

ABSTRACT

CLASSIFICATION OF STUDENT SCORE OF LAMPUNG UNIVERSITY USING NAÏVE BAYES AND RANDOM FOREST ALGORITHM

By

REKTI NURUL HIDAYAH

At the tertiary level, at the end of each semester, students receive learning outcomes for each course taken. The end of student assessment in a course is based on the total score obtained from each learning achievement and the Final Semester Examination (UAS), which has a predetermined weight. Classification is a supervised learning method that seeks to identify correlations between input features and target features. Two classification methods were used in this study, namely the naïve bayes and random forest methods and compared the two methods so that they could be categorized as good classification methods. The stages in this research are data selection, data preprocessing, building and testing the model, and model evaluation. As for the results of this study, it was found that the naïve bayes and random forest methods had the highest level of accuracy when using the splitting data compared to using the k-fold cross validation. Based on the evaluation results of the model with the confusion matrix, the random forest method is the best method for classifying Lampung University student grades when compared to the naïve bayes method with an accuracy rate of 99.38%. Furthermore, the performance of the two methods is better if after boosting with Gradient Boosting (GB), namely for the naïve bayes method it is 99.89% and the random forest method is 99.45%.

Keywords: Student, Final Grade, Classification, Naïve Bayes, Random Forest.

ABSTRAK

KLASIFIKASI NILAI MAHASISWA UNIVERSITAS LAMPUNG MENGUNAKAN ALGORITMA *NAÏVE BAYES* DAN *RANDOM FOREST*

Oleh

REKTI NURUL HIDAYAH

Ditingkat perguruan tinggi, setiap akhir semester, mahasiswa mendapatkan hasil belajar untuk setiap mata kuliah yang diambil. Akhir dari penilaian mahasiswa pada suatu mata kuliah didasarkan pada total nilai yang diperoleh dari setiap pencapaian pembelajaran dan Ujian Akhir Semester (UAS), yang memiliki bobot yang telah ditetapkan. Klasifikasi merupakan metode *supervised learning* yang berusaha untuk mengidentifikasi korelasi antara fitur input dan fitur target. Dua metode klasifikasi digunakan dalam penelitian ini yaitu metode *naïve bayes* dan *random forest* serta membandingkan kedua metode tersebut agar dapat dikategorikan sebagai metode klasifikasi yang baik. Tahapan dalam penelitian ini yaitu seleksi data, *preprocessing* data, membangun dan menguji model, serta evaluasi model. Adapun hasil penelitian ini diperoleh bahwa metode *naïve bayes* dan *random forest* memiliki tingkat akurasi tertinggi jika menggunakan *splitting* data dibandingkan dengan menggunakan *k-fold cross validation*. Berdasarkan hasil evaluasi model dengan *confusion matrix*, metode *random forest* merupakan metode terbaik untuk mengklasifikasi nilai mahasiswa Universitas Lampung jika dibandingkan dengan metode *naïve bayes* dengan tingkat akurasi sebesar 99,38%. Selanjutnya untuk performa dari kedua metode lebih baik jika sesudah dilakukan *boosting* dengan *Gradient Boosting* (GB), yaitu untuk metode *naïve bayes* sebesar 99,89% dan metode *random forest* sebesar 99,45%.

Kata Kunci: Mahasiswa, Nilai Akhir, Klasifikasi, *Naïve Bayes*, *Random Forest*.