Cleburne ISD

Utilizing Machine Learning to Distribute Math Resources

Jonathan Armstrong

Problem Statement

How can Cleburne ISD distribute math resources over the next academic year to promote optimum student growth and achievement?

Context: The assistant superintendent of data and campus improvement has provided information about all cleburne isd fifth grade students from the 20-21 academic year. This includes demographic, socioeconomic, and limited academic data as well as how each student performed on the Math STAAR assessment from failure to masters. The goal provided was to use this data to decide early on in the school year where to provide resources such as classroom utilization of a math specialist, extra time with paraprofessionals, educator training, and so on.

Scope of solution space: This project focuses only on where to distribute the resources for 1 academic year. It does not involve which resources are decided on or provide any long-term guidance.

Constraints within solution space:

-The data provided was < 500 entries.

-The testing data set aside and used to generate simulations of efficacy was < 100

-This project makes the following assumptions:

1) The resources in question are separate from RTI (response to intervention) and therefore are distributed on a classroom by classroom basis and not through pullout methods.

2) This project makes a blunt approximation of the current standard for resource delivery assuming that the classrooms with the top 20% highest number of sped/economically challenged students receive the resource. In this case economically challenged and sped were each counted as 1 point separately so that a student who was both economically challenged and special ed would contribute 2 points to their class total.
3) This project assumes that all resources have an equal impact on all students and classrooms. Specifically, the given resource would increase test performance on 20% of students in a classroom regardless of the particular features of any student.

Exploratory Data Analysis

Exploring the data was difficult at a student level due to it's granularity. However, by aggregating the data by campus it was possible to glean some overarching trends. First the obvious patterns were examined to make sure district math data matched state trends. Namely socio-economic and special education status played a large part in how each campus performed. Language did not have a significant impact on campus performance both when aggregated by campus and when examined on an individual student level.

One of the most surprising elements to surface during exploratory analysis was the significance each individual campus had on student performance in spite of overarching demographic trends. As is explored in the linked Tableau dashboard, campus number emerged as one of the most important predictors of student achievement. This spoke to the importance of fostering individual campus cultures across a district.

Tableau Dashboard

Machine Learning to Predict category

The data used for this project can be accessed here. This data has been scrubbed of identifying information.

Wrangling/Cleaning

Before training the model the data was assessed and cleaned. One of the most important features was staar progress 2019 so the 3% of students who were missing that feature were dropped. There were no students who were “new to Texas” in this set so this feature was dropped altogether. Discipline placement issues were reduced to three bins instead of the individual number of discipline incidents to reduce impact of outliers. All ordinal data was transformed into numerical form.

Access the jupyter notebook here.

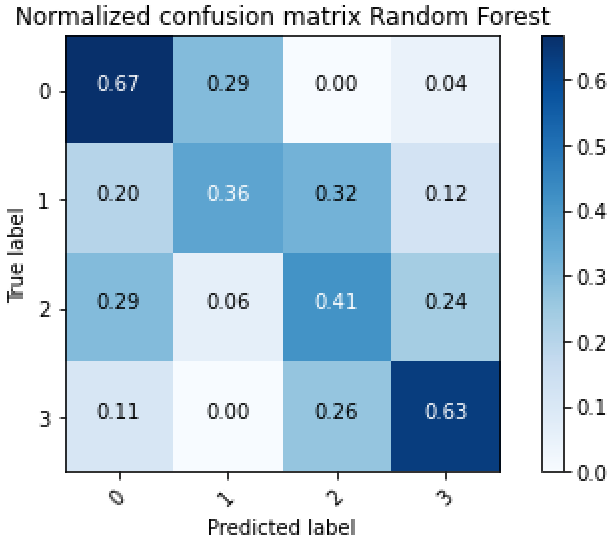
Pre-Preparation/MachineLearning Training

All numerical features were scaled using scikit’s standard scaler. Dummies were generated for all categorical features using pandas get_dummies function. The data was split with 80% used for training and the remaining 20% used for testing purposes.

Several models were tuned using a randomgridsearch. The top performing models were LogisticRegression, RandomForest, and KNearestNeighbor. Compare their performance here:

Name	Accuracy	f1-score
LogisticRegression	0.494	0.488
RandomForest	0.518	0.513
KNN	0.506	0.507

RandomForest was eventually chosen as the model to be utilized as it had the highest performance for choosing level 1 or “approaching” students as visualized in this normalized confusion matrix.



The most important features according to the RandomForest model were:

-STAAR progress from 2019

-Gifted/Talented status

-At Risk category

-Campus number

-Special Education Status

This coincides with the predictions made during exploratory analysis.

The associated jupyter notebooks: prep, training/selection

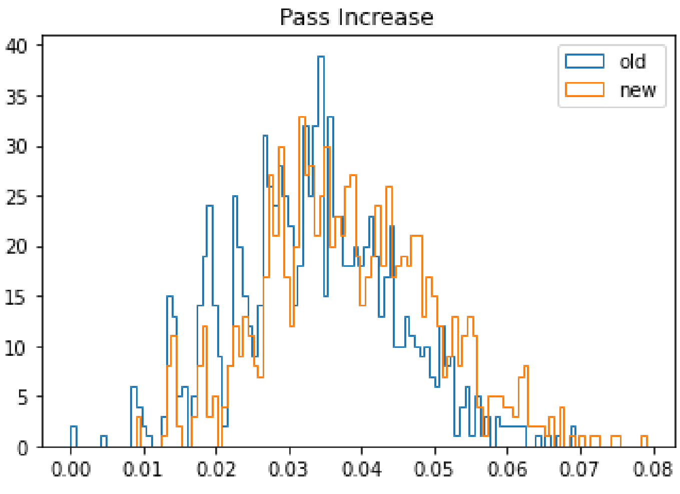
Simulating Model Efficacy vs Standard Distribution

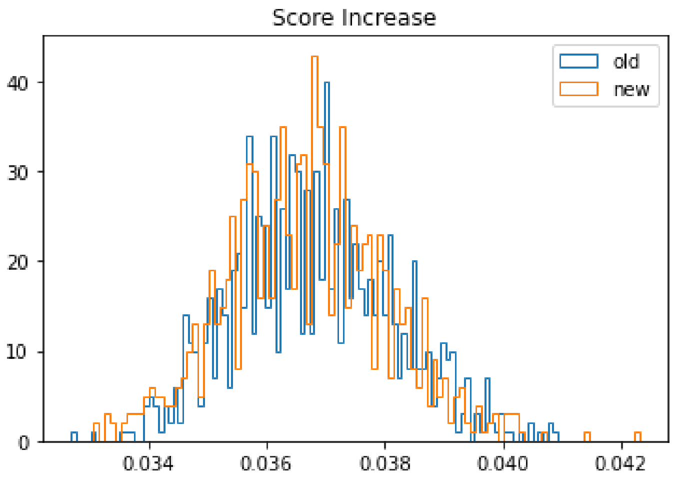
The prediction before any simulations were created was that both standard distribution of math resources and distribution based on the machine learning model would produce similar increases in the average growth of the target feature (STAAR placement). Since in both cases the same number of resources would be distributed and have the same impact.

However, the standard distribution would focus on the classrooms which would likely have the lowest performance based on economic and sped status, whereas the machine learning distribution would focus instead on classrooms with students predicted to be “approaching” the STAAR. Raising the level of students who’ve failed would only get them to “approach”, and would have less of an impact on the pass rate for the district than raising the level of students who already approach and with help could achieve “meets”.

1000 school districts were simulated by randomly generating groups of 25 fifth-grade classrooms. These simulated classrooms were themselves filled with groups of 20 students randomly selected from the testing set of the data. A resource was distributed to the top 20% of classrooms within a district that had the highest number of economically challenged and sped students. This was to approximate how resources would normally be distributed within a district to the classrooms with the highest need. Using this method the average pass rate for all the simulated districts rose by 3% while the average staar score earned by students rose by 4%.

Using the same method 1000 new districts were generated to test distribution through the use of the RandomForest model. Through this method resources were distributed based on the top 20% of classrooms within a district for their respective sum of predicted “approach” students. As expected the average staar scores for these simulated districts rose by 4% as well. However the pass rate for districts using this method also increased by 4%, a 33% improvement over the standard method. A t-test was performed on the distributions of increases between both methods which produced a p value of less than 0.05 which indicated statistical significance. This means that it was likely due to the new method actually producing higher results compared to the old method, instead of due to chance because of sampling. The histograms below illustrate the differences between the distributions.





The jupyter notebook used to create these simulations can be found here.

Application and Future Work

While the results of utilizing the machine learning algorithm in this case may not be with the logistical cost of rolling it out among the campuses, I believe the results here do show a proof of concept. Namely, by focusing on the students who are probably close to passing, the pass rate will improve at a higher rate than by focusing only on the students who are probably going to perform poorly.

The largest hindrance to the success of applying this machine learning model was it's lack of accuracy in prediction the desired category. This may be a result of the features used to train it and the size of the test sample. With historical student data that has been adjusted for seasonality, along with a new feature which would track a students previous year's staar achievement, a much more accurate model might be possible which would make targeting "approaching" students more efficient.