

## ABSTRAK

Jessica Esther Sutedja (00000006937)

### **PREDIKSI NILAI MAHASISWA MENGGUNAKAN TEKNIK-TEKNIK DATA MINING**

Skripsi, Fakultas Sains dan Teknologi (2019).

(xxiv + 74 halaman, 182 tabel, 38 gambar, 5 lampiran)

*Big data* adalah salah satu topik yang paling sering didiskusikan seiring perkembangan teknologi. Oleh karena itu, banyak metode dikembangkan untuk mendapatkan sebanyak mungkin informasi yang tersedia di set data yang kompleks. Salah satu penerapan dari metode-metode yang multiguna ini adalah pada bidang edukasi.

Di penelitian ini, teknik-teknik *data mining* akan diterapkan pada data yang dikumpulkan di Universitas Pelita Harapan untuk membangun model-model yang dapat memprediksi nilai akhir mahasiswa-mahasiswa tahun pertama dan kedua yang mengikuti mata kuliah Kalkulus dan Statistika. Teknik-teknik yang digunakan adalah Naïve Bayes, *C5.0 Decision Tree*, *Classification and Regression Tree (CART)* dan *Support Vector Machine (SVM)*.

Model-model yang didapat dari metode-metode ini telah diringkas dan dianalisa. Hasilnya menunjukkan bahwa SVM dan *decision tree* memiliki performa yang mirip satu dengan yang lain di berbagai kondisi, tetapi *decision tree* adalah model yang lebih mudah dimengerti. Naïve Bayes yang tidak menggunakan tak-siran Laplace memiliki performa yang lebih bagus dibandingkan dengan ketika tak-siran Laplace digunakan untuk membangun model, dengan akurasi dan nilai kappa yang lebih tinggi. Selain ujian-ujian, tugas dan kuis, jurusan, angkatan, kepercayaan diri dan kepuasan mahasiswa ketika menghadiri mata kuliah juga dapat memberikan efek pada performa mahasiswa di mata kuliah tersebut.

Kata Kunci: *big data, data mining, classification*

Referensi: 15 (1993-2017)

## **ABSTRACT**

Jessica Esther Sutedia (00000006937)

### **STUDENT GRADE PREDICTION USING DATA MINING TECHNIQUES**

Thesis, Faculty of Science and Technology (2019).

(xxiv + 74 pages, 182 tables, 38 figures, 5 appendix)

Big data is one of the most discussed topic along with the development of technology. As a result, many methods have been developed and used in order to extract as many useful information as possible from complex data sets. The versatile use of these techniques also bring a lot of interest in using them in various fields, one of them being educational field.

This thesis involves applying these data mining techniques to data sets collected in Universitas Pelita Harapan to create models for predicting students final grade in Calculus and Statistics courses taken by first and second year students. The techniques used are Naïve Bayes, C5.0 Decision Tree, Classification and Regression Tree (CART), and Support Vector Machine (SVM).

Models produced by these methods are summarized and analyzed. The results indicate that while SVM and decision tree models have very similar performances, decision tree models are more interpretable. Naïve Bayes with no Laplace estimator perform well compared to when Laplace estimator is used, producing generally higher test accuracy and kappa values. Aside from exams, quizzes, and assignments, the result shows that students' major, batch, confidence and satisfaction in classes taken may affect the students' performance in the courses.

Keywords: big data, data mining, classification

References: 15 (1993-2017)