

***Thesis Title : Prediction Modeling Students Dropout using Machine Learning  
Method in Higher Education***

## **ABSTRACT**

*In this era, as well as education, drop out is also an attractive topic for students, institutions or the government. This case becomes very important because dropouts can affect the graduate ability and university quality. Various studies have applied machine learning methods to find and identify dropout indicators, especially in higher education. In this study, the Learning Vector Quantization (LVQ) algorithm was used to select the features and analyze based on the importance of the two groups of variables namely demographic variables and academic performance. Furthermore, the K-Nearest Neighbor (K-NN), Naïve Bayes (NB), and Decision Tree (DT) methods are applied to perform predictive modeling process. In addition, the Ensemble Classifier method combines those single-classifier methods and use Gradient Boosting as a comparison method for finding the best modeling solution. The modeling process has been tested and evaluate with 10-fold cross-validation method and generate recall rate of 64.29% (NB), 64.84% (DT), 75.27% (KNN) while the Ensemble Stacking Classification with Gradient Boosting method is greater about 79.12% with the best accuracy rate of 98.82%. The variables includes student attendance, assignment scores, mid-test scores, final scores, total credits, GPA, student's area, parents' income, parent's education, gender and age were found as predictors of student dropouts.*

***Keywords:*** *machine learning, prediction modeling, dropout, ensemble stacking classification*

## Judul Tesis: Pemodelan Prediksi Mahasiswa Drop Out Menggunakan Metode Machine Learning Pada Perguruan Tinggi

### ABSTRAK

Di era saat ini, sama halnya dengan pendidikan, *dropout* juga menjadi topik yang menarik bagi mahasiswa, institusi ataupun pemerintah. Kasus ini menjadi sangat penting dikarenakan *dropout* dapat mempengaruhi kemampuan lulusan dan kualitas universitas. Beragam penelitian sudah menerapkan metode *machine learning* untuk menemukan dan mengidentifikasi indikator terjadinya dropout khususnya di perguruan tinggi. Pada penelitian ini, algoritma *Learning Vector Quantization* (LVQ) digunakan untuk memilih fitur yang tepat dan menganalisa berdasarkan *importance value* dari dua kelompok variabel yaitu variabel demografik dan performansi akademik. Selanjutnya, metode *K-Nearest Neighbor* (K-NN), *Naïve Bayes* (NB), dan *Decision Tree* (DT) diterapkan pada proses pemodelan prediksi.

Selain itu, metode *Ensemble Classifier* akan menggabungkan tiga metode tersebut dengan menerapkan *Gradient Boosting* sebagai metode pembanding untuk menemukan solusi pemodelan terbaik. Proses pemodelan sudah diuji dengan metode evaluasi *10-fold cross-validation* dan menghasilkan nilai *recall rate* sebesar 64,29 % (NB), 64,84% (DT), 75,27% (KNN) sementara metode *Ensemble Stacking Clasification* menggunakan *Gradient Boosting* lebih baik sekitar 79,12% dengan nilai akurasi terbaik yaitu 98,82%. Indikator yang digunakan adalah persentase kehadiran mahasiswa, nilai tugas, nilai UTS, nilai UAS, total pencapaian SKS, IPK, daerah asal siswa, pendapatan orang tua, tingkat pendidikan orang tua, jenis kelamin dan usia sebagai prediktor putus sekolah siswa.

**Kata kunci:** *machine learning, pemodelan prediksi, dropout, ensemble stacking classification*