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 Preparation

1 | Gather

- Public Datasets
- Academic Years 2013 to 2019

Illinois State Board of Education

ILLINOIS REPORT CARD

Data Preparation

2 | Compile

- ➔ 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 Preparation

3 | Analyze

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

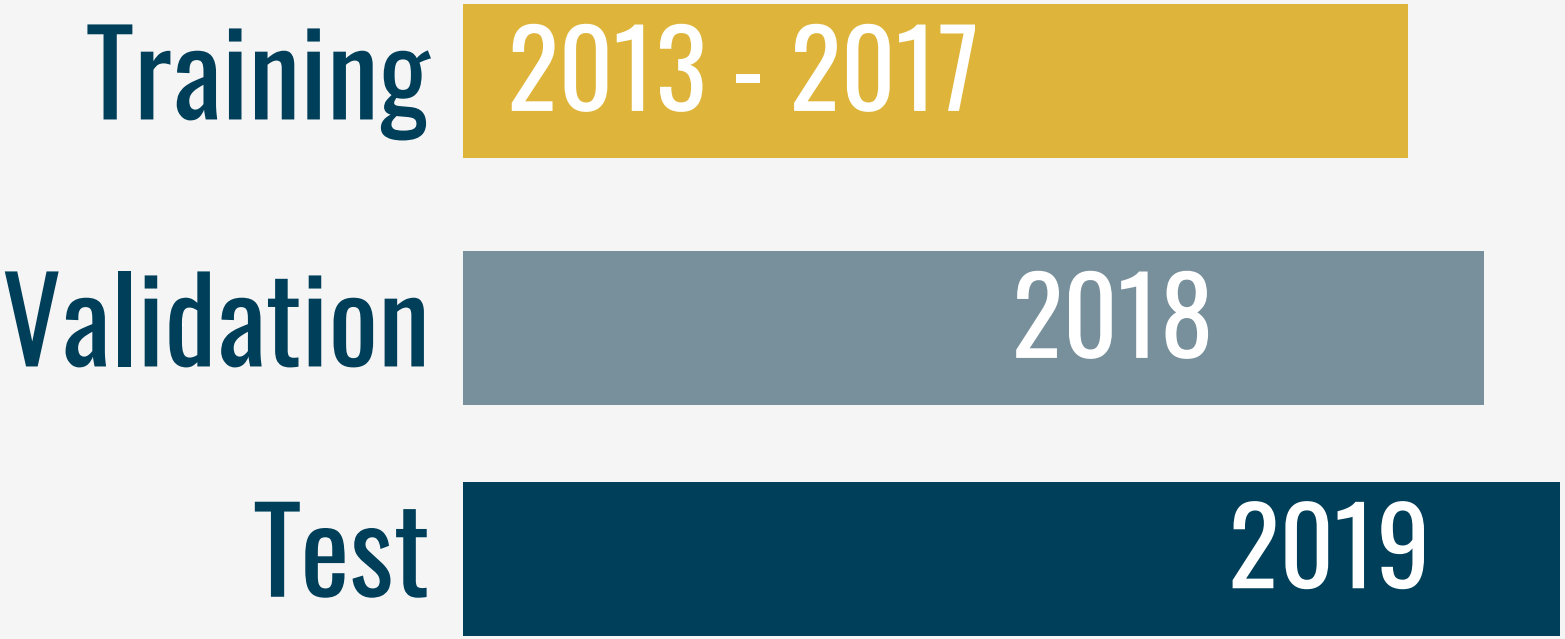
Target Variable

4-year High School Graduation Rate

Top correlated features include:

1. Dropout Rate
2. Student Attendance Rate
3. Teacher Retention Rate

Split Data into Training, Validation, and Test Sets



Machine Learning Models

First Simple Model

Linear Regression

Result

Adjusted R-squared: 0.441

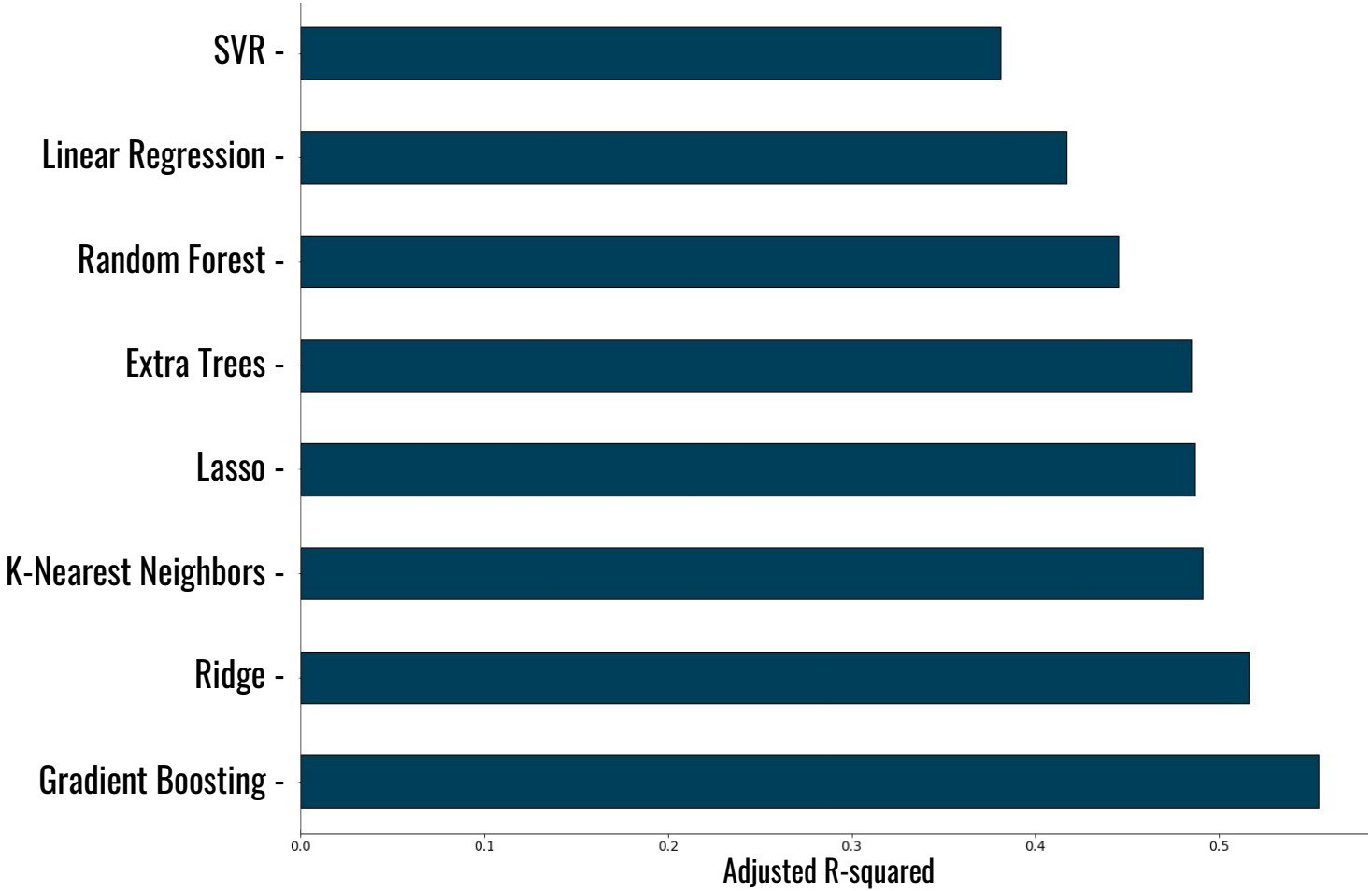
What does that mean?

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

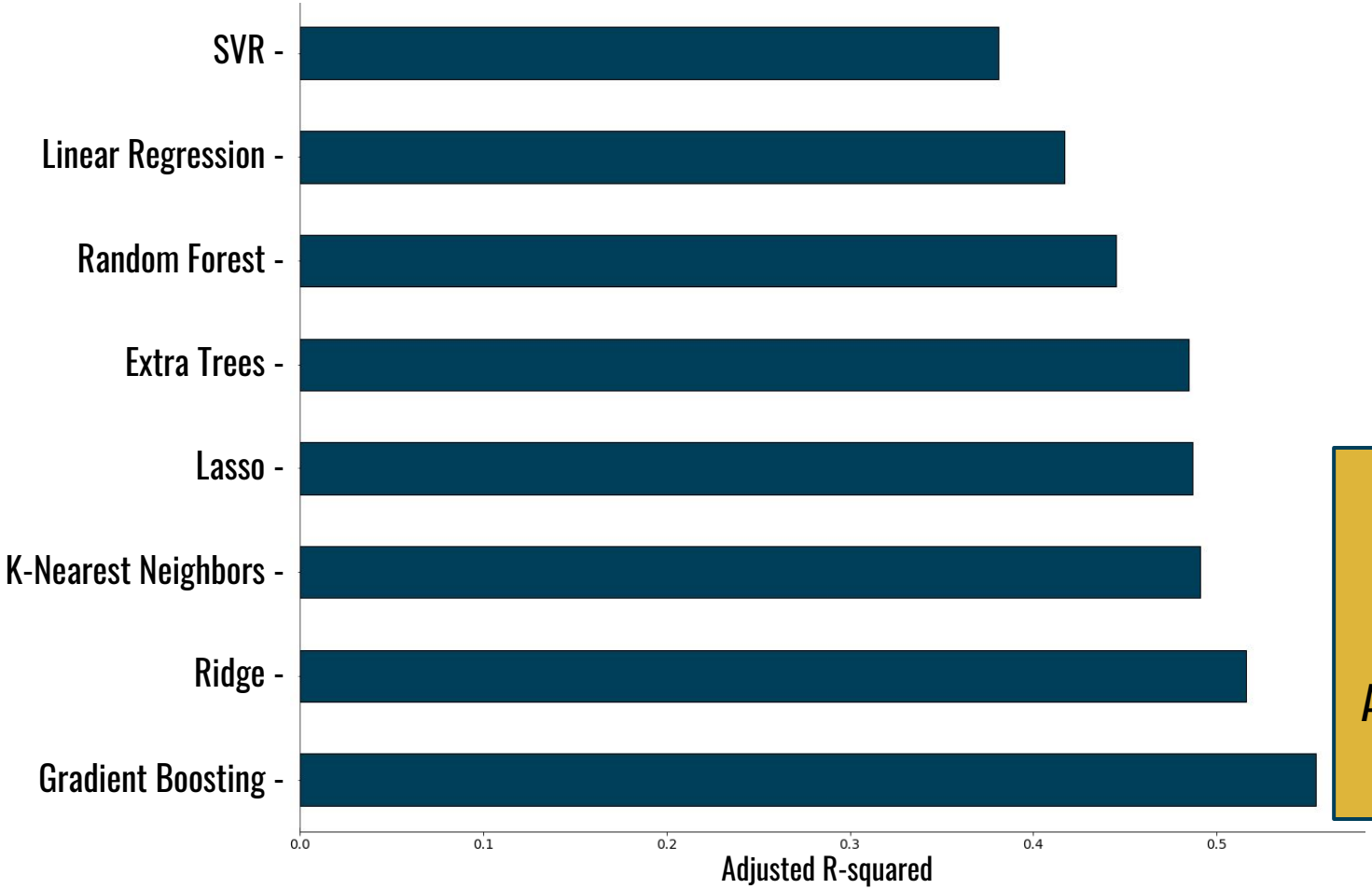
It's about 44.1% now.

I'm going to try to make it better at predicting.

Comparing Model Performance - based on Adjusted R-squared value



Comparing Model Performance - based on Adjusted R-squared value



Best performing

Gradient Boosting Regressor

Adjusted R^2 value: 0.554

Final Model

Gradient Boosting
Optimized by Grid Search

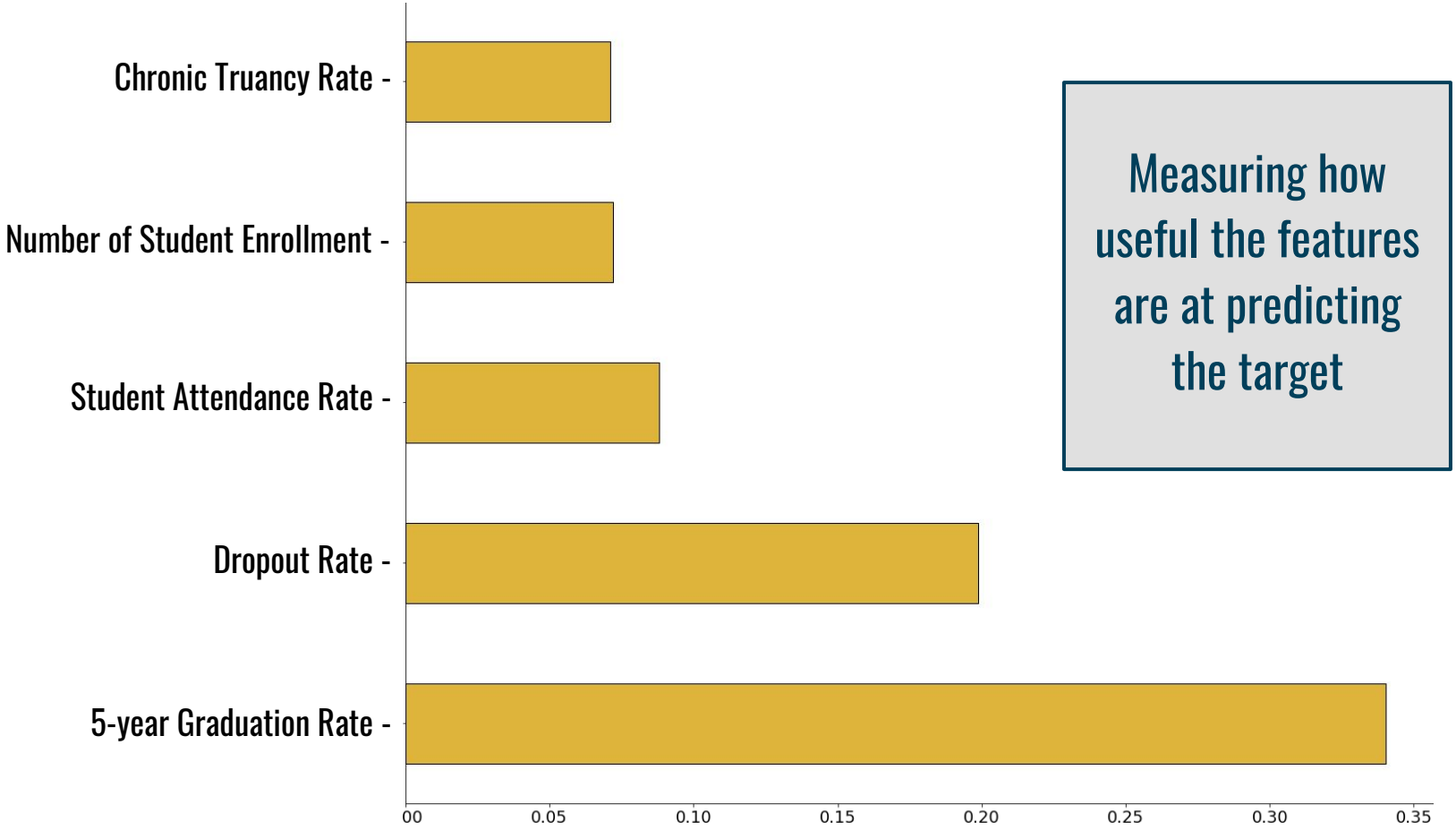
Result

Adjusted R-squared: 0.688

The model is making better predictions

44.1%  68.8%

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
  - ◆ remove ‘5-year Graduation Rate’ from model
  - ◆ 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