Understanding Low Completion Rates at Community Colleges

William C. Webb

Community Colleges Increase Access to Higher Education & Economic Mobility

- ~Half of low-income Americans first enroll in a community college to pursue higher education
- High school graduates face much lower earning potentials without a post-graduate certificate or degree[1,2]

1. Belfield, C. and T. Bailey. 2017. The Labor Market Returns to Sub-Baccalaureate College: A Review
2. Mann-Levesque, L. 2019. Improving community college completion rates by addressing structural and motivational barriers.

Community Colleges Train Skilled Workers

- Many skilled occupations require an Associate's Degree or certificate for entry-level positions
- Demand for skilled workers on the rise
- Growth rate for 43 skilled occupations projected to rise[3]

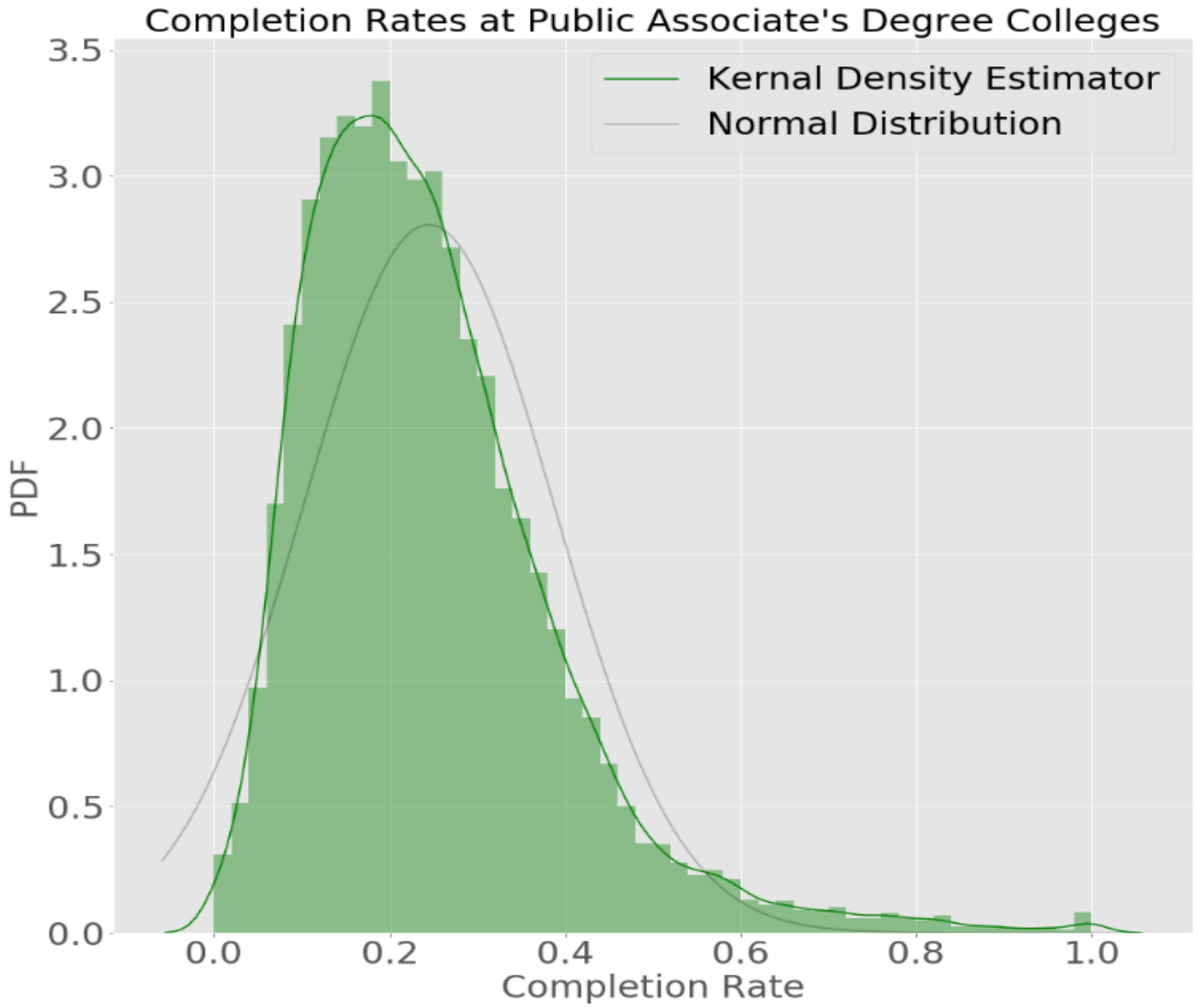
3. United States Bureau of Labor Statistics. 2019. Occupational Outlook Handbook.

Community Colleges Have Low Completion Rates

- First-time degree/certificate-seeking student completion rates within six years:
    - Community colleges: 23% completion rate
    - Four-year public colleges: 47% completion rate[4]

4. National Center for Education Statistics. 2019. Digest of Education Statistics.

Completion Rates Vary Across Community Colleges



Why Do Completion Rates Vary?

- Possible reasons:
    - Student demographics
    - Quality of instruction
    - Level of student services
    - Others?
- Explaining factors related to variation could help community colleges increase completion rates

Who Might Care About Completion Rates?

- Prospective students
- Families of prospective students
- Admissions departments
- Faculty

- Recruiting departments
- Curriculum designers
- Administrators

College Scorecard

- Published by U.S. Dept. Education[5]
- Annual survey of all colleges offering U.S. financial aid
- >2200 categories of data collected per college
- Data include different measures of completion rate

5. U.S. Department of Education. 2019. College Scorecard Data.

Explaining Variation in Completion Rate

- Used College Scorecard data to investigate variation in completion rates
- Completion rate measure used:
  - "Completion rate for first-time, full-time students at less-than-four-year institutions (150% of expected time to completion"

Methods – Data Wrangling

- Accessed 2.46 GB College Scorecard Data
- 22 academic years, 1996-97 through 2017-18
- Jupyter notebooks, Python 3.7.4, pandas and other libraries
- Original pandas DataFrame contained 154,228 rows/cases and 1977 columns/features

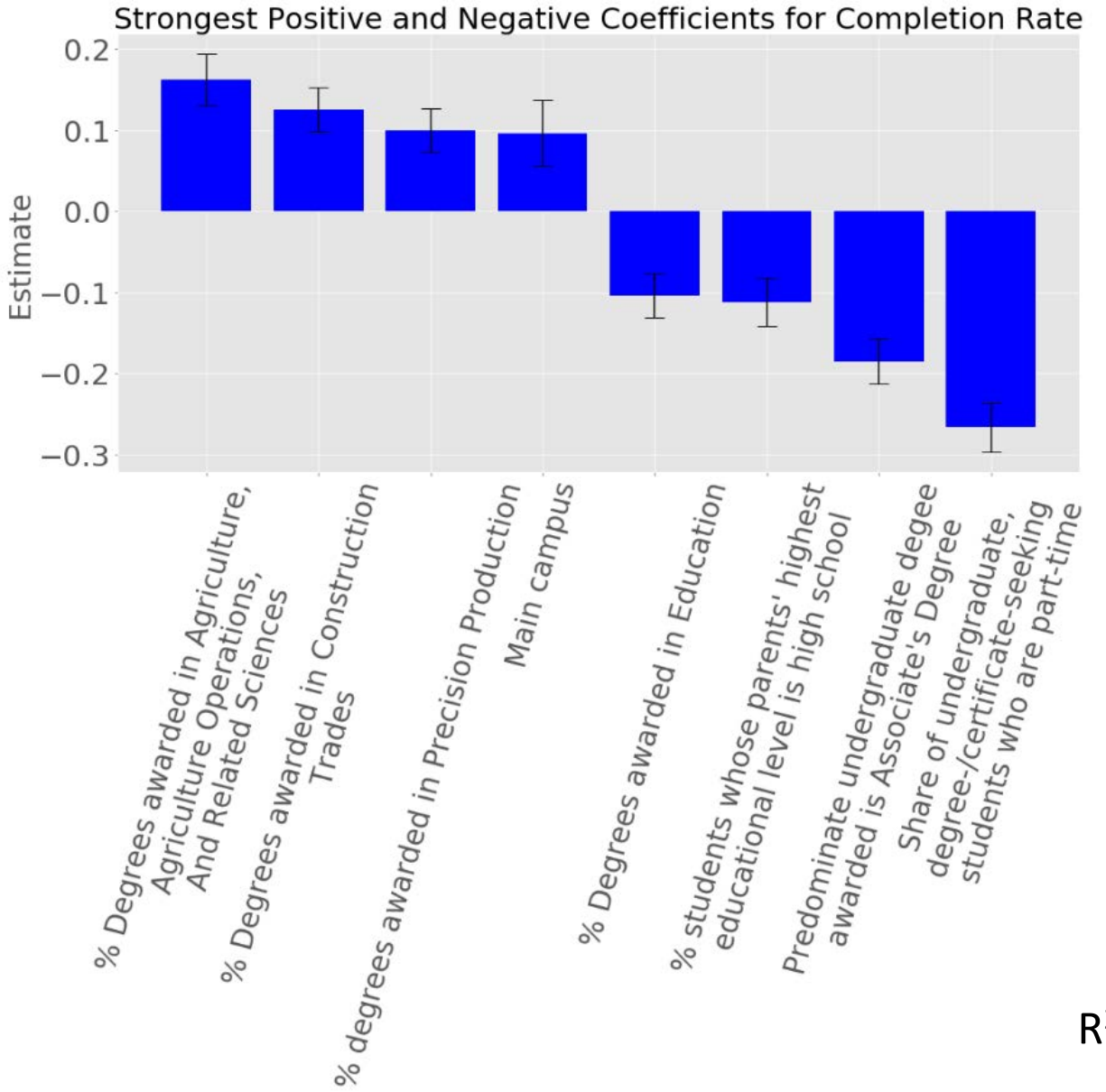
Methods – Data Cleaning

- Selected cases representing public, two-year colleges offering the Associate's Degree as the highest degree
- Cleaning included feature selection, case selection, addressing missing values, standardizing data, log-transforming data
- Resulted in DataFrame with 19803 cases and 52 features

Linear Regression & Machine Learning

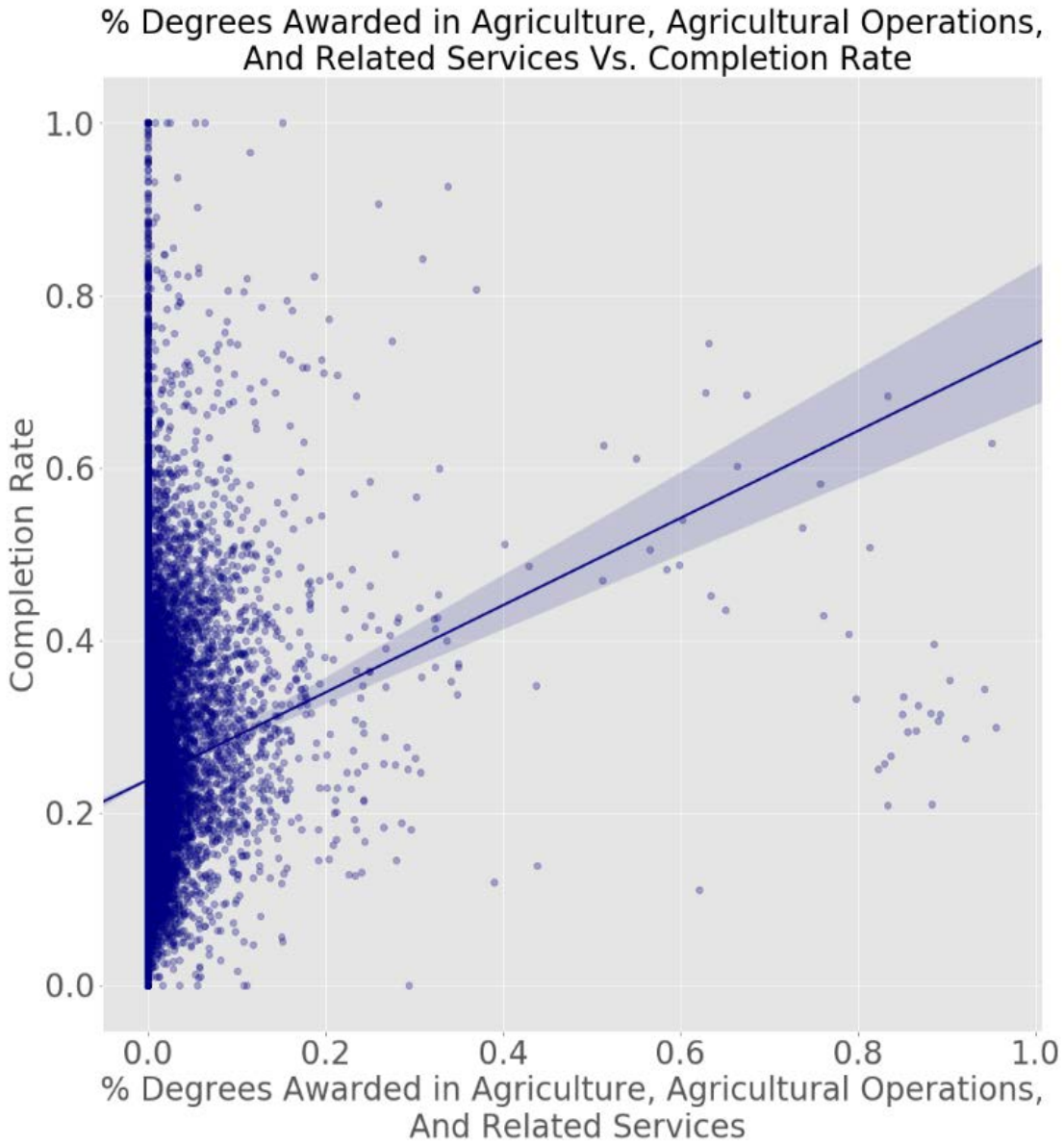
- Linear regression with completion rate as the target/response and remaining features as predictors
    - Used for statistical inference
- Compared seven regression-based machine learning algorithms in their abilities to predict completion rates of unseen data
    - 10-fold nested cross-validation & hyperparameter tuning

Results – Linear Regression

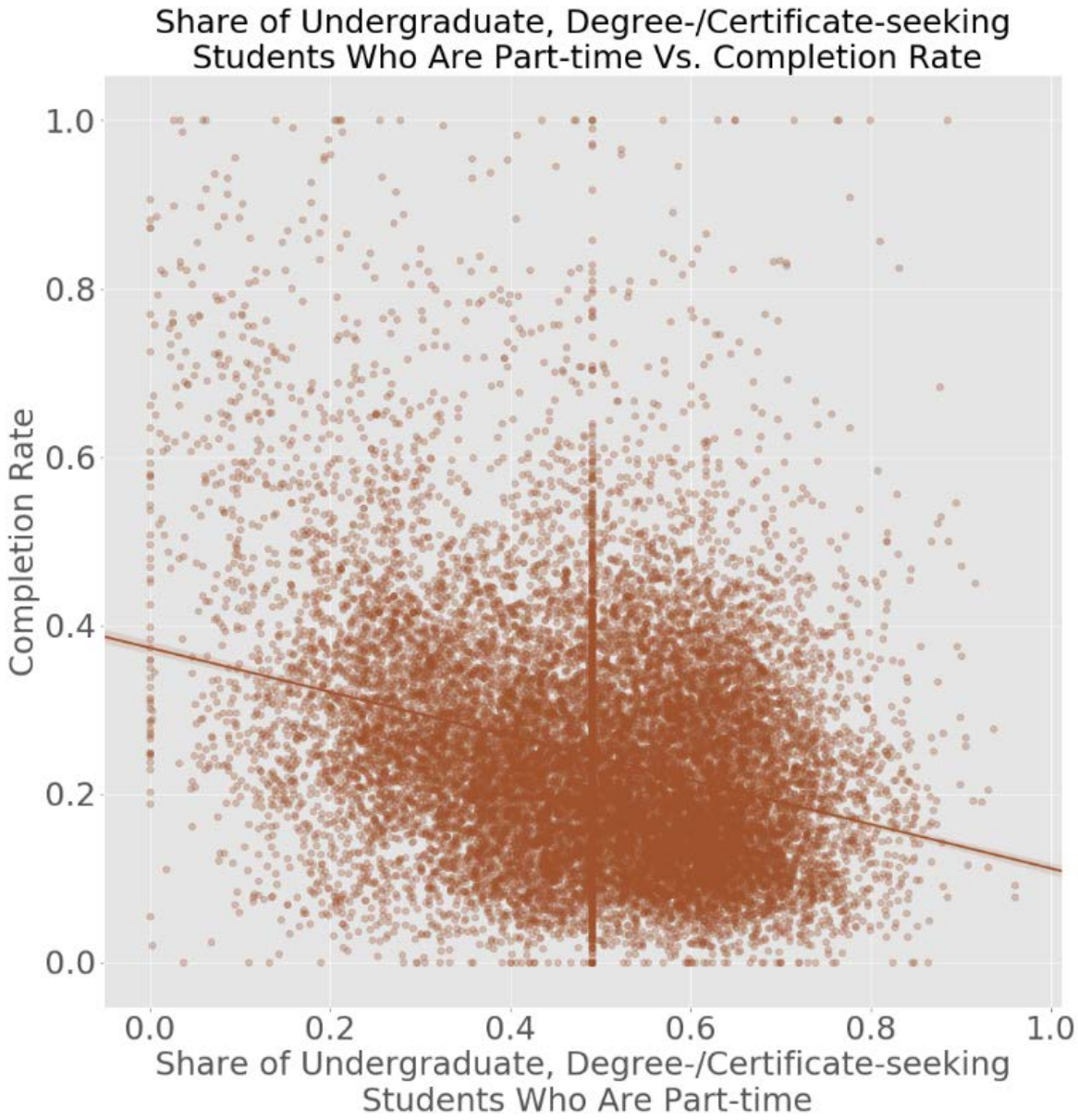


R^2_{adj} = 0.29, DF = 51, P <0.01

Strongest Positive Predictor



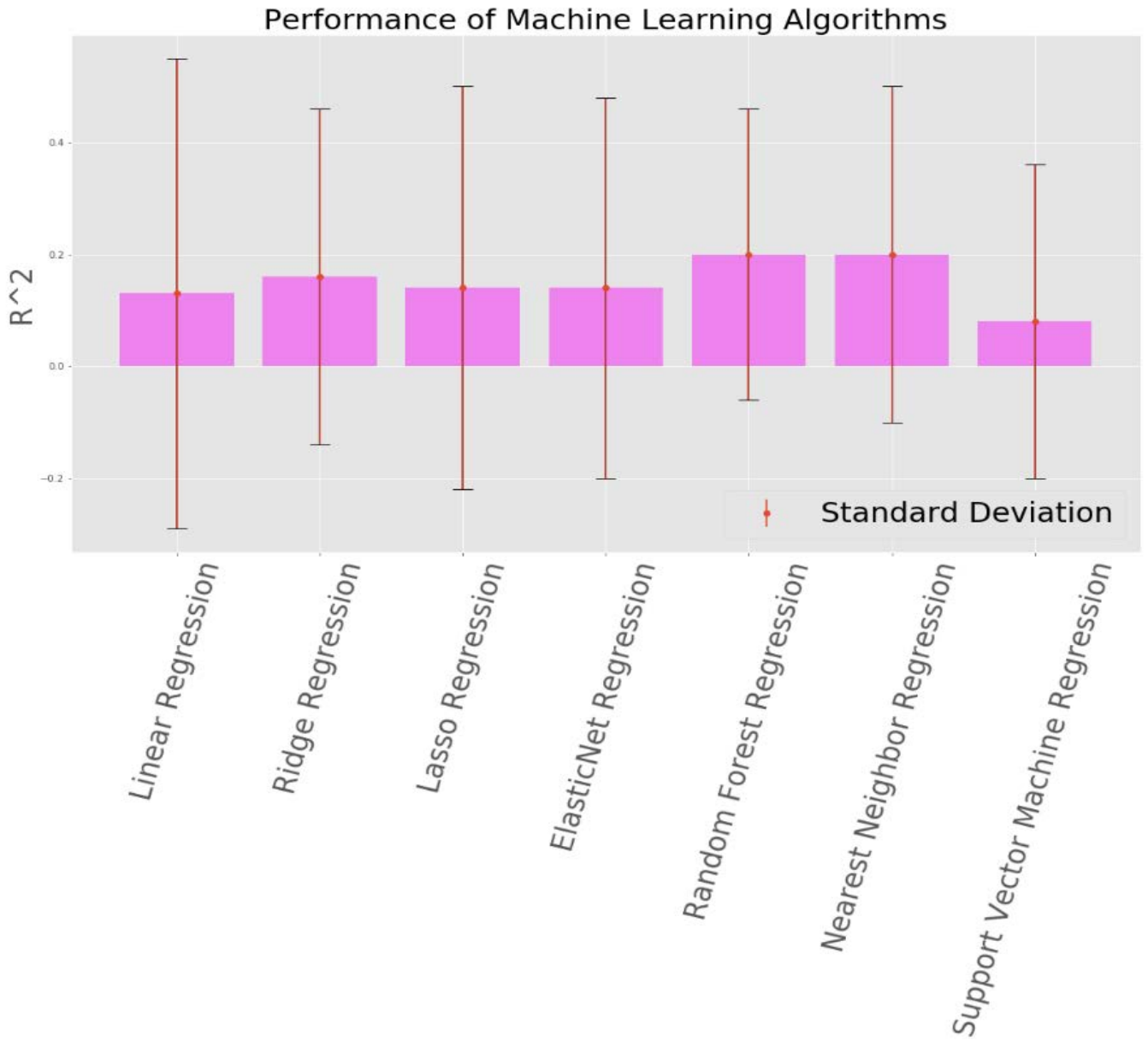
Strongest Negative Predictor



Machine Learning Performance

- Best: Random Forest Regression (R^2 = 0.20 ± 0.13 SD)

- Worst: Support Vector Machine Regression (R^2 = 0.08 ± 0.14 SD)



Conclusions

- Completion rate declines with % of part-time students seeking degrees or certificates
- Completion rate improves with % degrees awarded in technical and/or certificate-granting programs

Conclusions

- Linear regression and related machine-learning algorithms explain/predict less than 1/3 of variation in completion rate
- Advanced distribution-fitting could identify better-fitting models for the College Scorecard data