

Abstract

On-time graduation is an achievement for both students and university. For universities, especially study programs, the percentage of graduation on-time is a crucial point for accreditation. Therefore, it requires a good strategy. The percentage of graduation on time in college can be predicted with data mining and machine learning. The objective of this research is to provide earlier information regarding students who are at risk of not graduating on time. Thus, the study program can take appropriate action before it is too late. There are several classification methods that can be used for prediction. Our research combines Naïve Bayes with Principal Component Analysis (PCA). PCA is used to simplify the complex academic data. The PCA result has simpler structure then to be processed using Naive Bayes classification. This research uses data obtained from four batches of student academic performance data in Informatics Study Program, Telkom University. The dataset is partitioned by academic year to obtain time series data of each student. The combination of PCA and Naïve Bayes algorithms obtained better result compared to classification using Naïve Bayes only, with on average 6.04% higher accuracy.

Keywords: classification, graduation, machine learning, student, naïve bayes, principal component analysis.