

ABSTRAK

NIM : 41518010150
Nama : Rani Aprilya Putri
Pembimbing TA : Dr. Nenden Siti Fatonah
Judul Tugas Akhir : Perbandingan Metode Klasifikasi serta Analisis Faktor Akademis Pola Kelulusan Mahasiswa di Perguruan Tinggi

Perkembangan Teknologi informasi di berbagai bidang diikuti dengan berkembangnya data. Pangkalan data yang menyimpan data pengelolaan pelaksanaan pendidikan tinggi dari seluruh perguruan tinggi yang terintegrasi secara nasional yaitu PDDIKTI (Pangkalan Data Pendidikan Tinggi). Data PDDIKTI dapat digunakan untuk mendapatkan informasi yang bermanfaat. Banyaknya pemanfaatan data menyebabkan adanya cabang ilmu baru yang dapat menangani masalah informasi serta pola yang penting atau menarik dari sejumlah data yang besar, yaitu data mining. Dalam penelitian ini akan menerapkan data mining untuk mengetahui klasifikasi pola akademik kelulusan mahasiswa yang menjadi pengaruh besar dalam proses akreditasi. Pada penelitian ini penulis akan membandingkan metode klasifikasi Machine Learning dan Ensemble Learning, mengetahui faktor yang paling berpengaruh signifikan terhadap kelulusan serta analisis data pola akademik mahasiswa. Adapun metode algoritma yang digunakan untuk pengujian adalah *Decision Tree*, *Ensemble Learning Bagging (Random Forest)* dan *Ensemble Learning Boosting (SGB)*. Dan hasil proses pengujian klasifikasi pola data kelulusan mahasiswa dengan hasil akurasi terbaik adalah menggunakan metode algoritma *Ensemble Learning Bagging* atau *Random Forest* dengan melakukan *cross validation* dan *hyperparameter tuning (GSCV)* dengan akurasi 96.1%. Penggunaan *cross validation* dan *hyperparameter tuning* terbukti dapat mempengaruhi dan mengoptimalkan akurasi learning. Faktor yang paling mempengaruhi pola kelulusan mahasiswa adalah jumlah SKS, Total Cuti, IPK, IPS, Angkatan, Program Kelas dan Jurusan. Dan untuk data yang digunakan dalam penelitian ini menunjukkan persentase kelulusan mahasiswa yang kurang baik karena persentase mahasiswa yang lulus tidak tepat waktu perbandingannya di atas 50% pada 3 data angkatan 2016-2018.

Kata kunci:

Machine Learning, Ensemble Learning, Klasifikasi, PDDIKTI.

ABSTRACT

Name : Rani Aprillya Putri
Student Number : 41518010150
Counsellor : Dr. Nenden Siti Fatonah
Title : Perbandingan Metode Klasifikasi serta Analisis Faktor Akademis Pola Kelulusan Mahasiswa di Perguruan Tinggi

The development of information technology in various fields is followed by the development of data. The database that stores data on the management of the implementation of higher education from all nationally integrated universities is PDDIKTI (Higher Education Database). PDDIKTI data can be used to obtain useful information. The large number of data uses has led to a new branch of science that can deal with information problems and patterns that are important or interesting from large amounts of data, namely data mining. In this study, data mining was applied to determine the classification of student graduation academic patterns which became a major influence in the accreditation process. In this study, the authors will compare the classification methods of Machine Learning and Ensemble Learning, find out the factors that have the most significant influence on graduation and analyze data on student academic patterns. The algorithm methods used for testing are Decision Tree, Ensemble Learning Bagging (Random Forest) and Ensemble Learning Boosting (SGB). And the results of the process of testing the classification of student graduation data patterns with the best accuracy results using the Ensemble Learning Bagging or Random Forest algorithm method by doing cross validation and hyperparameter tuning (GSCV) with an accuracy of 96.1%. The use of cross validation and hyperparameter tuning is proven to influence and optimize learning accuracy. The factors that most influence the student's graduation pattern are the number of Credits, Total Leave, GPA, Social Studies, Academic Year, Class Programs and Majors. And for the data used in this study, the percentage of student graduation is not good because the percentage of students who graduate not on time is above 50% in the 3 data for the 2016-2018 batch.

Key words:

Machine Learning, Ensemble Learning, Classification, PDDIKTI.