Predicting High School Graduation Rates with Machine Learning

Presented by Dann Morr

Introduction

- Improving graduation rates is a constant concern for school district superintendents and administrators
- From 2013 to 2019, Illinois' 4-year high school graduation rates have fluctuated between 83% - 87%.

*from Illinois Report Card Trend Data

Project Goal

- ➔ Predict 4-year high school graduation rates using machine learning
- ➔ Provide a tool for superintendents to use for managing their district schools

Data

1 | Gathering

- Public Datasets
- Academic Years 2013 to 2019

Illinois State Board of Education

ILLINOIS REPORT CARD

Data

2 | Compiling

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

Feature Categories

1. Student Demographics
2. Instructional Setting
3. Teacher & Administrator Statistics
4. College & Career Readiness
5. Advanced Coursework

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

Target

4-Year Graduation Rate

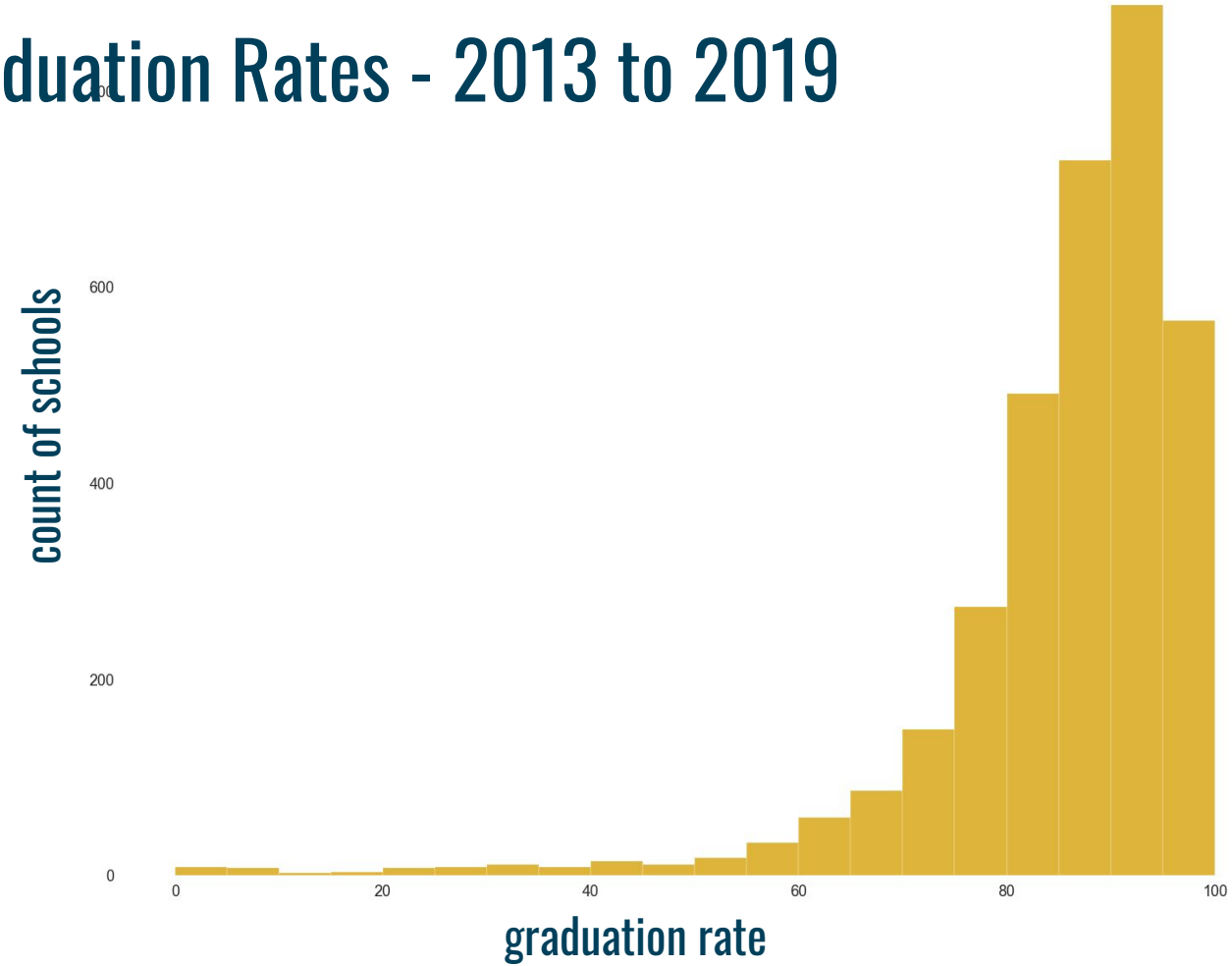
First year: 2016
Second year: 2017
Third year: 2018
Fourth year: 2019

Graduation: 2019

Distribution of Graduation Rates - 2013 to 2019

Avg graduation rate:
86%

Majority of schools have graduation rate above
88%



Machine Learning Models

First Simple Model

Linear Regression

Result

R-squared: 0.445

What does that mean?

First Simple Model

Linear Regression

Result

R-squared: 0.445

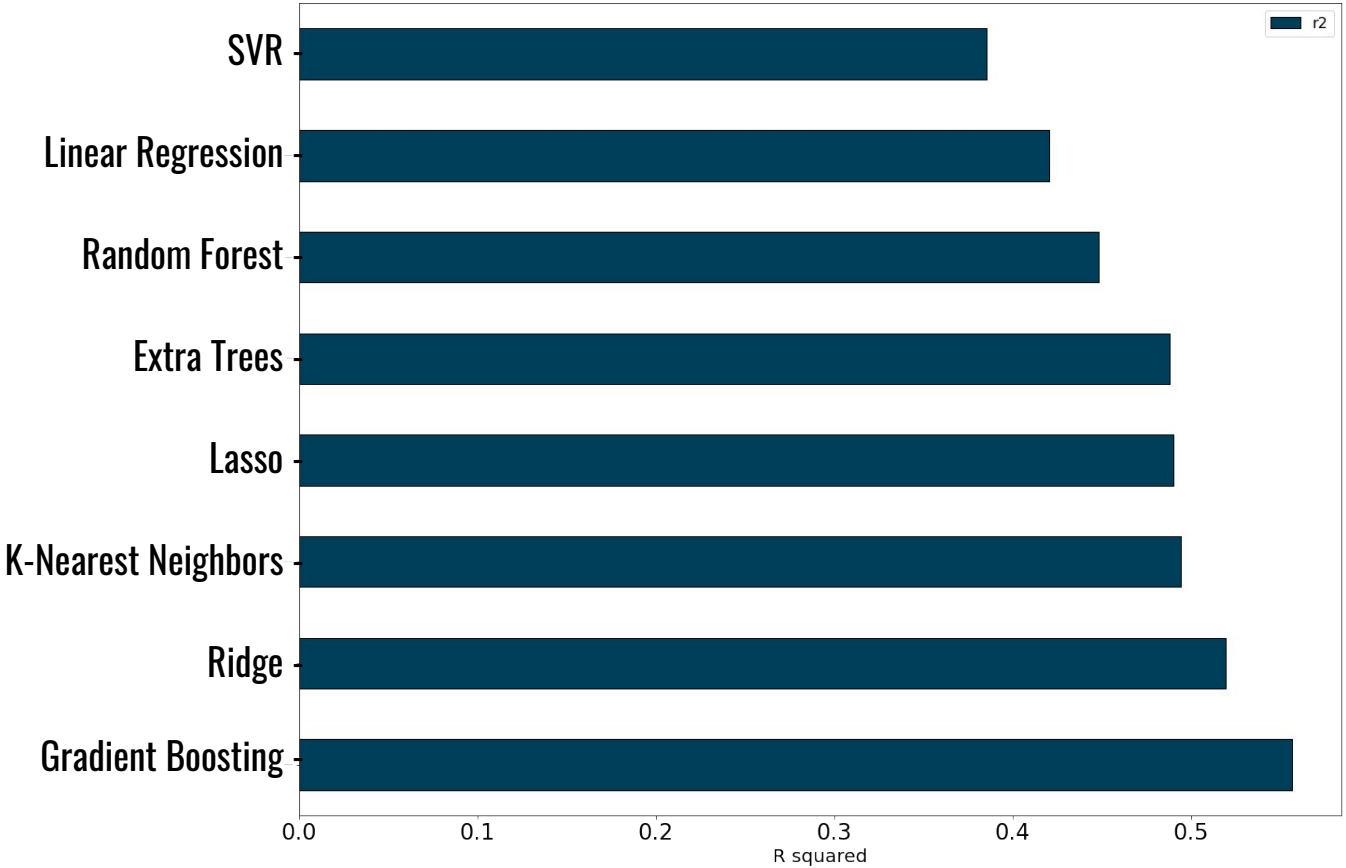
What does that mean?

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



Best performing

Gradient Boosting Regressor

R^2 value: 0.557

Final Model

Gradient Boosting

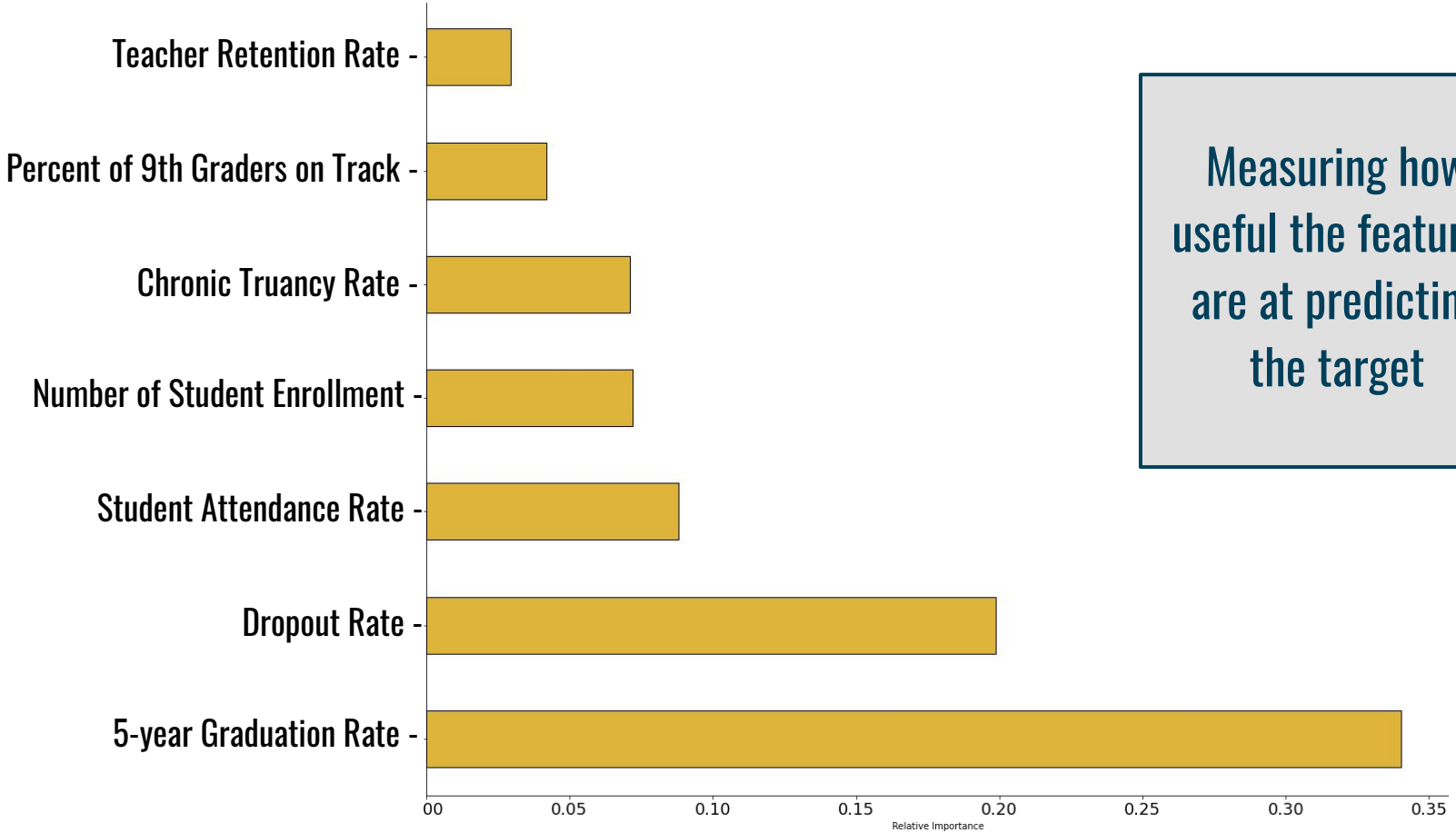
Result

R-squared: 0.684

The model is making better predictions

44.5%  68.4%

Features Identified as Most Important by the Final Model



Measuring how useful the features are at predicting the target

Conclusion

- ➔ Machine Learning Model shows proof of concept
- ➔ Possible applications:
  - monitor performance trends
  - identify cohorts in need of additional resources
  - project target goals for improved academic outcomes

Future improvement steps

- ➔ Evaluating additional models
  - ◆ will different models perform better on the 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