ABSTRAK

Pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi kaya akan data tapi minim in*form*asi. *Data mining* merupakan penambangan atau penemuan in*form*asi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. *Data mining* sendiri memiliki beberapa teknik salah satunya klasifikasi. Teknik klasifikasi terdiri dari beberapa metode, dan *decision tree* adalah bagian dari metode klasifikasi. Kemudian metode *decision tree* memiliki beberapa algoritma, algoritma C4.5 adalah salah satu dari algoritma yang dimiliki *decision tree*. Dengan memanfaatkan data nilai akademik, diharapkan dapat menghasilkan in*form*asi tentang tingkat kelulusan mahasiswa melalui teknik *Data mining*. Kategori tingkat kelulusan diukur dari lama studi yang terdiri dari dua kategori yaitu <= 5 tahun dan > 5 tahun dan data nilai akademik yang digunakan adalah 6 mata kuliah yang diantaranya algoritma, fisika1, fisika2, kalkulus1, kalkulus2, dan pengantar sistem komputer. Dalam algoritma C4.5 terdapat dua buah data, yaitu data *training* dan juga data *testing*. Data *training* adalah data akan diolah dalam pembentukan algoritmanya yang kemudian akan di uji dengan data *testing* yang hasilnya berupa in*form*asi hasil prediksi masa studi. in*form*asi yang ditampilkan berupa prediksi masa studi permahasiswa dan persentase kecocokan hasil prediksi dengan data asli mahasiswa dalam jumlah yang banyak dengan data *training* yang berbeda.

Kata kunci : *Data mining*, Klasifikasi, *Decision tree*, Algoritma C4.5, Data *training*, Data *testing*.

ABSTRACT

The rapid growth of the accumulation of data has created the conditions for data-rich but information poor. Data mining is the mining or the discovery of new information by looking for certain patterns or rules of a number of large amounts of data are expected to treat the condition. Data mining itself has one of several classification techniques. Classification technique consists of several methods, and is part of the decision tree classification method. Then the decision tree method has some algorithm, the algorithm C4.5 is one of the decision tree algorithm that is owned. By utilizing the data of academic value, expected to yield information about the graduation rates of students through data mining techniques. Category graduation rate is measured from the period of study consisting of two categories: <= 5 years and> 5 years and the academic value of the data used is 6 courses that include algorithms, fisika1, fisika2, kalkulus1, kalkulus2, and introduction to computer systems. The C4.5 algorithm there are two data, ie data training and testing of data as well. Training data is the data will be processed in the establishment of the algorithm which will then be tested by testing the data which results in the prediction information of the study period. information presented in the form of predictions permahasiswa study period and the percentage of predicted results match with the original data in a large number of students with different training data.

Keywords: Data mining, classification, Decision tree, C4.5 algorithm, training data, testing data.