

ABSTRAK

Penerimaan mahasiswa baru melalui jalur rapor, tentunya memiliki kriteria tersendiri agar calon mahasiswa dapat masuk ke perguruan tinggi. Berdasarkan nilai rapor tersebut diharapkan mampu menggambarkan calon mahasiswa yang dapat mengikuti pelajaran sehingga mampu mendapatkan prestasi yang maksimal. Masalah dalam penelitian ini yaitu mencari tahu apakah nilai rapor SMA dan SMK dapat dipergunakan untuk memprediksi prestasi akademik mahasiswa. Tujuan penelitian ini adalah untuk mengklasifikasi prestasi akademik mahasiswa berdasarkan nilai rapor menggunakan algoritma *Random Forest*, mengetahui akurasi terbaik yang didapatkan dari kinerja algoritma *Random Forest* dalam klasifikasi prestasi akademik mahasiswa berdasarkan nilai rapor, serta mengetahui variabel atau atribut yang mempengaruhi hasil akurasi tertinggi. Data yang digunakan adalah data mahasiswa angkatan 2016-2018 dari beberapa program studi di Universitas Sanata Dharma yang diterima melalui jalur nilai rapor. Penelitian ini akan dilakukan dengan penggunaan teknik klasifikasi menggunakan metode *Random Forest*. Dalam penelitian ini dilakukan eksperimen dengan memvariasi jumlah *tree* dan nilai *k* dalam *k-fold cross validation*. Di samping itu juga dilakukan perbandingan hasil akurasi terhadap dataset yg dikenai deteksi *outlier* dengan dataset yang tidak dikenai deteksi *outlier*. Hasil akurasi tertinggi didapatkan dengan menggunakan *dataset* Program Studi Akuntansi, Manajemen, Pendidikan Bahasa dan Sastra Indonesia, dan Sastra Inggris yang menghasilkan akurasi sebesar 100%. Hasil akurasi tersebut didapatkan dengan dan tanpa penerapan deteksi *outlier* untuk dataset prodi Manajemen dan Akuntansi. Sementara itu hasil akurasi untuk dataset prodi Pendidikan Bahasa dan Sastra Indonesia didapatkan dengan penerapan deteksi *outlier*, sedangkan untuk dataset prodi Sastra Inggris didapatkan tanpa penerapan deteksi *outlier*. Penggunaan *dataset* Gabungan semua prodi menghasilkan akurasi terbaik sebesar 64.71% dengan penerapan deteksi *outlier*, sedangkan penggunaan *dataset* SMA yang berisi nilai rapor SMA tanpa ada atribut Prodi menghasilkan akurasi terbaik sebesar 62.77% tanpa penerapan deteksi *outlier*. Dalam proses uji data tunggal, model yang dihasilkan dalam penggunaan *dataset* gabungan dan *dataset* SMA dapat digunakan pada uji data tunggal untuk semua prodi. Akan tetapi, model yang dihasilkan dari penggunaan *dataset* prodi tidak dapat digunakan untuk uji data tunggal untuk semua prodi.

Kata Kunci : Prediksi, Algoritma *Random Forest*, Nilai Rapor

ABSTRACT

Admission of new students through report cards, of course, has its own criteria so that prospective students can enter college. Based on the value of the report card, it is expected to be able to describe prospective students who can take lessons so that they are able to get maximum achievement. The problem in this study is to find out whether high school and vocational school report cards can be used to predict student academic achievement. The purpose of this study was to classify student academic achievement based on report cards using the Random Forest algorithm, find out the best accuracy obtained from the performance of the Random Forest algorithm in classifying student academic achievements based on report cards, and determine the variables or attributes that affect the highest accuracy results. The data used is student data from the 2016-2018 class from several study programs at Sanata Dharma University who were accepted through the report card grade. This research will be conducted using a classification technique using the Random Forest method. In this study, experiments were carried out by varying the number of trees and the value of k in k -fold cross validation. In addition, a comparison of the accuracy results for datasets subject to outlier detection is also carried out with datasets that are not subject to outlier detection. The highest accuracy results were obtained using the dataset of the Accounting, Management, Indonesian Language and Literature Education Study Program, and English Literature which resulted in an accuracy of 100%. The accuracy results are obtained with and without the application of outlier detection for the Management and Accounting study program dataset. Meanwhile, the accuracy results for the Indonesian Language and Literature Education Study Program dataset were obtained by applying outlier detection, while the English Literature study program dataset was obtained without the outlier detection application. The use of the combined dataset of all study programs produces the best accuracy of 64.71% with the application of outlier detection, while the use of the SMA dataset which contains SMA report cards without any study program attributes produces the best accuracy of 62.77% without the application of outlier detection. In the single data test process, the model generated using the combined dataset and SMA dataset can be used in a single data test for all study programs. However, the model generated by using the study program dataset cannot be used to test a single data for all study programs.

Keywords : Prediction, Random Forest Algorithm, Report Value