

Project Prospectus

Stefano Bastianelli, Brooke Brehm, Derek Caramella, Miguel Novo Villar, & Kenzie Potter

I. INTRODUCTION

We aim to model the way government expenditures and labor appropriation impacts secondary education graduation rates. We hypothesize a subtle marginal return from municipality expenditures contribute to graduation rates following a threshold; we aim to contribute to the existing literature that advise local municipalities to develop an optimal taxation scheme. Following rigorous preprocessing, we will make observations about the graduation rate by evaluating descriptive statistics, features correlations, and other visualizations. Moreover, we will utilize optimized machine learning models using grid searches for models such as: ElasticNet, SVR (Support Vector Regression), full connected neural networks, and ensemble models, we plan to predict graduation rates for New York public schools based on the school conditions described in the data set.

II. DATA

The data set chosen for this project is the 2020 Report Card Database from the New York State Department of Education. The data contains six tables: teaching certification, school groupings, experience of teachers and principles, expenditures per pupil, accountability status, and high school graduation rates; each data set contains between 5542 and 158920 records for the 2019-2020 school year in New York public schools. Each tables provides features describing teachers, students, schools, and districts by utilizing foreign key relationships. The teaching certification features include amounts and percentages of teachers in each school, poverty levels, experienced and inexperienced teachers and principles, expenditures from various government municipalities. Moreover, the data describes each school with an accountability status: good standing, potential target district, potential target support and improvement (TSI), potential comprehensive support and improvement (CSI), TSI, CSI, and closing school. Graduation rates for each school are grouped by all students, race, year cohort (4, 5, 6, and combined years to graduate), income level, and English as a second language for comparison, though not all schools have data for the entire set of groupings.

III. LITERATURE REVIEW

It might seem controversial that teacher quality plays an important role in student performance, but it is a persistent finding that qualified teachers directly impact educational outcomes (Boyd, Lankford, Loeb, Rockoff, Wyckoff, 2008). Different paths into teaching and changing requirements have shaped the educational landscape and directly affected the learning condition across students in the state of New York - based on teacher experience, income and standardized test

results. Analyzing, predicting and grouping different patterns across schools could potentially help policy makers update requirements and allocate resources more efficiently to create a more balanced and better educational system in the state of New York.

IV. HYPOTHESIS

We aim to answer questions we have about the percentage of students successfully graduating from New York schools with the goal of more students graduating. Does a \$1000 increase in expenditure per student have a noticeable positive affect on how many students graduate? Does the number of teachers instructing outside of their area of certification negatively impact graduation rates? Is there a number threshold of those teachers outside of certification for negatively or positively impacting the number of students graduating?

V. EXPECTATIONS AND RESULTS

Our initial expectation was that schools with higher expenditures per student will have higher graduation rates. We expected the number of teachers instructing outside of their areas to negatively impact the graduation rates of schools. So far, through our data preprocessing and model testing, we are uncertain if these initial expectations will hold. There does not seem to be a positive correlation between increase in expenditure and graduation rate, in fact there may be a negative correlation between the two variables. There also seems to be a negative correlation between graduation rate and the number of uncertified teachers. So far, we have finished the preprocessing of the data and have started creating models and predictions as well as data visualizations. We are hoping that our models will give us better insight on the relationships between our explanatory variables and graduation rate.

VI. TIMELINE

Goal	Date	Participants
Data Preprocessing Pipeline	4/15/2022	Caramella
Prospectus	4/15/2022	Brehm, Novo Villar, Potter
Data Visualizations	4/24/2022	Bastianelli, Brehm, Caramella, Novo Villar, Potter
Presentation	4/24/2022	Bastianelli, Brehm, Caramella, Novo Villar, Potter
Data Modelling/Prediction	5/2/2022	Bastianelli, Brehm, Caramella, Novo Villar, Potter
Final Paper	5/2/2022	Bastianelli, Brehm, Caramella, Novo Villar, Potter

Fig. 1. Project Timeline Detailing Goals, Completion Dates, and Participating Team Members

REFERENCES

- [1] Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). Narrowing the gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management*, 27(4), 793–818. <https://doi.org/10.1002/pam.20377>