

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

ABSTRAK

Dalam penerimaan mahasiswa baru Universitas Sanata Dharma (USD), seringkali jumlah mahasiswa yang mendaftar ulang lebih sedikit dari kapasitas penerimaan mahasiswa baru yang ditetapkan oleh USD. Namun sebaliknya dapat juga terjadi mahasiswa baru yang mendaftar ulang melebihi kuota. Untuk mengantisipasi masalah tersebut, USD perlu mengetahui prediksi calon mahasiswa baru yang tidak melakukan daftar ulang.

Banyak cara yang dapat digunakan untuk melakukan prediksi calon mahasiswa baru yang tidak melakukan daftar ulang. Salah satu cara yang digunakan adalah *data mining* menggunakan algoritma *Naive Bayesian*.

Dalam skripsi ini dibangun sebuah sistem untuk memprediksi apakah seorang mahasiswa baru akan mendaftar ulang atau tidak. Data yang akan digunakan dalam proses prediksi adalah data PMB USD tahun 2009 untuk pendaftaran melalui jalur reguler tes tertulis dari gelombang 1 sampai gelombang 3 meliputi prioritas pilihan pada program studi tempat calon mahasiswa diterima, gelombang masuk pendaftaran, jenis kelamin, jurusan SMA, program studi tempat diterima, status daftar ulang yang dilakukan dan nilai final tes masuk berdasarkan program studi. Data-data ini digunakan sebagai himpunan data pelatihan dan data pengujian. Data pelatihan digunakan untuk melakukan penghitungan nilai probabilitas prior dan *likelihood* sedangkan himpunan data pengujian digunakan untuk melakukan penghitungan nilai probabilitas posterior. Pengujian dilakukan menggunakan metode *fivefold cross-validation* dan *tenfold cross-validation* dengan data sebanyak 1304 baris. Prosentase keakuratan prediksi yang dihasilkan untuk metode *fivefold cross-validation* adalah 65,57% dan metode *tenfold cross-validation* 65,64%.

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

ABSTRACT

USD often has less new students than the standar capacity of the admission. Nevertheless, there is also a possibility of over quota. Henceforth, USD needs to anticipate the unpredictable situation.

Several ways could be done to predict the numbers of new students who do not enroll. One of them is data mining by using Naive Bayesian algorithm.

In this thesis, the writer developed a system to predict whether a new student will enroll or not. The data was taken from Sanata Dharma University Admission year 2009. The data consists of regular admission test which includes the priority of the choosen study program, the enrollment period, sex, the previous major in SMA, the study program in which the student is accepted, enrollment status and the final test score based on the choosen study program. The data were used as training data set and test data set as well. Training data set was used to calculate prior probability and likelihood while the test data set was used to calculate posterior probability. The system was test using fivefold cross-validation and tenfold cross-validation with 1304 data row. The accuracy of the prediction using fivefold cross-validation method is 65,57%, while the accuracy of the tenfold cross-validation method is 65,64%.