ECON 573: Lochana Arumugam, Manu Srinivas, Nikolas Lewis, Ngudiankama R. Mfulama

Impact of School Resources and Demographics on District Test Score Performance

A Regional Analysis Using Classification Methods

This project investigates the question: How do variations in school resources and district-level demographics affect a school district's likelihood of performing above the national average in standardized test scores? In particular, we focus on whether districts perform above or below the national average, which we operationalize as standard deviations (SD) from the U.S. mean test score. Next, we analyze how key inputs such as funding gaps, student enrollment, and student demographic composition play in this outcome. We hypothesize that both financial adequacy, measured as whether a district meets or exceeds predicted per-pupil spending targets, and key demographic factors like poverty rates and racial/ethnic composition would significantly influence test score outcomes.

This research would have both theoretical and policy implications. Policymakers often debate whether increased funding or resource reallocation can reduce achievement gaps, and previous studies have typically examined continuous performance measures or aggregated national data. However, our work uses a binary classification - districts performing "Above" (1) versus "Below"(0) the national average - to help provide a clear decision tool for policymakers. Although prior research has documented links between student performance and socioeconomic background, there is less consensus on whether districts are using their resources efficiently and which specific factors predict outperforming the national benchmark (Dey 2021). We believe that understanding these relationships is crucial for informing funding formulas, targeted interventions, and policy reforms. Moreover, our emphasis on a regional approach (comparing South, Northeast, Midwest, and West districts (according to the Census split) contributes to the literature by aiming to find relationships across cultural bounds, localized funding policies, and demographic structures.

Our primary contribution is threefold. First, we use a new identification strategy by splitting our analysis across the regions of the United States, where the interplay of resources and demographics might differ markedly from national averages. We utilize the District Cost Database (DCD) from the School Finance Indicators Database (SFID), which includes a unique measure of adequate (or required) spending, that is, the amount of per-pupil spending a district would need to meet national average math and reading scores. Second, we transform the DCD's outcome gap variable into a binary outcome, which allows us to frame the analysis as a classification problem rather than a continuous outcome model. This aids in our interpretation since we can compare and contrast each predictor and see if it contributes to or diminishes the outcome gap. Finally, we implement and compare a suite of classification models, including logistic regression (with stepwise selection), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), classification trees, random forests, boosting, support vector machines (SVM), and k-nearest neighbors (kNN), to predict performance relative to the national benchmark. We further use cross-validation techniques to select optimal probability thresholds and ensure that our models do not overfit, to assess how well various inputs predict whether a district will outperform or fall short of the national benchmark.

Our main results show that spending adequacy and student poverty rate are among the strongest predictors of positive academic performance. We also find that some districts outperform expectations despite below-average funding, suggesting that non-financial factors may also play a meaningful role. These findings extend the literature by providing a predictive framework for identifying high-performing districts and emphasizing the importance of both funding equity and contextual district characteristics in shaping student achievement. While some model performances varied by method, the overall results reinforce the potential for data-driven decision-making in education policy.

Statistically, by the end of our analysis, our results suggest that even marginal improvements in resource allocation (as measured by the funding gap) and adjustments in demographic-related educational policies can have substantial impacts on district performance. The consistent findings across multiple models underscore that targeted fiscal and socio-economic interventions may shift a district's likelihood of performing above or below the national standard. Overall, the evidence supports the view that effective resource allocation and tailored policy responses are critical to enhancing educational outcomes, with ensemble methods like random forests providing particularly strong predictive insights concerning our regions.

Literature Review:

A growing body of research has explored how disparities in school resources and demographic characteristics influence student academic outcomes, particularly standardized test scores. Eric Dey (2021) examined international PISA(Program for International Student Assessment) data and found a strong, positive relationship between the quality of educational resources, such as instructional materials, infrastructure, and technology, and student math achievement. The study emphasizes that better-resourced schools tend to yield stronger academic performance.

McMahon (2000) conducted a district-level analysis in South Florida and found that poverty was the most significant predictor of student underachievement. While certain targeted spending, like reducing class sizes, showed some promise, the overall impact of poverty outweighed the effects of resource-based interventions. Similarly, the Commonwealth Institute (2023) documented systemic challenges faced by high-poverty schools, including fewer resources, higher student-to-staff ratios, and less experienced teachers. These schools also experienced lower standardized test scores, higher dropout rates, and diminished accreditation levels.

At the national level, Reardon et al. (2022) used longitudinal data from 2009 to 2019 to identify trends in academic achievement across U.S. school districts. They found that districts with more experienced teachers and lower levels of teacher turnover showed significantly greater gains in student test scores. These findings highlight the importance of both instructional quality and school stability in closing achievement gaps.

More recently, Gautam et al. (2024) applied machine learning methods to examine educational equity in Georgia's K–12 system. Their model identified socioeconomic status, particularly free/reduced lunch eligibility, as the most significant predictor of test score performance, followed by per-pupil faculty spending. These findings reinforce the consistent theme that both demographic and financial resource variables are deeply tied to educational outcomes.

The Model:

As we aim to model the probability that a school district performs above the national average, we build on the basic assumptions that a school district's performance is influenced by two main factors. The first is demographic composition. This includes metrics such as total enrollment, poverty rate, and the percentage of special education students etc. The second was the resource gap. This is the difference between actual spending per pupil and the required/adequate spending. While we do not develop a clear-cut formal theoretical model, our empirical approach rests on several key behavioral, informational, and institutional assumptions about how education systems operate and how resources influence student achievement.

We assume that state-level academic performance (as measured by the National Assessment of Educational Progress) is influenced by the interaction between (1) resource inputs, including per-pupil spending, special education services, and access to qualified teachers, and (2) contextual demographic factors, such as student poverty rate, racial/ethnic composition, and the share of English language learners. Consistent with prior literature, we also assume there are diminishing marginal returns to spending, particularly in higher-income districts, along with the fact that schools aim to use up their funding so as not to have a budget decrease either.

A central institutional feature of our model is the grouping of states into districts and then benchmarking the district performance to the national average. Rather than modeling student achievement as a continuous outcome, we use the outcome gap variable from the District Cost Database (DCD), which measures a district's average test score in standard deviations from the national mean. We transform this into a binary outcome variable, “class_outcome”, coded as 1 if the district performs above the national average and 0 otherwise. This allows us to model the outcome as a discrete classification problem.

The primary specification is a logistic regression

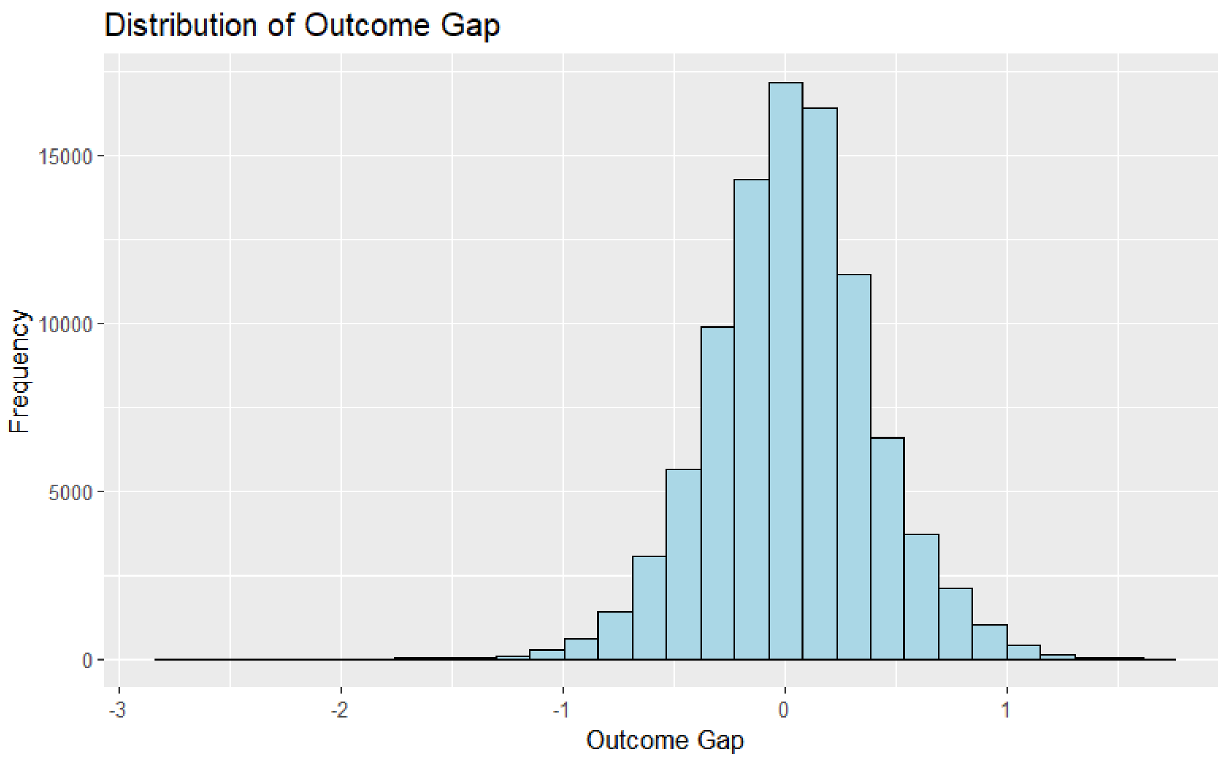
$$\log\left(\frac{\Pr(\text{Above}_i)}{1-\Pr(\text{Above}_i)}\right) = \beta_0 + \beta_1 \text{FundingGap}_i + \beta_2 \text{Enroll}_i + \beta_3 \text{Pov}_i + \beta_4 \text{IEP}_i + \beta_5 \text{ELL}_i + \sum_{k=6}^{K} \beta_k X_{ik} + \varepsilon_i,$$

We supplement this logistic model with additional classification algorithms, including classification trees and random forests. This helps us test the importance of our predictors and find whether more flexible, non-parametric methods yield better predictive accuracy or highlight nonlinearities missed by the logistic regression.

Assumptions and Sensitivities

- We assume that district-level spending decisions and demographic characteristics are exogenous to the test score performance observed in the current year. In practice, this may be violated if districts increase spending in response to past low performance.
- We assume cross-district comparability of test scores, made possible by the standardized outcome gap variable in the DCD. However, any residual unobserved heterogeneity in state-level testing standards may bias the estimates.
- The binary transformation of the outcome gap into class_outcome simplifies the analysis but may lead to information loss, particularly among districts close to the national average.
- We assume the log-odds functional form is appropriate for estimating the probability of a district performing above average, but we test this assumption using non-parametric methods (e.g., random forest). The logistic regression makes the following assumptions:
    - Observations are independent of each other
    - There is little to no multicollinearity
- We also do not differentiate states by their years, as exploratory analysis and diagnostic plots indicate that the distribution of the outcome gap is approximately normal across frequencies. However, we acknowledge that if there are unobserved state‐year specific

trends, this assumption might overlook subtle variations. This is demonstrated by the plot below.



Overall, our empirical model focuses on prediction and policy relevance rather than structural identification. Nonetheless, the model is grounded in a realistic conceptual framework in which we assume and then test that educational outcomes depend on both resource adequacy and socioeconomic context.

Data Description:

Our analysis uses data from the District Cost Database (DCD), part of the School Finance Indicators Database. This dataset provides annual district-level information on K-12 public school budget gaps, student demographics, and standardized test score outcomes across roughly

159,035 U.S. school districts. For our study, we use data from 2009 to 2021, which offers the most complete and recent cross-sectional snapshot at the time of our analysis.

After removing missing observations and extraneous identifiers (year, district, NCES ID, etc.), our working dataset consists of 94,211 observations. Our analysis is based on a cross-sectional database rather than a panel, even with data going back to 2008-2009. It isn't a panel since each district included variables over a certain period, and we did not compare and contrast different periods with each other. A core strength of the dataset is that it includes standardized test performance scores, expressed as deviations from the national average (outcomegap), and an associated predicted spending value (predcost), which estimates the cost required for a district to achieve national average performance. This allows us to directly compare actual performance and funding adequacy across districts nationwide. However, there are a few important limitations:

- Districts in Vermont, Hawaii, and Alaska are excluded from the dataset due to data quality concerns or their structural uniqueness.
- The outcome gap variable, which we use to construct our binary outcome variable pos_outcome, is not available for every district, particularly in smaller or irregular reporting areas.
- The dataset lacks district-level instructional practice data (e.g., teacher experience, curriculum quality), which may confound the relationship between spending and outcomes, as well as individual schools or grades.

In an ideal setting, we would pair this dataset with additional longitudinal measures of student-level outcomes and local education policies. However, the current dataset provides sufficient variation and scope for a district-level, national benchmarking analysis.

Descriptive Statistics

The tables below present all the descriptive statistics for the primary variables used in our analysis across each of the four regions. This includes our outcome variable class_outcome, which is coded as above if a district's average test score exceeds the national average, and below otherwise. We also include key independent variables, such as the funding gap, poverty rate, and the percentage of students in special education (iep) or classified as English language learners (ell). Demographic variables reflect the racial and ethnic composition of the district’s student body.

Northeast:

```
    stabbr         ppcstot          predcost       fundinggap       outcomegap          enroll             pov         
 NJ      :4977   Min.   : 1510   Min.   : 3282   Min.   :-57732   Min.   :-1.18724   Min.   :   101   Min.   :0.01207  
 PA      :4459   1st Qu.:13149   1st Qu.: 6516   1st Qu.:  4405   1st Qu.: 0.04077   1st Qu.:   796   1st Qu.:0.05752  
 MA      :2613   Median :15304   Median : 7931   Median :  7017   Median : 0.27363   Median :  1718   Median :0.10008  
 NY      :1292   Mean   :15872   Mean   : 9036   Mean   :  6836   Mean   : 0.27501   Mean   :  2899   Mean   :0.11966  
 NH      :1218   3rd Qu.:17962   3rd Qu.:10044   3rd Qu.:  9689   3rd Qu.: 0.51821   3rd Qu.:  3401   3rd Qu.:0.16216  
 CT      :1211   Max.   :63592   Max.   :77423   Max.   : 55348   Max.   : 1.65754   Max.   :990145   Max.   :0.52482  
 (Other):1155                                                                                                          
      iep              ell                amind               asian              black                hisp        
 Min.   :0.0000   Min.   :0.000000   Min.   :0.000000   Min.   :0.00000   Min.   :0.000000   Min.   :0.00000  
 1st Qu.:0.1413   1st Qu.:0.002737   1st Qu.:0.000000   1st Qu.:0.00679   1st Qu.:0.009651   1st Qu.:0.01758  
 Median :0.1661   Median :0.008935   Median :0.001091   Median :0.01760   Median :0.020351   Median :0.04304  
 Mean   :0.1688   Mean   :0.023747   Mean   :0.002621   Mean   :0.04036   Mean   :0.062900   Mean   :0.09662  
 3rd Qu.:0.1928   3rd Qu.:0.025199   3rd Qu.:0.002542   3rd Qu.:0.04594   3rd Qu.:0.055118   3rd Qu.:0.10345  
 Max.   :0.6489   Max.   :0.371608   Max.   :0.653846   Max.   :0.70488   Max.   :0.956654   Max.   :0.96233  

     multi               pac                white           class_outcome
 Min.   :0.000000   Min.   :0.000000   Min.   :0.0001006   Below: 3631  
 1st Qu.:0.005093   1st Qu.:0.000000   1st Qu.:0.7072380   Above:13294  
 Median :0.015517   Median :0.000396   Median :0.8658465                
 Mean   :0.021059   Mean   :0.001331   Mean   :0.7751095                
 3rd Qu.:0.030228   3rd Qu.:0.001414   3rd Qu.:0.9342985                
 Max.   :0.281897   Max.   :0.124260   Max.   :1.0000000                
```

Midwest:

```
     stabbr          ppcstot          predcost        fundinggap          outcomegap           enroll      
 IL      : 6115   Min.   : 4188   Min.   : 3759   Min.   :-63065.0   Min.   :-2.78564   Min.   :   101  
 OH      : 4850   1st Qu.: 9265   1st Qu.: 7652   1st Qu.:  -889.5   1st Qu.:-0.09500   1st Qu.:   542  
 MI      : 4469   Median :10362   Median : 9236   Median :  1225.0   Median : 0.09115   Median :  1066  
 MO      : 3788   Mean   :10873   Mean   :10310   Mean   :   562.8   Mean   : 0.08305   Mean   :  2345  
 WI      : 3492   3rd Qu.:11839   3rd Qu.:11474   3rd Qu.:  2842.0   3rd Qu.: 0.27238   3rd Qu.:  2270  
 IA      : 3136   Max.   :83114   Max.   :79961   Max.   : 45914.0   Max.   : 1.46695   Max.   :405644  
 (Other):10749                                                                                          
      pov                iep              ell                amind              asian         
 Min.   :0.009836   Min.   :0.0000   Min.   :0.0000000   Min.   :0.0000000   Min.   :0.000000  
 1st Qu.:0.096601   1st Qu.:0.1195   1st Qu.:0.0004091   1st Qu.:0.0005198   1st Qu.:0.001916  
 Median :0.140533   Median :0.1424   Median :0.0059278   Median :0.0023774   Median :0.005128  
 Mean   :0.155274   Mean   :0.1457   Mean   :0.0287845   Mean   :0.0144953   Mean   :0.014141  
 3rd Qu.:0.197889   3rd Qu.:0.1683   3rd Qu.:0.0252525   3rd Qu.:0.0057471   3rd Qu.:0.011339  
 Max.   :0.664588   Max.   :0.6981   Max.   :0.6466666   Max.   :1.0000000   Max.   :0.517176  

     black               hisp             multi                pac                white        class_outcome 
 Min.   :0.000000   Min.   :0.00000   Min.   :0.000000   Min.   :0.0000000   Min.   :0.0000   Below:13486  
 1st Qu.:0.004405   1st Qu.:0.01596   1st Qu.:0.009804   1st Qu.:0.0000000   1st Qu.:0.8018   Above:23113  
 Median :0.010122   Median :0.03257   Median :0.020833   Median :0.0000000   Median :0.9074                
 Mean   :0.048831   Mean   :0.06811   Mean   :0.026739   Mean   :0.0009014   Mean   :0.8268                
 3rd Qu.:0.025398   3rd Qu.:0.06882   3rd Qu.:0.035952   3rd Qu.:0.0009731   3rd Qu.:0.9464                
 Max.   :0.996848   Max.   :0.96333   Max.   :0.554455   Max.   :0.2674961   Max.   :1.0000                
```

South:

```
     stabbr         ppcstot          predcost        fundinggap       pos_outcome          enroll      
 TX      :8121   Min.   : 1099   Min.   : 4706   Min.   :-64634   Min.   :0.0000   Min.   :   103  
 OK      :3822   1st Qu.: 8439   1st Qu.: 9625   1st Qu.: -6063   1st Qu.:0.0000   1st Qu.:   762  
 AR      :2317   Median : 9333   Median :12009   Median : -2614   Median :0.0000   Median :  1905  
 GA      :1791   Mean   : 9680   Mean   :13912   Mean   : -4232   Mean   :0.3468   Mean   :  6440  
 KY      :1540   3rd Qu.:10496   3rd Qu.:15810   3rd Qu.:  -406   3rd Qu.:1.0000   3rd Qu.:  4676  
 MS      :1318   Max.   :43085   Max.   :74618   Max.   : 33571   Max.   :1.0000   Max.   :357579  
 (Other):7261                                                                                    
      pov               iep               ell                amind               asian               black        
 Min.   :0.02059   Min.   :0.01882   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.00000  
 1st Qu.:0.17351   1st Qu.:0.09762   1st Qu.:0.005181   1st Qu.:0.001012   1st Qu.:0.001776   1st Qu.:0.01053  
 Median :0.23253   Median :0.12193   Median :0.020892   Median :0.002717   Median :0.004706   Median :0.05306  
 Mean   :0.23910   Mean   :0.12729   Mean   :0.046522   Mean   :0.038098   Mean   :0.009961   Mean   :0.16254  
 3rd Qu.:0.29673   3rd Qu.:0.15016   3rd Qu.:0.057168   3rd Qu.:0.007299   3rd Qu.:0.010102   3rd Qu.:0.22833  
 Max.   :1.00000   Max.   :0.44898   Max.   :0.941423   Max.   :0.909574   Max.   :0.496712   Max.   :0.99891  

      hisp             multi                pac           
 Min.   :0.00000   Min.   :0.000000   Min.   :0.0000000  
 1st Qu.:0.02896   1st Qu.:0.008455   1st Qu.:0.0000000  
 Median :0.07538   Median :0.020356   Median :0.0002258  
 Mean   :0.17369   Mean   :0.029328   Mean   :0.0010221  
 3rd Qu.:0.22048   3rd Qu.:0.036746   3rd Qu.:0.0010480  
 Max.   :1.00000   Max.   :0.979167   Max.   :0.1977901  
```

West:

```
> summary(df_west)
     stabbr         ppcstot          predcost        fundinggap        outcomegap           enroll             pov        
 CA      :7076   Min.   :   403   Min.   : 3791   Min.   :-83985   Min.   :-1.9612   Min.   :   101   Min.   :0.0000  
 WA      :1729   1st Qu.:  8293   1st Qu.: 8461   1st Qu.: -4827   1st Qu.:-0.4110   1st Qu.:   558   1st Qu.:0.1158  
 OR      :1338   Median :  9808   Median :10912   Median : -1156   Median :-0.1594   Median :  1837   Median :0.1760  
 AZ      :1332   Mean   : 10395   Mean   :13013   Mean   : -2617   Mean   :-0.1425   Mean   :  6060   Mean   :0.1931  
 MT      : 859   3rd Qu.: 11796   3rd Qu.:15106   3rd Qu.:  1308   3rd Qu.: 0.1072   3rd Qu.:  5692   3rd Qu.:0.2539  
 ID      : 743   Max.   :130849   Max.   :94433   Max.   :122599   Max.   : 1.5836   Max.   :687534   Max.   :0.7230  
 (Other):1431                                                                                                           
      iep              ell               amind               asian               black                hisp             multi          
 Min.   :0.0000   Min.   :0.00000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.0000   Min.   :0.000000  
 1st Qu.:0.0971   1st Qu.:0.02027   1st Qu.:0.003143   1st Qu.:0.004405   1st Qu.:0.003759   1st Qu.:0.1044   1st Qu.:0.009414  
 Median :0.1189   Median :0.08231   Median :0.007389   Median :0.011942   Median :0.008914   Median :0.2567   Median :0.026131  
 Mean   :0.1187   Mean   :0.14143   Mean   :0.042790   Mean   :0.043143   Mean   :0.022913   Mean   :0.3460   Mean   :0.037629  
 3rd Qu.:0.1407   3rd Qu.:0.21204   3rd Qu.:0.019017   3rd Qu.:0.038775   3rd Qu.:0.022125   3rd Qu.:0.5475   3rd Qu.:0.051522  
 Max.   :0.9839   Max.   :0.95674   Max.   :1.000000   Max.   :0.747499   Max.   :0.610323   Max.   :1.0000   Max.   :0.990291  

      pac              white        class_outcome 
 Min.   :0.000000   Min.   :0.0000   Below:9532   
 1st Qu.:0.000000   1st Qu.:0.2291   Above:4976   
 Median :0.001908   Median :0.5518                
 Mean   :0.003561   Mean   :0.5040                
 3rd Qu.:0.004255   3rd Qu.:0.7718                
 Max.   :0.106302   Max.   :1.0000                
```

Econometric Model:

Our primary hypothesis is that a district's test score performance, measured as its "outcome gap" (the number of standard deviations by which a district's performance differs from the U.S. national average), is closely related to school resources and demographic characteristics. To operationalize this, we construct a binary indicator, class_outcome, which takes the value "Above" if the outcome gap is positive and "Below" otherwise. Formally, we define it as:

$$\text{class\_outcome}_i = \begin{cases} \text{"Above"} & \text{if outcomegap}_i > 0, \\ \text{"Below"} & \text{if outcomegap}_i \leq 0. \end{cases}$$

The Basic Model:

We estimate the logistic regression model in its simplest form as mentioned in the theoretical model:

$$\log\left(\frac{\Pr(\text{Above}_i)}{1-\Pr(\text{Above}_i)}\right) = \beta_0 + \beta_1\text{FundingGap}_i + \beta_2\text{Enroll}_i + \beta_3\text{Pov}_i + \beta_4\text{IEP}_i + \beta_5\text{ELL}_i + \sum_{k=6}^{K}\beta_k X_{ik} + \varepsilon_i,$$

where the X_{ik} represents the various demographic percentages (e.g., for American Indian, Asian, Black, Hispanic, multiracial, and Pacific Islander students), with one category (e.g., White) omitted to avoid perfect collinearity.

Justification of Specification:

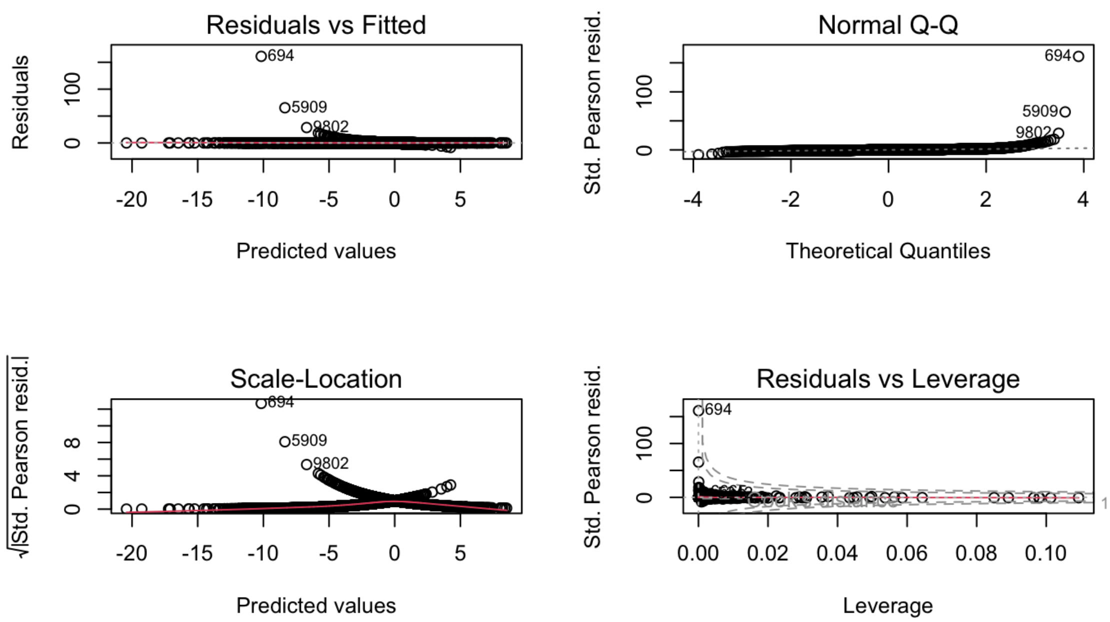
This logistic specification is appropriate because our dependent variable is binary. Its log-odds formulation permits us to interpret coefficient estimates as changes in the odds of a district performing above the national average. The model is motivated by prior empirical work in education economics that models student or school outcomes as functions of resource inputs and demographic context (e.g., Reardon et al., 2022; McMahon, 2000; Gautam et al., 2024). The use of the FundingGap variable, derived from the National Education Cost Model, allows us to focus not just on raw spending, but on whether districts meet the spending *targets* necessary for equitable outcomes - an innovation that increases comparability and policy relevance.

We do not include school- or district-level instructional strategies, administrative practices, or teacher experience due to data limitations; however, we believe that by including key controls (enrollment and several demographic variables), we aim to isolate the effect of the funding gap on performance. Unobserved factors are assumed to be uncorrelated with the included regressors, or at least not large enough to bias our findings significantly.

To ensure robustness, we perform stepwise selection (using AIC) to drop inconsequential predictors. Interesting results were obtained using this method, especially due to our decision to split our analysis regionally. In addition, we test the specification using various machine-learning methods (e.g., random forests, SVM) and compare their performance via cross-validation. This multi-method approach reduces the likelihood that our results are driven by model mis-specification and focuses on a higher accuracy of classification. We also do not use instrumental variables (IV) due to a lack of plausible instruments that satisfy the exclusion restriction. As such, the interpretation of β_1 is descriptive and policy-relevant but not strictly causal.

Methodology:

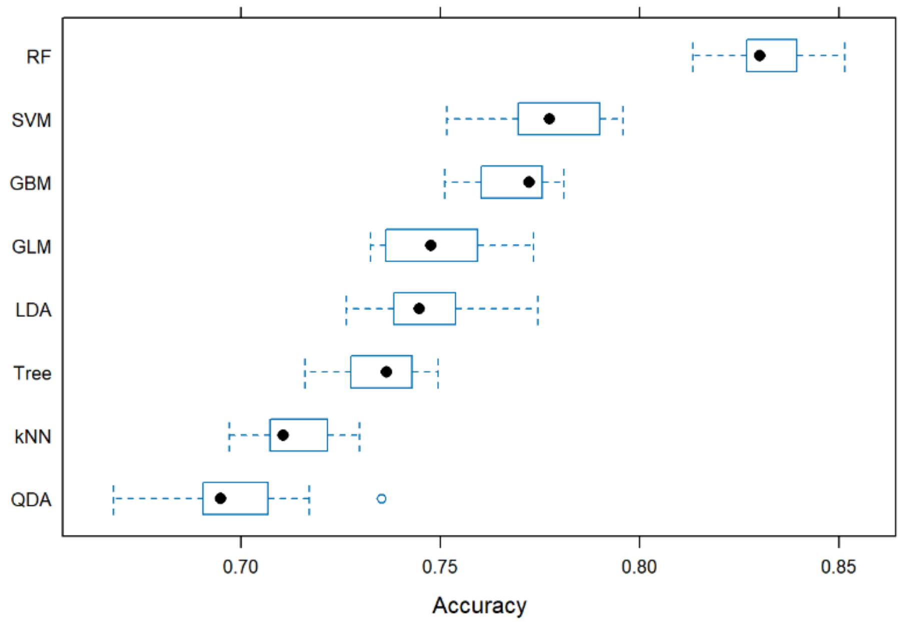
The first model to set the baseline was the logit model, and then it was remade using stepwise selection to ensure that we chose essential predictors for our region. We are aware that identification of causal effects relies on the assumption that, after controlling for enrollment and a comprehensive set of demographics, the remaining variation in the funding gap is exogenous. We recognize the possibility of omitted variable bias; however, extensive sensitivity analysis and cross-validation reduce the likelihood that unobserved heterogeneity drives our results.



Additionally, a key simplifying assumption in our analysis is that the distribution of the outcome gap does not vary significantly by year or state. As confirmed by our histograms and Q–Q plots, the outcome gap is approximately normally distributed across observations. Hence, we pool the data over time and do not include state–year interactions in our baseline specification.

Following the model with all predictors, we ran a cross-validation for threshold optimization whilst checking our model performance across test and training data, across multiple different models. Instead of using the conventional 0.5 cutoff for classification, we implement a 10-fold cross-validation on the training set to select an optimal probability threshold. Our CV routine examines thresholds from 0.30 to 0.70 in 0.05 increments and then chooses the best for classification, which varies across regions.

We also compare the logistic regression with alternative classifiers. The Random Forest model, for instance, achieves a higher accuracy and AUC compared to the others. Cross-validation across models (GLM, LDA, QDA, Tree, RF, GBM, SVM, and kNN) shows that Random Forest and SVM outperform simpler models in predictive accuracy without evidence of overfitting. Just in case there was overfitting in Random Forests, we also plotted the learning curve, plotting the Out-Of-Bag error against the fraction of training data, suggesting that as training size increases, the OOB error converges with the cross-validated error (both around 11–12%), confirming adequate generalization. Finally, we compared the accuracy across all of the models for each region. An example for the Southern region is depicted below.



Results

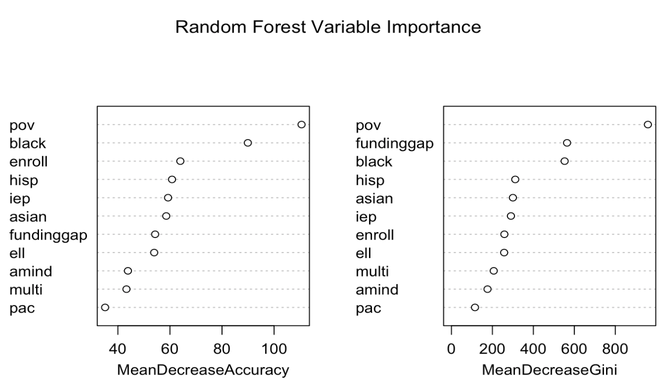
Prior to creating the models, the data was sectioned into four regions (West, Midwest, South, Northeast). This was because of the concern that there would be an overwhelming amount of heterogeneity across the entire country. As the results show, for different regions, the impact that particular variables have on the testing outcome gap varies. Perhaps this is an indication that economic policy regarding education should vary by geographical region (state level) rather than on a federal level.

Our main outcome of interest was the outcome gap between the state's educational outcomes versus the national average. The best model in terms of prediction accuracy was utilizing random forests. The interpretation of the plot on the left is that as the variable is randomly permuted, it measures how much the model's accuracy decreases. Higher values mean that the variable is more important for prediction, since when it is left out of the model, the performance suffers greatly. The Gini plot demonstrates how useful the value is in splitting the data at each node of the tree. Together, these two parameters show relative importance to each model.

Northeast:

Among all models tested, the random forest model proved to be the best performer for the Northeastern region, achieving an accuracy rate of 90.8%. The model also produced variable importance plots to highlight which features were most influential in predicting whether a district performed above or below the national average.

The Mean Decrease Accuracy (MDA) plot (left) illustrates how much each variable contributes to overall prediction accuracy. In this case, socioeconomic variables, such as poverty rate and racial/ethnic composition, were more influential than variables like the funding gap. These findings suggest that demographic context plays a stronger role in driving predictive accuracy than financial adequacy alone.



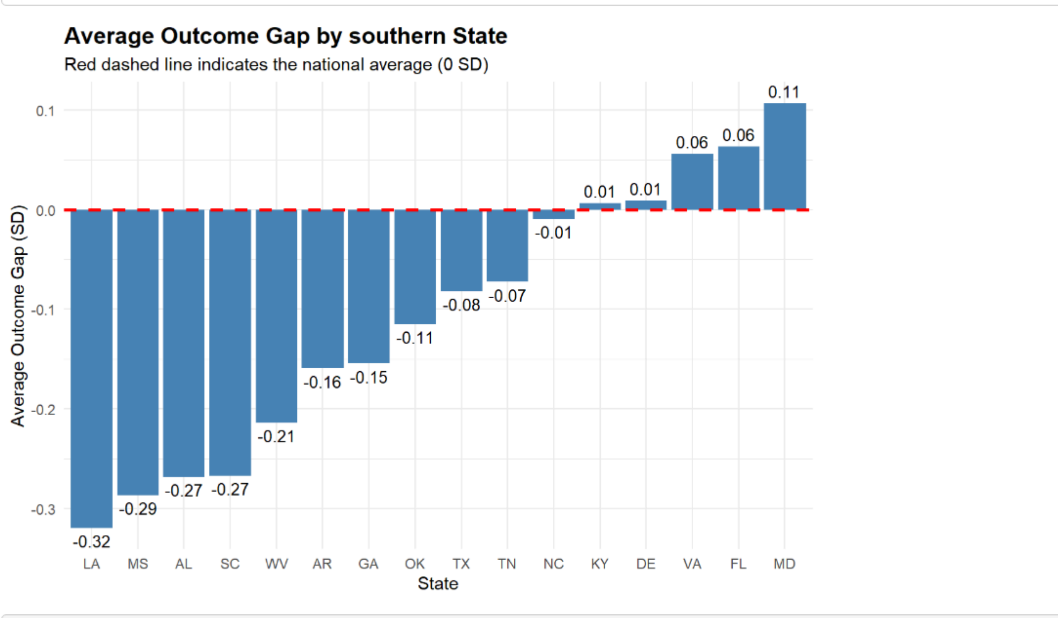
However, the Mean Decrease Gini (MDG) plot revealed that the funding gap was one of the most important variables for node purity, meaning it played a major role in creating effective splits within the decision trees. This distinction underscores an important nuance: a variable can be structurally important for class separation (MDG) without being highly predictive of model accuracy (MDA). Variables that rank highly on both metrics represent "gold standard" predictors; they help both separate classes and improve predictions.

This pattern may reflect broader structural differences among suburban districts, which tend to be more financially stable and well-resourced. These trends are consistent with the logistic regression results, where the funding gap variable had a negative coefficient, suggesting that districts spending less than the predicted adequate level are less likely to outperform the national average. Additionally, the collinearity between racial composition and poverty may influence both prediction and interpretability.

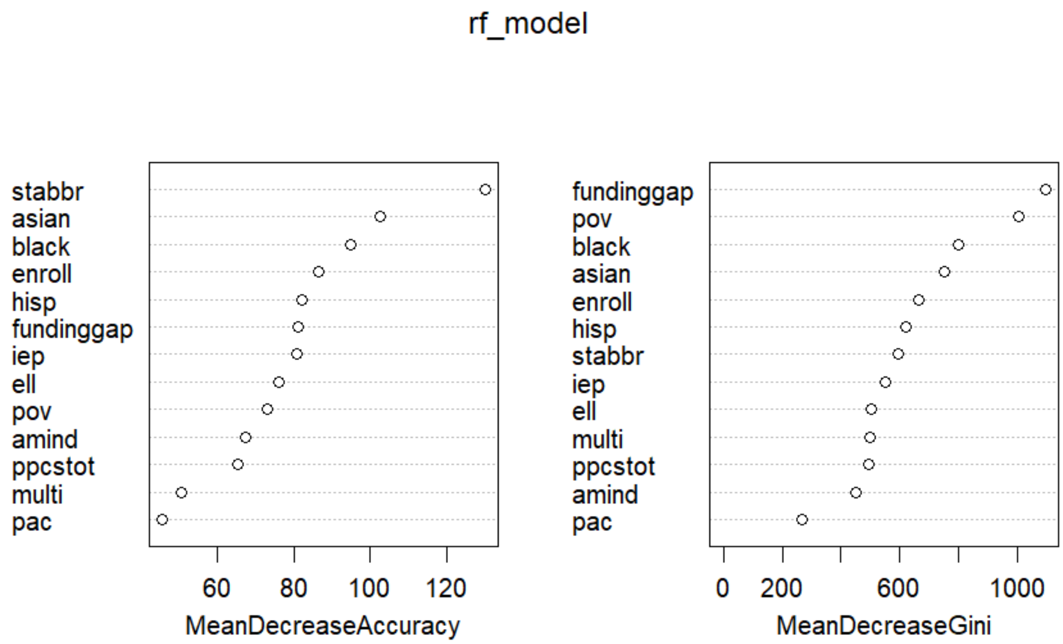
```
Coefficients:
              Estimate Std. Error z value
(Intercept)  5.790e+00  1.716e-01  33.747
fundinggap  -1.129e-05  7.691e-06  -1.468
pov         -1.718e+01  4.846e-01 -35.440
iep         -8.953e+00  6.471e-01 -13.835
ell         -5.964e+00  9.381e-01  -6.358
amind       -8.642e+00  2.437e+00  -3.546
asian        1.731e+01  1.177e+00  14.700
black       -7.434e+00  3.260e-01 -22.802
hisp        -7.766e-01  2.465e-01  -3.151
multi       -2.574e+00  1.178e+00  -2.185
            Pr(>|z|)
```

While random forests provide a powerful framework for assessing variable importance, logistic regression and marginal effects analysis offer greater interpretability for policymakers. Logistic regression tells us how a one-unit change in the predictor impacts the log odds of whether the state's educational outcomes are above or below the national average. We estimate how a one-unit change in a given variable (e.g., poverty rate) affects the log odds of a district performing above the national average. These coefficients are depicted in the plot above. This makes it easier to understand not just which factors matter but how much they matter and in what direction.

South



Our main outcome of interest was the outcome gap between the state's educational outcomes versus the national average. The best model in terms of prediction accuracy was utilizing random forests. At an accuracy of 85%, the most important predictors were found to be the state dummies, funding gap, black and asian races, and the poverty rate. The interpretation of



the plot on the left is that as the variable is randomly permuted, it measures how much the model's accuracy decreases. Higher values mean that the variable is more important for prediction since when it is left out of the model, the performance suffers greatly. The Gini plot demonstrates how useful the value is in splitting the data at each node of the tree. Together, these two parameters show relative importance to each predictor.

While random forests provide a good structure for understanding variable importance, another interpretation that could be useful for policymakers would be the marginal effects of one variable on the response. For this, using the logistic regression model proves invaluable. Below is the logistic regression for the outcome gap on all predictors for the southern region. As shown, all predictors were statistically significant in their prediction. It's worth noting that the poverty rate, special education, and the black and asian dummies had the largest substantive impacts(measured in the magnitude of the coefficient). Specifically, poverty rate, special education learners, and black had strong negative coefficients, while asian had a strong positive coefficient. This suggests that there could be evidence of racial discrimination in what states get funding in the south, as well as an association between poverty rate and lower educational outcomes.

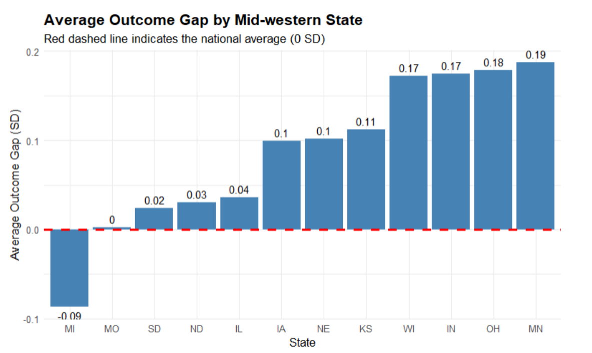
```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  2.016e+00  1.061e-01  18.998  < 2e-16 ***
fundinggap   4.117e-05  8.378e-06   4.915 8.89e-07 ***
enroll       1.144e-05  1.318e-06   8.677  < 2e-16 ***
pov         -5.335e+00  3.299e-01 -16.171  < 2e-16 ***
iep         -6.324e+00  5.742e-01 -11.013  < 2e-16 ***
ell         -3.040e+00  5.220e-01  -5.823 5.77e-09 ***
amind       -1.863e+00  2.088e-01  -8.920  < 2e-16 ***
asian        4.193e+01  1.919e+00  21.857  < 2e-16 ***
black       -4.485e+00  1.802e-01 -24.886  < 2e-16 ***
hisp        -1.934e+00  1.394e-01 -13.868  < 2e-16 ***
multi        2.000e+00  4.813e-01   4.154 3.26e-05 ***
pac         -1.122e+01  5.395e+00  -2.080   0.0376 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

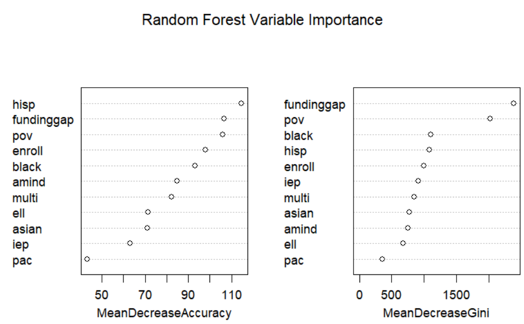
Midwest

For the midwestern states the best model for classification was the Random Forest model with the following results. It is important to note that only numeric variables were determined in the Random Forest model, thus, every midwestern district was observed irregardless of state.

Accuracy	0.807
Confidence Interval	0.7995, 0.8143
Sensitivity	0.724
Specificity	0.855
AUC	0.886

Compared to other regions, the accuracy was not as high. This could be because of the breadth of the region (as this region contained the most observations) or there may still be much heterogeneity within the school districts of the midwest. The specificity was higher, implying that the model was more successful in correctly predicting if a district would underperform. In terms of variables, the funding gap seems to be the most important variable. This makes intuitive sense as the state funding likely affects the resources available to the student at the district level. Poverty is also another important predictor that supports the notion that economic factors are the most important in predicting whether a district will perform higher or lower than the national average. The proportion of English learners (ell) and the proportion of students that are pacific





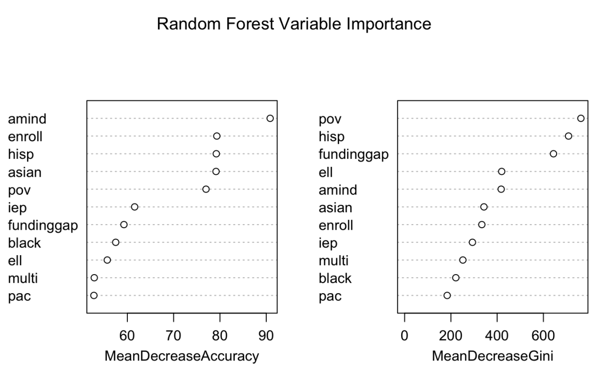
islanders (pac) are not as impactful, likely due to the racial makeup of the midwestern region.

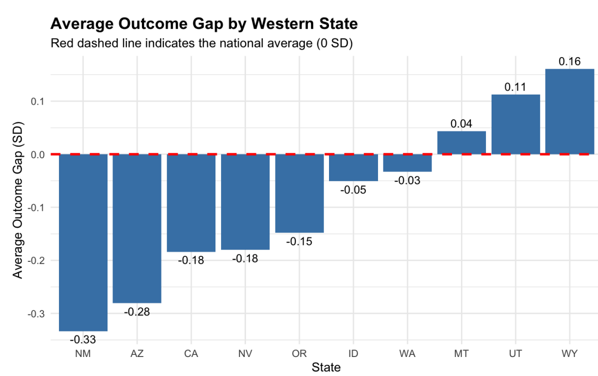
For the sake of interpretability, here is the generalized linear model (post stepwise subset selection) with the p-values for each predictor variable.

```
Call:
glm(formula = formula(model_step), family = binomial, data = train_midwest)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.703e+00  9.163e-02  29.504  < 2e-16 ***
fundinggap   1.677e-04  8.585e-06  19.539  < 2e-16 ***
enroll       8.271e-06  2.035e-06   4.065 4.81e-05 ***
pov         -8.346e+00  3.284e-01 -25.412  < 2e-16 ***
iep         -3.355e+00  4.459e-01  -7.525 5.29e-14 ***
ell          1.169e+00  6.063e-01   1.928   0.0538 .
amind       -4.511e+00  3.801e-01 -11.868  < 2e-16 ***
asian        1.125e+01  9.876e-01  11.390  < 2e-16 ***
black       -3.388e+00  2.888e-01 -11.731  < 2e-16 ***
hisp        -4.497e+00  3.056e-01 -14.716  < 2e-16 ***
multi       -3.565e+00  7.356e-01  -4.846 1.26e-06 ***
pac         -2.515e+01  6.306e+00  -3.988 6.67e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 33725  on 25620  degrees of freedom
Residual deviance: 24928  on 25609  degrees of freedom
AIC: 24952

Number of Fisher Scoring iterations: 6
```

To conclude for the Midwest, poverty and funding to the districts (which impacts school resources) seem to have the most impact on a district’s performance relative to the national standard. This seems to be a common theme throughout all of the regions, however, the differences in demographics should be considered to propose effective economic policies.

West:





Specifically in the western region, our analysis reveals that variations in school resources and demographic characteristics significantly affect district test score performance relative to the national benchmark. If the goal is to improve test score performance, focusing on the top-ranked variables (e.g., enrollment size, native american, etc) as they are ranked higher in the MDA and permuting those variables would decrease the model accuracy, while poverty seems to play the most significant role in splitting the data. This confirms that socioeconomic status and resource allocation are major determinants of a district's likelihood of performing above or below the national average. In our logistic regression model, key variables - particularly the funding gap - were highly significant. Overall, the logistic model achieved an accuracy of about 80.14% and an AUC of 0.8825. Cross-validation revealed that optimizing the probability threshold to 0.45 (instead of the conventional 0.50) improved the classification sensitivity to 82.79% and specificity to 73.59%, indicating that even small shifts in the decision boundary can lead to better discrimination. Comparing across models, ensemble methods outperformed linear classifiers. The random forest model, for example, attained an accuracy of 87.84% with an AUC of 0.9431 and an out-of-bag error of 11.91%, suggesting minimal overfitting and robust predictive power. In contrast, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) delivered lower performance with very skewed misclassification rates of high false negatives with accuracies of 79.54% and 69.64% (AUCs of 0.8774 and 0.8362, respectively).

Conclusion:

Our study demonstrates that both school resources and demographic characteristics play a significant role in determining whether a district performs above or below the national average. By constructing a binary classification framework and utilizing a range of econometric and machine-learning models, including logistic regression, random forests, and SVM, we provide robust evidence that factors such as the funding gap, enrollment, and socio-economic variables are key determinants of district performance. Notably, the random forest model achieved the highest classification rate across the board, underscoring its superior discriminative power relative to more traditional methods.

These empirical findings have clear policy implications. For example, the high variable importance of poverty rates and funding gaps suggests that targeted interventions to alleviate resource deficiencies and support economically disadvantaged districts could markedly improve educational outcomes. Our regional analysis further reveals that local institutional environments may influence performance differently than national aggregates, a finding that invites more granular, context-specific policy approaches. Despite these compelling insights, our analysis is subject to certain limitations. The use of cross-sectional data and pooling across years, while justified by diagnostic checks, may mask important temporal dynamics. Future research could extend our work by incorporating panel-data methods and exploring the causal impacts of policy changes over time.

Overall, our methodological approach, with robust cross-validation and model diagnostics, provides confidence in these results. The evidence suggests that even relatively small improvements in resource allocation and targeted socio-economic policies can contribute to a district's likelihood of achieving above-average educational outcomes, making a strong case for concerted policy attention in these areas.

Appendix:

State Regions:
Northeast: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, Pennsylvania

Midwest: Indiana, Illinois, Michigan, Ohio, Wisconsin, Iowa, Nebraska, Kansas, North Dakota, Minnesota, South Dakota, Missouri

South: Delaware, District of Columbia (not included due to lack of observations), Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee

West: Arizona, Colorado, Idaho, New Mexico, Montana, Utah, Nevada, Wyoming, Alaska, California, Hawaii, Oregon, Washington

Years of Observation: *2009-2021*

Number of schools analyzed across all districts : *159,035 (94211 after cleaning)*

Variable Definitions
List of Definitions can be found of page 8 of the *DCD Codebook*
https://www.schoolfinancedata.org/wp-content/uploads/2024/04/DCD_Codebook_2024.pdf

APA Citations:
Dey, N. (2021). Resource and performance: Investigating the relationship between school resource quality and student math achievement using PISA data. New England College. Retrieved from
https://www.nec.edu/wp-content/uploads/Dey-Resource-and-Performance_2021.pdf

McMahon, W. W. (2000). The impact of school resources on the learning of inner city children. Journal of Education Finance, 25(4), 463–475. Retrieved from https://eric.ed.gov/?id=EJ598987

The Commonwealth Institute. (2023). Unequal opportunities: Fewer resources, worse outcomes for students in schools with concentrated poverty. Retrieved from https://thecommonwealthinstitute.org/tci_research/unequal-opportunities-fewer-resources-worse-outcomes-for-students-in-schools-with-concentrated-poverty/

Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., & Saliba, J. (2022). Trends in academic achievement across the U.S.: 2009 to 2019 (CEPA Working Paper No. 22-02). Stanford Center for Education Policy Analysis. Retrieved from https://cepa.stanford.edu/sites/default/files/wp22-02-v102022.pdf

Gautam, R., McCombs, L., & Pennington, M. (2024). Exploring educational equity in Georgia: A machine learning analysis of K–12 achievement disparities. arXiv preprint arXiv:2402.01710. https://arxiv.org/abs/2402.01710