The result of experiment with 1400 records using fivefold cross-validation method is 61.64 %.