

ABSTRAK

Penelitian ini bertujuan untuk mencari pola klasifikasi lulusan siswa SMP yang tidak melanjutkan ke jenjang SMA dengan menggunakan algoritma C4.5, yaitu salah satu metode pohon keputusan yang ada dalam teori *data mining*. Data lulusan siswa SMP diperoleh dari 11 SMP pada dua kecamatan di Kabupaten Timor Tengah Utara, yaitu Kecamatan Kota Kefamenanu dan Kecamatan Miomaffo Timur, untuk tahun kelulusan 2008/2009, 2009/2010, 2010/2011. Data yang digunakan berisi informasi tentang informasi nilai ujian nasional, keterangan (lulus/ tidak lulus), jenis kelamin, usia siswa, anak ke berapa dalam saudara, jumlah saudara, pekerjaan orang tua, penghasilan orang tua, jarak rumah dengan SMA terdekat dan status melanjutkan / tidak melanjutkan. Sistem yang telah dibangun diuji dengan menggunakan *3-fold cross validation* dan *5-fold cross validation*. Tingkat akurasi total untuk keseluruhan data yang valid pada setiap tahun kelulusan serta data gabungan ketiga tahun dengan *3-fold cross validation* adalah 69,08%, 79,73%, 78,34%, 72,31% sedangkan tingkat akurasi dengan *5-fold cross validation* adalah 66,95%, 80,57%, 79,91% dan 76,74. Adapun tingkat akurasi berdasarkan data uji pada *fold* yang memiliki akurasi tertinggi pada *3-fold cross validation* sebesar 73,93%, 82,01%, 81,50% , 76,74% sedangkan tingkat akurasi tertinggi pada *5-fold cross validation* sebesar 73,04%, 83,13%, 82,19% serta 79,71%. Dapat disimpulkan bahwa atribut yang mempengaruhi siswa tidak melanjutkan sekolah adalah: 1) keterangan (lulus/tidaklulus), 2) pekerjaan orang tua yang terjadi secara beragam, 3) siswa yang memperoleh nilai terbaik pada beberapa mata pelajaran yang diujikan namun tidak melanjutkan sekolah, 4) siswa berjenis kelamin perempuan.

Kata Kunci: C4.5, *data mining*, *cross-validation*

ABSTRACT

This study aims to explore the pattern classification of graduate students who do not continue to junior high school level using the C4.5 algorithm, one of decision tree methods in the theory of data mining. Data obtained graduate junior high school students from 11 junior high schools in two districts in North Central Timor, the District and the District Miomaffo Kefamenanu City East for graduation year 2008/2009, 2009/2010, 2010/2011. The data used contains information about the national test scores, statement (passed / not passed), gender, age of the student, how many in your child, the number of siblings, parental employment, parental income, distance between home and school nearby and status continue / not continue. The system has been built tested using 3-fold cross validation and 5-fold cross validation. The accuracy rate of the overall total for valid data in each year of graduation and three years of data combined with 3-fold cross-validation was 69.08%, 79.73%, 78.34%, 72.31%, while the level of accuracy with 5-foldcross-validation was 66.95%, 80.57%, 79.91%, and 76.74. As for the accuracy of the test data based on the fold that have the highest accuracy on 3-fold cross-validation of 73.93%, 82.01%, 81.50%, 76.74%, while the highest level of accuracy in 5-fold cross validation by 73 , 04%, 83.13%, 82.19% and 79.71%. It can be concluded that the attributes that affect students do not attend school are: 1) information (pass / no pass), 2) parents work occurring variety, 3) students who obtain the best value on some subjects tested, but did not go to school, 4) female students.

Keywords: C4.5, *data mining*, *cross-validation*