



SESIoN

SocioEconomic Status Impact on New York State Academic
Performance

Dave Dyer, Nick Orangio, Ben Spivey, Kshitij Srivastava, Vuong Tran

Introduction

- Socioeconomic status (SES) and its impact on academic performance is well studied and documented – lower SES is often associated with lower academic performance and vice versa^{1,2,5,6,7}
- Research on SES and academic performance is heavily academic and utilizes classical statistical techniques^{3,4} – by applying modernized machine learning techniques and interactive data visualization, we will improve the insights into the relationships between SES and academic performance
- We intend to improve on and add to previous research on this topic through the thoughtful application of machine learning and data visualization
- We will focus on the state of New York, however we suspect our findings will be applicable and reproducible for other states in the United States



Heilmeier Questions



WHAT ARE YOU TRYING TO DO

Model and visualize the impact of SES on academic performance in New York State, creating an engaging and interpretable product for end users



HOW IS IT DONE TODAY

Heavily academic, dry research papers with classical statistical methods and without interactive visualizations



WHAT IS NEW IN YOUR APPROACH

Modern machine learning techniques and engaging, interactive data visualizations using D3



WHO CARES?

Health-minded citizens, researchers, parents, teachers, school administrators, and lawmakers



WHAT IMPACT WILL IT MAKE?

Effectively informing stakeholders will lead to more effective policymaking, leading to effective policy and increased academic performance

Heilmeier Questions



WHAT ARE THE RISKS?

The ecological fallacy, interpretability of analysis, and correlated factors in models



HOW MUCH WILL IT COST

Outside of man-hours, costs are expected to be negligible as the project uses publicly available data and computing



HOW LONG WILL IT TAKE

Lit survey, proposal, presentation – 1 week
Data cleaning / analysis – 90 ± 155 hrs
ML / modeling – 90 ± 155 hrs
Visualization – 144 ± 248 hrs



WHAT ARE THE CHECKS FOR SUCCESS?

Cleaned data, a preliminary model, and a functioning D3 choropleth by end of October

What are you trying to do?

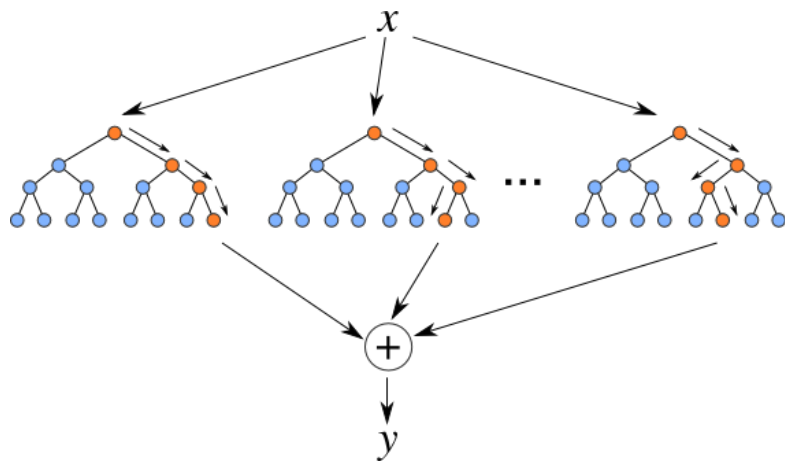
- Model and visualize the impact of SES on academic performance in New York State, creating an engaging and interpretable product for end users
- Improve on current models using machine learning and elegant, interactive data visualization
- Apply state of the art techniques to a field that is heavy in academic/economic research, resulting products hard to interpret for the end user
 - Modern, interpretable machine learning techniques
 - Interactive and engaging data visualization using JavaScript / D3

How is it done today?

- Current academic performance models are statistics based (ANOVA, t-test, regression, etc)^{3,4}
- Current models can fail to capture random effects and confounding factors and do not utilize more capable machine learning techniques
- SES is typically measured by factors such as:²
 - Free and Reduced Lunch Ratio (FRL)
 - Family education levels
 - Occupation
 - Income
- Academic performance is typically measured by:²
 - GPA
 - Standardized testing
- Data is presented in a research paper, often very dry and without interactive visualization

What's new in your approach?

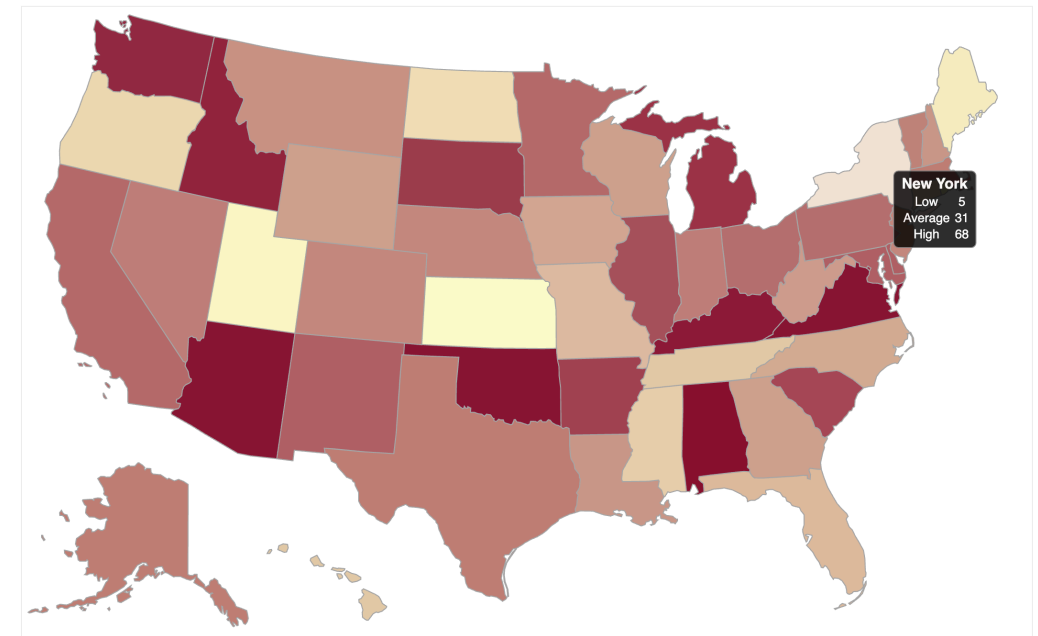
- We will apply modern machine learning techniques (i.e. regularized regression, regression trees, random forests, etc)
- We will create an interactive visualization that users can engage with to better understand the impact of SES on academic performance, ultimately having a larger impact on the end user than a research paper



Random forest regression

<https://levelup.gitconnected.com/random-forest-regression-209c0f354c84>

US State Map



Interactive choropleth in D3

<http://bl.ocks.org/NPashaP/a74faf20b492ad377312>

What difference will it make?

- Our project intends to broaden the scope of awareness of factors such as SES that impact academic performance in a United States education system that is increasingly failing relative to other leading countries
- Health-minded citizens, parents, teachers, school administrators, researchers and lawmakers will find this project informative and influential in regards to financial decisions and budgeting to achieve high levels of academic performance^{1,8,10}
- Effectively informing stakeholders is a crucial step in effective policymaking



What are the risks and payoffs?

- The ecological fallacy – a misinterpretation of results that apply aggregate findings to individual outcomes when it is not appropriate to do so²
- Interpretability of “black box” machine learning models to the intended audience of the project⁹
- Confounding variables due to suppressed data



How much will it cost?

- Per A Simple Cost Model for Machine Learning Applications [insert reference], a straightforward net present value cost model can be applied for machine learning applications:

$$NPV = C_0 + \sum \frac{C_t}{(1+r)^t}$$

- Where C_0 is the cost of equipment and C_t is the decision cost
- We are working with freely available data and intend to use compute and storage we already own, therefore the NPV and cost should be \$0

• Plan of Activities & Roles •

Timeline

- Literature survey, proposal and proposal presentation – 1 week
- Data loading, cleaning, design, and preliminary modeling – 90 ± 155 hrs
- Statistics and machine learning modeling – 90 ± 155 hrs
- D3 visualization, interactivity, testing, iteration – 144 ± 248 hrs¹¹

Checks for Success

- Cleaned data, preliminary model, functioning choropleth in D3 by end of October

Roles

- Documentation, design, organization, presentation, oversight – Nick, Dave, and Shitij
- D3 visualization and Python modeling – Vuong, Ben, Shitij, Dave, and Nick (whole team)

References

- [1] Jinnai, Y. (2014). "Direct and indirect impact of charter schools' entry on traditional public schools: New evidence from North Carolina." *Economic Letters*, 124, 452-456.
- [2] Sirin, S. R. (2005). "Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research." *Review of Educational Research*, 75(3), 417-453.
- [3] Hearn, J. C. (1988). "Attendance at Higher-Cost Colleges: Ascribed, Socioeconomic, and Academic Influences on Student Enrollment Patterns." *Economics of Education Review*, 7(1), 65-76.
- [4] Farooq, M.S., Chaudhry, A.H., Shafiq, M., Berhanu, G. (2011). "Factors Affecting Students' Quality of Academic Performance: A Case of Secondary School Level." *Journal of Quality and Technology Management*, VII(II)01-04.
- [5] Benner, A. D., Boyle, A. E., Sadler, S. (2016). "Parental Involvement and Adolescents' Educational Success: The Roles of Prior Achievement and Socioeconomic Status." *Journal of Youth and Adolescence*, 45, 1053-1064.
- [6] Zwick, R. (2012). "The Role of Admissions Test Scores, Socioeconomic Status, and High School Grades in Predicting College Achievement." *Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana*, 49(2), 23-30.
- [7] Battle, J., Lewis, M. (2008). "The Increasing Significance of Class: The Relative Effects of Race and Socioeconomic Status on Academic Achievement." *Journal of Poverty*, 6(2), 21-35.
- [8] Lee, J. O., Kosterman, R., Jones, T.M., Herrenkohl, T.I., Rhew, I.C., Catalano, R.F., Hawkins, J.D. (2016) "Mechanisms linking high school graduation to health disparities in young adulthood: a longitudinal analysis of the role of health behaviours, psychosocial stressors, and health insurance." *Public Health*, 139, 61-69.
- [9] Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., Kagal, L. (2018). "Explaining Explanations: An Overview of Interpretability of Machine Learning." *IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, Turin, Italy, 80-89.
- [10] Domanico, Ray (2020). "NYC Student Achievement: What State and National Test Scores Reveal." *The Manhattan Institute for Policy Research* <https://eric.ed.gov/?id=ED604331>.

References

- [11] Singla, K., Bose, J., Naik, C. (2018) "Analysis of Software Engineering for Agile Machine Learning Projects." *15th IEEE India Council International Conference (INDICON)*, Coimbatore, India, 1-5, doi: 10.1109/INDICON45594.2018.8987154.
- [12] Berkowitz, R., Moore, H., Astor, R.A., Benbenishty, R. (2017). "A Research Synthesis of the Associations Between Socioeconomic Background, Inequality, School Climate, and Academic Achievement" *Review of Educational Research*, 87(2), 425–469.
- [13] Winters, M. (2012). "Measuring the effect of charter schools on public school student achievement in an urban environment: Evidence from New York City." *Economics of Education Review*, 31, 293-301.
- [14] West, M. (2016). "Schools of choice: expanding opportunity for urban minority students." *Education Next*, 16(2).
- [15] Domingos, P. (1998). *How to Get a Free Lunch: A Simple Cost Model for Machine Learning Applications*. Workshop on AI Approaches to Time Series Problems, 15th National Conference Artificial Intelligence (AAAI-98) and International Conference on Machine Learning (ICML-98).