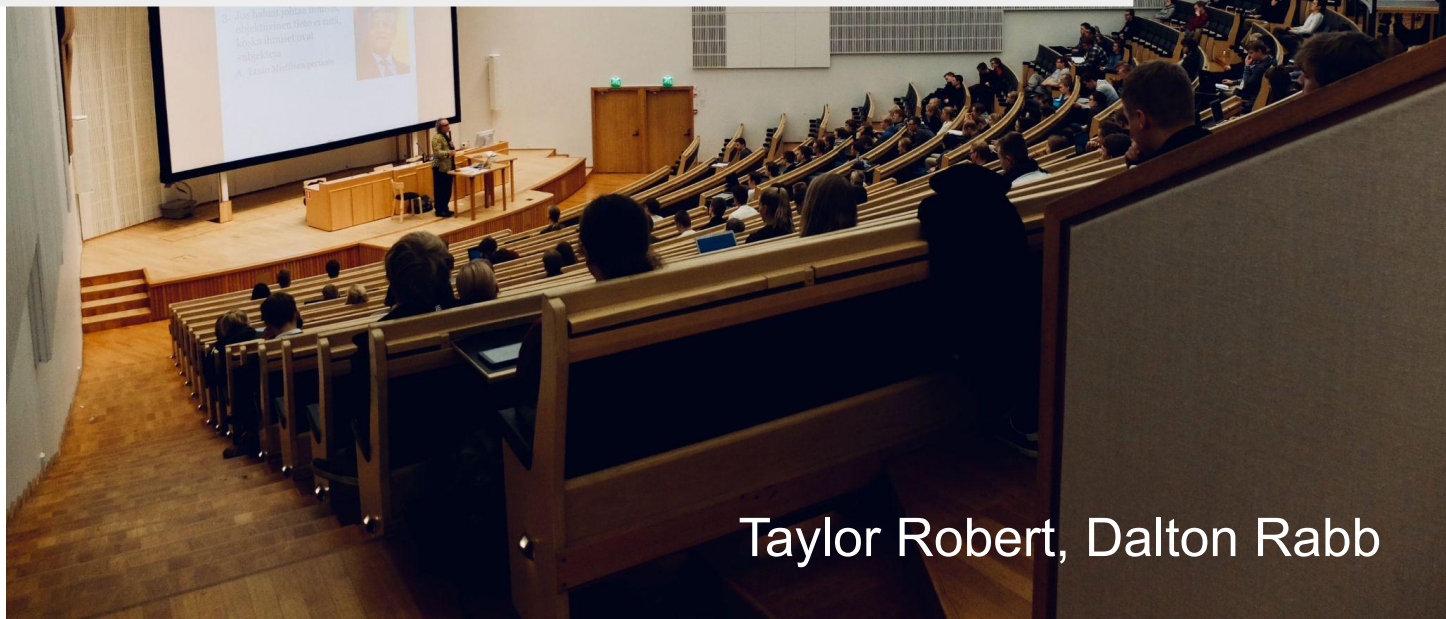


Predicting Student Performance



Taylor Robert, Dalton Rabb

Purpose

Predict future student and school performance using environmental data.

Understand the predictive advantage gained by using pretest data.



Data Source



Publicly available data on Kaggle, with 2133 available records.

This dataset features information regarding school setting, teaching method and classroom size.

Most features of this data set were categorical, and were modeled using one-hot encoding.

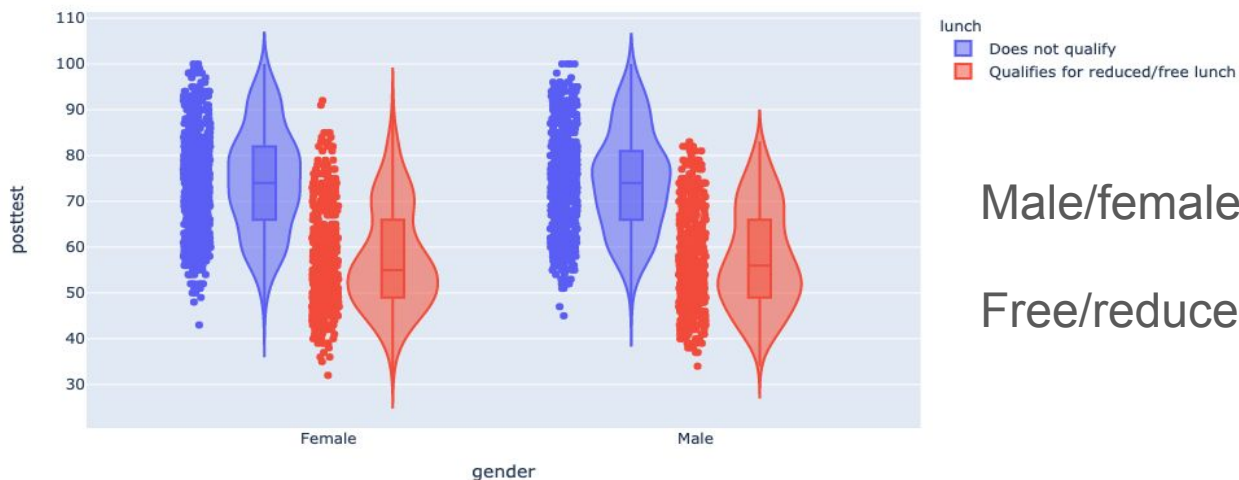


Data Features

Column Header	Description	Notes
school	Name of the school the student is enrolled in.	categorical
school_setting	The location of the school	categorical
school_type	The type of school. Either public or non-public	categorical
classroom	The type of classroom	categorical
teaching_method	Teaching methods: Either experimental or Standard	categorical
n_student	Number of students in the class	continuous
student_id	A unique ID for each student	not relevant
gender	The gender of the students: male or female	categorical
lunch	Whether a student qualifies for free/subsidized lunch or not	categorical
pretest	The pretest score of the students out of 100	Base score
posttest	The posttest score of the students out of 100	Target variable to predict

Exploratory Analysis - Observations

Posttest Scores and Economic Status

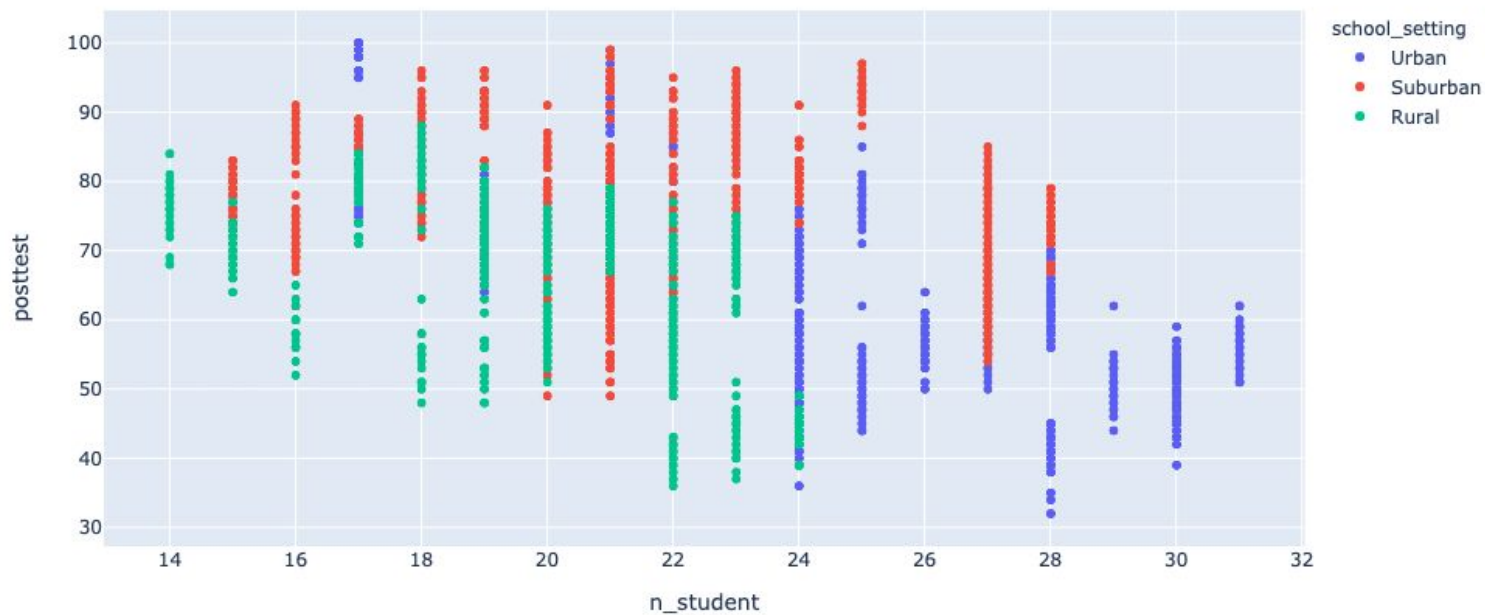


Male/female no noticeable effect

Free/reduced lunch noticeable effect

Exploratory Analysis - Observations

Class Size, Test Scores and School Setting



Exploratory Analysis

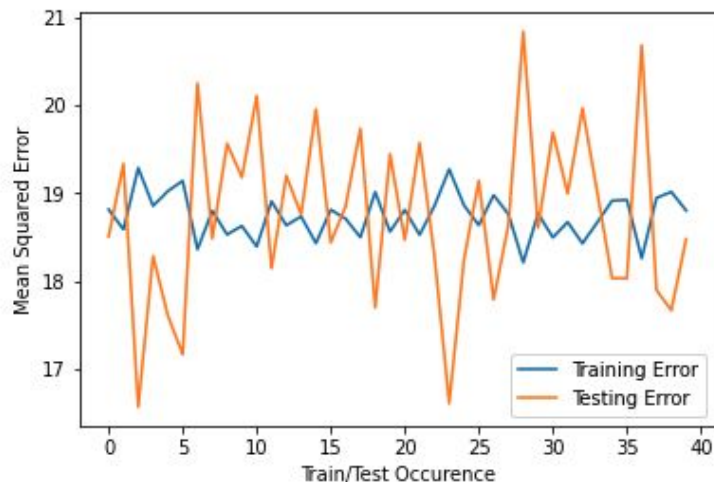
First, we explored a baseline regression model to understand how well the school system currently predicts scores using only pretest data.

We will use these results as a metric against which we can measure the success of modeling.

Baseline Model Performance

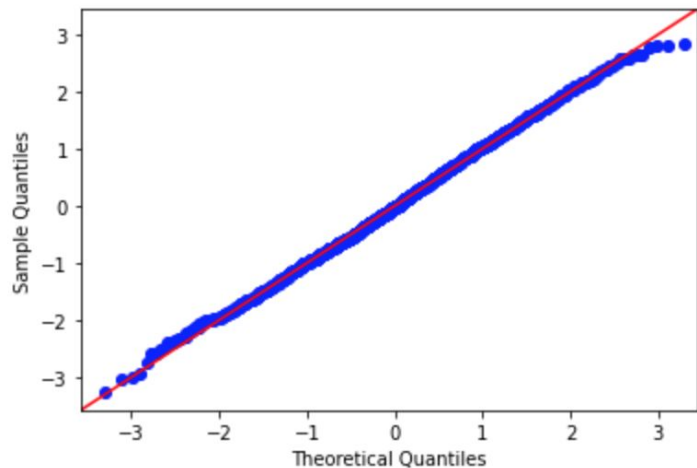
RMSE: 4.3285

R^2 : 0.9042



Environmental Factor Model

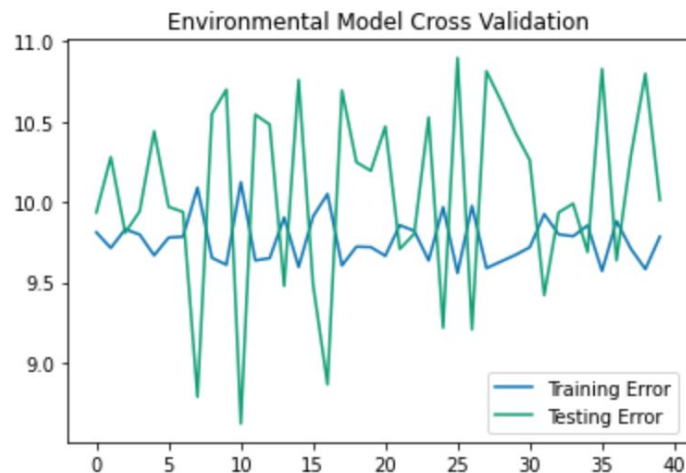
- This model used features of the dataset other than pretest data.



Env Model Performance

RMSE: 5.714

R^2 : 0.83



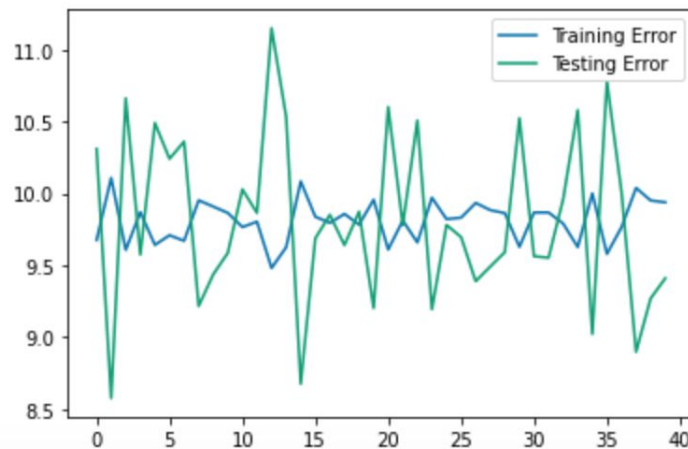
Pretest & Environmental Factor Model

- Began by using all available environmental predictors within the dataset
- After initial model creation chose to selectively drop predictors if the P value indicated statistical insignificance

Pretest + Env Model Performance

RMSE: 3.1310

R^2 : 0.95



Model Performance Comparison

Model Description	RMSE	R ²
Pretest Predictor	4.33	0.90
Environmental Predictors	5.71	0.83
Pretest & Environmental	3.31	0.95

Summary

- Pretesting provides powerful information to predict student performance
- Using pretest scores with environmental predictors can further improve the predictive power of a model
- 83% of the variance in student posttest performance can be explained by modeling environmental factors alone



Recommendations

- Model the effects of specific resource allocation:
 - Ex. predict effect of lowering class sizes or retraining teachers in different teaching styles.
 - Collect data by student over the life of the student - pretest scoring might be statistically insignificant in light of last year's posttest scores.
 - Once a continuously successful predictive model is made, allocate funds to students in negatively impactful environments.
 - Have over-performing schools share what teaching methods work with their underperforming counterparts.

THANK YOU

For your time and attention.

Project Repository: [GitHub repo link](#)

Authors presenting today can be contacted using the following information:

Taylor Hale Robert taylorhalerobert@gmail.com
Dalton Rabb drabb138@gmail.com

