

Deteksi Dini Mahasiswa Berisiko Gagal Berdasarkan Data Sistem Informasi Mahasiswa dan Aktivitas Tutorial Online Studi Kasus: Universitas Terbuka = Student At-Risk Failure Early Detection Based on Student Information System and Online Tutorial Activities Data Case Study: Universitas Terbuka

Fitria Amastini, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20514089&lokasi=lokal>

Abstrak

Universitas Terbuka (UT) menyediakan student support services untuk meningkatkan hasil pembelajaran mahasiswa dan persistensi mahasiswa untuk tetap menyelesaikan studinya di Pendidikan Jarak Jauh. Namun, fakta lapangan menunjukkan rata-rata IPK dan IPS mahasiswa Sarjana dan Diploma Angkatan 20161 s/d 20182 masih di bawah standar IPK tuntutan pasar kerja (2.75). Solusi permasalahan tersebut adalah mendeteksi dini mahasiswa berisiko gagal menggunakan metode klasifikasi data mining berdasarkan data aktivitas Tutorial Online (Tuton) dan data pribadi mahasiswa. Pengklasifikasian mahasiswa berisiko gagal berdasarkan nilai IPS agar dapat mendeteksi lebih awal tidak hanya di semester awal tetapi juga di semester berikutnya. Selain itu, nilai IPS memiliki korelasi positif yang kuat terhadap nilai IPK sehingga nilai IPS dianggap dapat sebagai indikasi awal dari risiko kegagalan. Algoritma klasifikasi untuk model deteksi dini mahasiswa berisiko menggunakan naïve bayes, logistic regression, SVM, decision tree (CART, C5.0), random forest, dan adaboost. Tahap awal pengujian model menggunakan data aktivitas Tuton masa 20182-20191. Pembagian data training dan data testing menggunakan Stratified K-fold sebanyak 10 kali iterasi dan melakukan eksperimen metode tanpa sampling class imbalance dan metode random undersampling (50P:50N, 70P:30P, 66P:33P, 60P:40N) pada data training. Pada tahap awal pengujian model menunjukkan F1-score di minggu ke-empat tidak berbeda signifikan dengan minggu ke-delapan sehingga dianggap sebagai waktu yang tepat untuk mengintervensi lebih awal agar mahasiswa dapat berjuang di tugas berikutnya. F1-score tertinggi dari tahap awal pengujian model adalah tanpa sampling class imbalance di data training dengan algoritma random forest (90.20%), adaboost (89.20%), dan decision tree CART (88.10%). Ketiga algoritma terbaik akan diuji kembali pada tahap akhir menggunakan data testing aktivitas Tuton masa 20192. Hasil tahap akhir kinerja model deteksi dini mahasiswa berisiko kegagalan berdasarkan F1-score menunjukkan algoritma adaboost dengan nilai tertinggi (84.7%) diikuti oleh algoritma random forest (83.8%). Berdasarkan pengukuran recall, CART menunjukkan nilai tertinggi (99.9%) tetapi mengalami overfitting terhadap kelas positif sehingga tidak lebih baik dibandingkan melakukan intervensi ke seluruh mahasiswa. Kinerja terbaik untuk model deteksi dini mahasiswa berisiko gagal di UT adalah menggunakan algoritma adaboost.

.....Universitas Terbuka (UT) provides student support services to improve student academic outcomes and student persistence for their completion in Distance Education. However, the average cumulative and semester GPA of Bachelor and Diploma programs from academic year 20161-20182 show below labor market standard GPA (2.75%). Solution to this problem is early detection on academic failure risk through the implementation of classification data mining to predict student at-risk academic failure using Online Tutorial (Tuton) activities data and student's personal information. Classification student at-risk academic failure based on their semester GPA in order to early detect not only on the initial semester but also on the

following semester. Furthermore, semester GPA has a strong positive correlation to cumulative GPA so that semester GPA is considered as an early indication of the risk of failure. The classification algorithm for student at-risk failure early detection model using naïve bayes, logistic regression, SVM, decision tree (CART, C5.0), random forest, and adaboost. The initial model testing stage use data from Tuton activities on 20182-20191. Splitting method of training data set and testing data set using Stratified K-fold in 10 times iteration and experimenting without class imbalance sampling and random undersampling method (50P:50N, 70P:30P, 66P:33P, 60P:40N) on training data set. On The initial model testing stage shows that F1-scores on fourth week are not significantly different from the eighth week so early intervention on fourth week is the right time for student to study harder on the next assignments. The highest F1-score from the initial model testing stage is without sampling imbalance on training data set using random forest (90.20%), adaboost (89.20%), and decision tree CART (88.10%). The three best algorithms will be tested again on the final testing stage using Tuton activity on 20192 as testing data set. The F1-score results on the final student at-risk of failure early detection model stage shows that adaboost algorithm highest performance (84.7%) and followed by random forest (83.8%). Based on recall results, CART showed the highest performance (99.9%) but tend to positive class overfitting so that it was no better than intervening all of students. The best performance for student at-risk of failure early detection models at UT is using adaboost algorithm.