## horizontal line



College Retention Rate Predictor

09.25.2016

**─**

Trunojoyo Anggara, Ph.D.

# Overview

My project is to mine the College Scorecard database and see the what are the variables affecting 6-year college graduation rate.

# Goals

1. Figuring out variables that infer or predict undergraduate student retention rate
2. Applying various machine learning classifiers: Naive Bayes, Neural Network, Logistic Regression, and decision tree/random forests into the various variables available through the College Scorecard database

# Specifications

The investigation will be done using the 2012 data from the College Scorecard. I am splitting the dataset into 70% training set and 30% testing set. I am categorizing the 6 year graduation rate into 6 categories:

1. Higher than 70%
2. From 60-69.99%
3. From 50-59.99%
4. From 40-49.99%
5. From 30-39.99%
6. Less than 30%

# Milestones

## Cleaning up dataset

The first milestone will be in cleaning up the variables that will be used as predictors for various colleges retention rate.

## Final Report

The report will be in the form of showing the accuracy of various machine learning classifiers in predicting the category of each university based on the various variables that are affecting the 6-year college graduation rate.