

ABSTRAK

Nama : Ariana Yunita
Program Studi : Doktor Ilmu Komputer
Judul : Analisis Data Prestasi Akademik Siswa Menggunakan Pemelajaran Mesin
Pembimbing : Harry Budi Santoso, Ph.D
Prof. Zainal A. Hasibuan, Ph.D
Prof. Dr.Eng. Wisnu Jatmiko

Selama ini upaya deteksi keberhasilan dan kegagalan cukup terlambat. Hal ini mengakibatkan mahasiswa yang akhirnya gagal, walaupun mereka memiliki potensi, kehilangan banyak kesempatan. Oleh karena itu, prediksi prestasi akademik diperlukan di periode awal masa studi untuk menghindari kegagalan mahasiswa. Penelitian mengenai prediksi yang merupakan salah satu domain kajian dari *Learning Analytics* yang telah dilakukan oleh para peneliti. Namun, penelitian ini tetap menjadi tantangan karena banyak faktor yang memengaruhi mahasiswa, dan saat ini data tersebar di mana-mana. Untuk mengatasi masalah tersebut, penelitian ini bertujuan untuk memprediksi dan memberi rekomendasi untuk meningkatkan prestasi akademik mahasiswa. Penelitian ini mengadopsi “*End-To-End of Life Cycle Automation System*” sebagai kerangka penelitian dengan “*Everything is Data*” sebagai ide utamanya. Dengan gagasan ini, data dapat dikumpulkan, diproses dan menghasilkan pola dan wawasan yang berarti bagi setiap pemangku kepentingan untuk mengambil keputusan. Studi ini mengumpulkan data mahasiswa dan institusi pendidikan tinggi dari berbagai sumber, termasuk *database* pendidikan tinggi nasional, *data crawling*, dan *database* terkait lainnya. Studi ini menghasilkan faktor mahasiswa yang berkontribusi terhadap prestasi akademik dari literatur dan dari praktik terbaik. Faktor-faktor yang berkontribusi terhadap prestasi akademik mahasiswa dari sisi teoretis didefinisikan sebagai delapan faktor: demografi, pembelajaran mandiri, teknologi, keterlibatan kursus mahasiswa, sosial budaya, aktivitas di kelas, sistem pendidikan, dan kepribadian. Selanjutnya, dari sisi praktis, studi ini menghasilkan cikal bakal ekosistem *big data* pendidikan tinggi di Indonesia. Data mentah yang digunakan adalah data mahasiswa dari 15 perguruan tinggi di Indonesia, dari tahun 2011 hingga 2019, mewakili wilayah Indonesia Timur, Indonesia Tengah, dan Indonesia Barat. Studi ini juga menghasilkan profil mahasiswa yang dihasilkan dari ekosistem *big data* yang diberi label menjadi: Mahasiswa Berisiko Sangat Tinggi, Mahasiswa Berisiko Tinggi, Mahasiswa Gagal, Mahasiswa Berisiko Berfluktuasi, Mahasiswa Berisiko Sedang, Mahasiswa Berisiko Rendah, dan Mahasiswa Berisiko Sangat Rendah. Prediksi temporal juga dilakukan, dan dari semester pertama hingga semester keempat menggunakan Random Forest dengan tingkat akurasi antara 83-92%. Berdasarkan struktur pohon dan ciri-ciri penting yang dihasilkan pada tahap penelitian sebelumnya, rekomendasi dirumuskan bagi pemangku kepentingan terkait: perguruan tinggi, dan pimpinan program studi atau dosen. Ulasan pakar juga dilakukan untuk melengkapi analisis. Model deskriptif, prediktif dan preskriptif ini diharapkan dapat bermanfaat bagi pemangku kepentingan terkait untuk secara pribadi melakukan intervensi pada

mahasiswa yang berisiko dan mendukung mahasiswa dengan prestasi akademik yang lebih tinggi. Studi ini menunjukkan bahwa beberapa fitur institusi, seperti jumlah publikasi di setiap institusi dan kualitas dosen, dapat berkontribusi pada prestasi akademik mahasiswa.

Kata kunci: data berskala besar, penggalian data, pembelajaran mesin, Pendidikan tinggi, model prediksi, preskripsi

ABSTRACT

Name : Ariana Yunita
Study Program : Doctorate of Computer Science
Title : Data Analytics on Student's Academic Achievement Using Machine Learning
Supervisor : Harry Budi Santoso, Ph.D
Prof. Zainal A. Hasibuan, Ph.D
Prof. Dr.Eng. Wisnu Jatmiko

During this time, university students who pass or fail are being detected at the end of the academic period. By that time, helping them is too late. As a result, the failed university students lose a lot of opportunities. Furthermore, it is common practice for government, institutions, or lecturers to apply "one-size-fits-all" policies and strategies, even each student has diverse characteristics. Therefore, anticipating proper action to students' achievement and predicting students' academic achievement is required in advance to avoid students' failure. Research regarding prediction, which is one of the fields of Learning Analytics, is common. However, this research remains a challenge since many factors influence students, and currently, in a big data era, university students' data are scattered everywhere. To overcome these problems, this research aims to predict and prescribe students' academic achievement. This study adopts the "End-To-End of Life Cycle Automation System" as a research framework. The main idea is "Everything is Data" and the data can be collected, processed and produce patterns and meaningful insights for any stakeholder to make decisions. This study collected students' and higher education institutions' data from multiple sources, including the national higher education database, data crawling, other related databases, etc. This study results student's factors contributing to academic achievement from literature and from best practices. Factors contributing to student's academic achievement from the theoretical side were defined as eight factors: demography, self-regulated learning, technology, student course engagement, socio culture, activity in classroom, educational system, and personality. Furthermore, from the practical side, this study results a preliminary big data ecosystem of higher education in Indonesia. Raw data were student's data from fifteen universities in Indonesia, from 2011 until 2019, represent East Indonesia, Central Indonesia, and West Indonesia. This study also results in students' profiles generated from the big data ecosystem that were labeled into: Very High-Risk Students, High-Risk Students, Fail Students, Fluctuating Risk Students, Moderate Risk Students, Low-Risk Students, and Very Low-Risk Students. The temporal prediction was also conducted, and from the first semester until the fourth semester, using Random Forest, the accuracy was between 83-92%. Based on the tree structure and important features resulting in the previous research stage, the prescription was created for related stakeholders: Higher Education institutions, and the head of study programs or lecturers. Experts review was also conducted to strengthen and complete the analysis. The descriptive, predictive and prescriptive model expectedly might be useful for related stakeholders to personally intervene in at-risk students and support students with higher academic achievement. This study implies that several features of the institution, such as the number of publications in each institution and the quality of the lecturer, might contribute to students' academic achievement.

Keywords:

big data, Indonesia, data mining, machine learning, higher education, predictive model, prescription