Predicting High School Graduation Rates for Illinois Schools

Presented by Dann Morr

Project Goal - Predict the graduation rate future cohort

What can variables regarding:

- Instructional Setting
- Previous Cohorts' Experience

tell us about the Graduating Rate of Seniors?

Data

1 | Gathering

Datasets for

Academic Years
2012-13 thru 2018-19

Illinois State Board of Education

ILLINOIS REPORT CARD

Data

1 | Gathering

Datasets for

Academic Years
2012-13 thru 2018-19

2 | Compiling

- ➔ Identify common features
- ➔ Filter data
- ➔ Combine into one matrix

Data

1 | Gathering

Datasets for

Academic Years
2012-13 thru 2018-19

2 | Compiling

- ➔ Identify common features
- ➔ Filter data
- ➔ Combine into one matrix

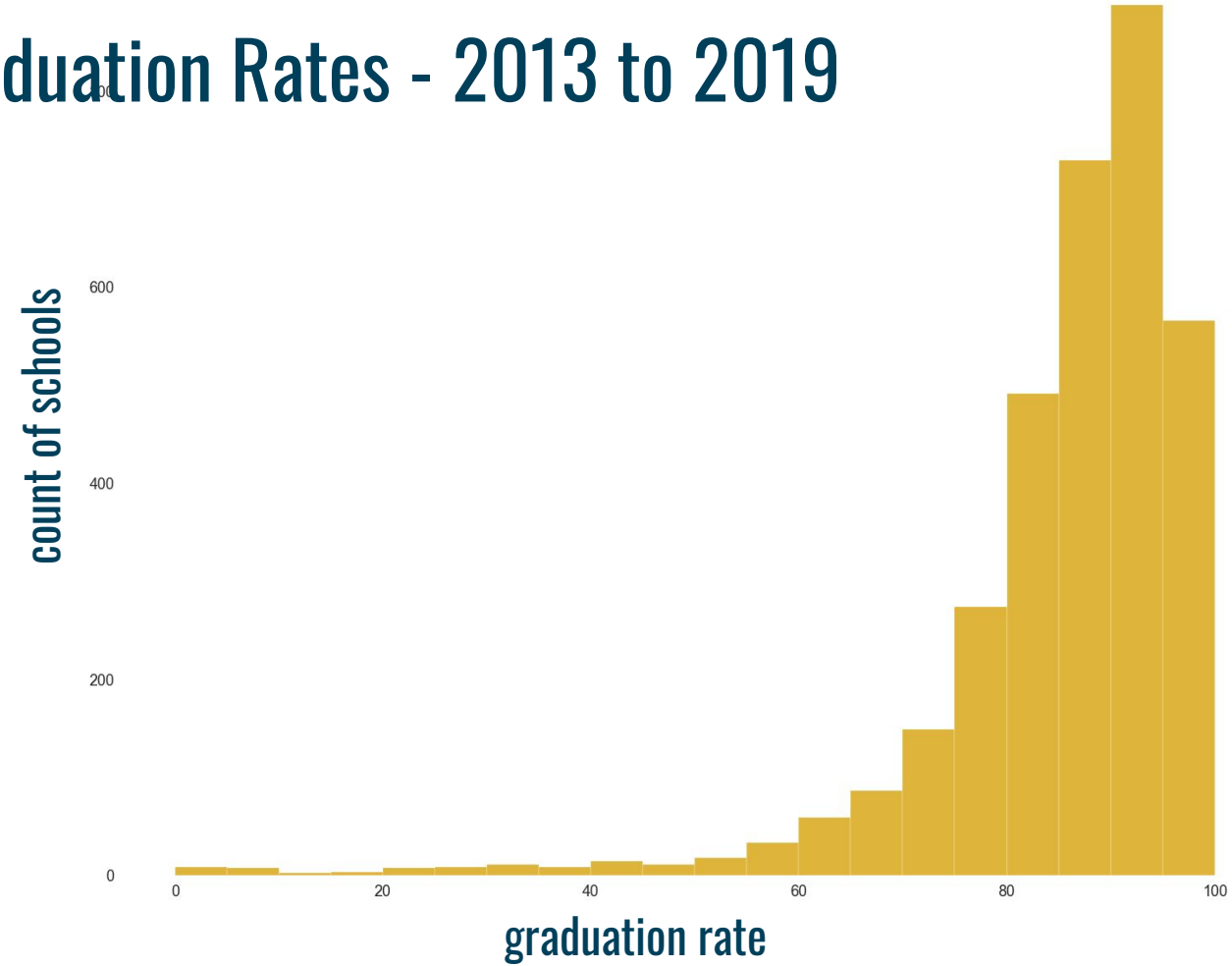
3 | Analyzing

- ➔ Identify target variable
- ➔ Find correlations
- ➔ Select models

Distribution of Graduation Rates - 2013 to 2019

Avg graduation rate:
85%

Majority of schools have graduation rate above
88%



Definitions

4-Year Graduation

First year: 2016
Second year: 2017
Third year: 2018
Fourth year: 2019

Graduation: 2019

5-Year Graduation

First year: 2015
Second year: 2016
Third year: 2017
Fourth year: 2018
Fifth year: 2019

Graduation: 2019

First Simple Model

Linear Regression

Result

Adjusted R-squared: 0.445

First Simple Model

Linear Regression

Result

Adjusted R-squared: 0.445

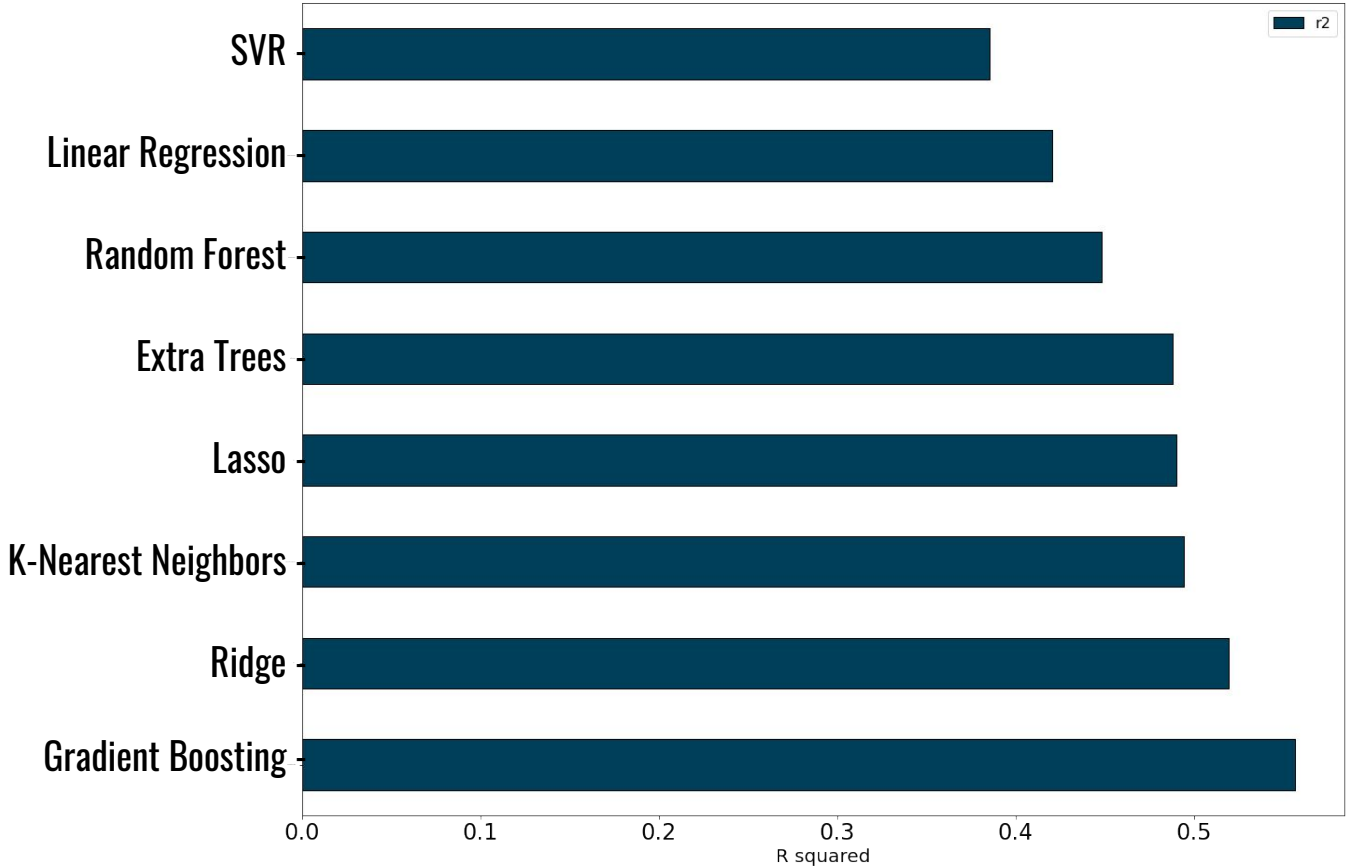
What does that mean?

Think of it as the percent of the answer that is explained by the model.

It's about 44.5% now.

I'm going to try to make it better

Comparing Model Performance - based on R squared value



Gradient Boosting returns a score of 0.557

Final Model

Gradient Boosting

Result

Adjusted R-squared: 0.684

Final Model

Gradient Boosting

Result

Adjusted R-squared: 0.684

It got better, didn't it?

It was at 44.5%

Now it's at 68.4%

How can we go higher?

Future improvement steps

- Hyper-tuning additional models
  - will different models perform better on test data?

- Feature selection and engineering
  - add features from district-level reporting
  - add school/municipality financial data

Contact

Dann Morr

email	dannmorr@gmail.com
github	@dannmorr
linkedin	linkedin.com/in/dannmorr