

# Machine Learning Engineer Nanodegree Capstone Project

---

## Retention and Graduation Rate in U.S. Postsecondary Education

Kai-Sheng Wang

April 23, 2020

### Input Data

- The raw dataset ("Most-Recent-Cohorts-Scoreboard-Elements") was downloaded in April 2020 from the "College Scorecard" page on Data.gov. <https://catalog.data.gov/dataset/college-scorecard>
- A dictionary that describes the elements in the dataset is available on the "College Scorecard" page on the website of U.S. Department of Education. <https://collegescorecard.ed.gov/data/documentation/>
- The raw dataset contains 6806 rows and 190 columns

### Software and Libraries Used

- python 3.7.4
- numpy 1.16.5
- pandas 0.25.1
- matplotlib 3.1.1
- scikit-learn 0.21.3

### List of Files Submitted for Final Project

- proposal.pdf (project proposal)
- report.pdf (project report)
- retention\_and\_graduation\_rate\_in\_us\_postsecondary\_education.ipynb (development python codes used for the project)
- README.pdf
- scorecard\_data.csv (raw input data)
- CollegeScorecardDataDictionary.xls (dictionary of elements in the dataset)