

ABSTRAK

Data mining adalah sebuah proses mengekstrak pola yang penting atau menarik dari sejumlah data yang sangat besar. Salah satu tugas dalam data mining adalah klasifikasi, yaitu sebuah proses untuk menemukan model yang membedakan objek ke dalam kategori atau kelas yang sesuai, dengan tujuan untuk memperkirakan kelas dari objek yang kelasnya tidak diketahui sebelumnya. Ada beberapa metode untuk melakukan klasifikasi, salah satunya adalah klasifikasi dengan menggunakan pohon keputusan (*decision tree*).

Pada tugas akhir ini diimplementasikan salah satu algoritma yang digunakan untuk membuat pohon keputusan yaitu algoritma ID3 Quinlan. Implementasi algoritma ini akan digunakan untuk mengklasifikasikan latar belakang mahasiswa-mahasiswa yang memiliki nilai IPK ≥ 3.00 dan nilai IPK < 3.00 . Data mahasiswa diambil dari data Penerimaan Mahasiswa Baru dan data akademik mahasiswa Tahun Angkatan 1998 sampai dengan 2002, Program Studi Ilmu Komputer, Jurusan Matematika, Fakultas MIPA, Universitas Sanata Dharma. Data-data ini digunakan sebagai data *training set* dan data *test set*. Hasil dari implementasi tersebut diperoleh aturan-aturan untuk mengklasifikasikan mahasiswa-mahasiswa yang memiliki nilai IPK ≥ 3.00 dan nilai IPK < 3.00 , dengan tingkat prosentase kesalahan prediksi berkisar 20.45 % sampai 42.97 %.

ABSTRACT

Data mining is a process to extract interesting or important patterns from large amount of data. One of tasks in data mining is clasification. Clasification is a process to find model that differentiate objects into appropriate class or category, in order to predict the class of objects whose class is unkown. There are several methods of classification. One of them is decision tree.

This thesis implemented one of algorithms used to build a decision tree called ID3 Quinlan algorithm. The implementation will be used to classifly background of students whose $GPA \geq 3.00$ and $GPA < 3.00$. The data was collected from students acceptance data and students academic data year 1998-2002 of Computer Science Study Program, Mathematics and Natural Sciences Faculty, Sanata Dharma University. The data was used as training set and test set as well. The result were set of rules to classify students whose $GPA \geq 3.00$ and $GPA < 3.00$. based on several experiments with different composition of training set and test set, the best misprediction rate was 20.45 % , while the worst misprediction rate was 42.97 %.