

Literature Review: Using Artificial Neural Networks and Support Vector Machines to Predict School Performance on the A-F Report Card

Tom Curran

April 23, 2018

University of Chicago

Perspectives on Advances Computation Research (MACS 30200)

Evaluating school performance has become one of the most hotly contested topics in education policy in recent memory. Stemming from the controversial No Child Left Behind (NCLB) policy enacted in the early part of the 21st century, School Administrators, Policy Makers and Teachers alike have been highly critical of approaches to evaluating school performance. While most states have different systems of evaluating school and school district performance, most states rely heavily on the results of state test scores, which is a consequence (and criticism) of NCLB. Teachers and School Administrators argue that in relying on test scores alone is an unfair advantage to wealthy children. Their criticism is well founded and research shows that the effects of poverty have a highly significant effect on student outcomes and performance. This becomes very apparent in states like Oklahoma, where school districts are graded on an A-F scale based on their state test outcomes for the academic year.

However, the reality on both sides of the argument is that there is a very complex relationship between organizational behavior, socio-economic status and education policy that plays a role in performance of a school and school district. The relationship between the fields and the many variables within is not always linear and more importantly hard to isolate. With the advent of techniques like Artificial Neural Networks (ANN) and Support Vector Machines we can delineate those complex relationships that dictate school performance. While there is an abundance of program evaluation literature, it tends to use overly simplified linear regression models, missing the key relationships as previously mentioned. There is an emerging field within education research called Education Data Mining and Learning Analytics that employ these new techniques (ANN and SVM). However, their focus has been primarily to predict individual student performance. I plan to adapt the prior research and models to predict the outcomes of individual schools and districts.

I am primarily interested in finding if a school Ad Valorem (Property Tax) plays a significant role in the performance of a school. Using an ANN and SVM will allow to identify the role of Ad Valorem taxes as well as the importance of Ad Valorem's interactive effects with other important socio-economic and demographic variables.

One reason as to why the Ad-Valorem tax is of interest is because it provides schools with more revenue, which often goes towards purchasing more teacher and human capital. The prevailing school of thought in education is that smaller student teacher ratio leads to greater gains in in student achievement, particularly among low-income students. One of the earliest research papers on the relationship between class size and student achievement was conducted by Gene Glass and Mary Lee Smith in 1979 [Cite]. Research studies like the Tennessee STAR experiment furthered Glass and Smith?s findings, pushing policy makers to focus on reducing class size as a mechanism for promoting student achievement [CITE].

However, there has also been significant literature that counter?s the class size argument. The research conducted by Caroline M. Hoxby shows that natural variation in populations result in class size having little to no effect on student achievement [CITE]. Christopher Jepsen and Steven Rivkin find that there is a significant tradeoff between teacher quality and class size. While Jepsen and Rivkin find that class size does impact achievement the achievement is negatively affected when there the share of teachers in a school have little or no experience [CITE].

Both sides of the program evaluation literature borrow heavily from previous education policy research. As such, their methodologies rely almost entirely on regression models. As previously mentioned, using regression techniques like linear regression, logistic regression, diff-in-diff and others may not fully capture causal mechanisms. The variation of outcomes produced by these models suggests at a deeper underlying relationship that cannot be captured by traditional program evaluation methods. As such, I believe it prudent to apply new techniques and methodologies to deepen our understanding of program and institutional effectiveness. As such, much of the literature used as a foundation for the project borrows from data mining, artificial intelligence and machine learning, specifically using literature in the emerging field of Educational Data Mining. Educational data mining (EDM) and learning analytics (LA) typically refer to a variety of techniques quantitative researchers use with the education field. EDM and LA were born as an alternative to Bayesian and frequentist approaches to education data. In some

research fields, the term "data mining" is a negative one, where hypothesis is formed around the data whereas a hypothesis should be formed prior to conducting data analysis. However, data mining has a positive historical precedent rooted in exploratory data analysis that has its own set of techniques for generalizability and validation [CITE]. Furthermore, EDM and LA borrow their methodologies from a variety of other fields including bioinformatics, and business.

When it comes to EDM and LA, it is important to select the appropriate tools so that proper methodologies can be used in educational data. Slater et al suggest several such tools for each step of the EDM process, from data pre-processing, to analysis and visualization. Among the author's suggested set of tools are Python and Jupyter notebook. Slater et al share several insights as to the benefits of using Python and the Jupyter environment. Advantages include python's vast resources for manipulating data sets for pre-processing, python notebook's record of analyses conducted and perhaps most importantly, its computational power, especially when it comes to large data sets. Slater et al continue their article by summarizing the benefits and drawbacks to several tools when conducting text mining, network analysis or visualizations projects. Slater et al's review of EDM and LA tool kits provide a detailed landscape for tools to choose from that is best suited for the project at hand.

The survey of tools is complimented by Behdad Bakhshinategh et al's work Educational data mining applications and tasks: A survey of the last 10 years. Bakhshinategh breaks down EDM applications into distinct subgroups, among them are what he calls "Decision Support Systems". In essence, Bakhshinategh et al make the case for data mining and learning analytics to help educators and administrators make decisions in an education context. This subgroup's ultimate goal is to engage stakeholders by using data to provide feedback, creating alerts, generating recommendations, and enhancing courseware. In Bakhshinategh says that the target of the decision support systems is primarily focused on the teacher but can be applied to administrator. In the case of this research paper, the target population are policy makers. Since decision support is concerned with any and all stakeholders, policy makers are a critical component to that, it is reasonable

to apply Bakhshinategh et al's framework to the question at hand for entire schools.

Modeling and predicting performance is particularly challenging in the context of education. Particularly because there are many opportunities to introduce bias into a model. Lou et al provide a statistical framework in which to reduce variables that they call "discrimination aware classifiers". The issue becomes prevalent when creating rule-based groups of students that necessarily depend on sensitive characteristics. For example, creating achievement levels that necessarily depend on student's gender and/or demographic identity. As Lou et al state, "It is desirable to keep the sensitive attribute during the training of a classifier to avoid information loss but decrease the undesirable correlation between the sensitive attribute and the class attribute when building the classifier" [CITE]. In other words, it becomes increasingly necessary to use sensitive attributes in order to avoid information loss but creating decision rules from sensitive information can lead to discriminatory practices. As such, Lou et al suggest these sensitive attributes be used as an "information carrier and not a distinguishing factor". As such, Lou et al provide a method for developing a Discrimination Aware Association Rule classifier (DAAR) that measure the discrimination severity of a given rule. These DAAR rules are used to predict student achievement in a computer science class. They show that there is limited loss of predictive accuracy between non-discriminatory aware rules and the DAAR method [CITE].

DAAR rules are of particular importance given the unit of observation for this research paper. Much like students in Lou et al's paper, districts and schools in Oklahoma vary widely in terms of their financial characteristics, teaching staff, and most importantly students along with many other characteristics. For example, in Oklahoma City Public Schools 31.9

% of students are English Language Learners (ELL) [CITE], while the school district directly to the north of OKCPS has only 4.4\{\}% ELL students[CITE]. Even within districts there is great variation in populations. Within OKCPS there is great variation in school populations. Star Spencer High School has 2.5% ELL[CITE] students whereas U.S. Grant High School has 30.8% ELL students [CITE]. Lou et al's methods for

minimizing discriminatory variables becomes increasingly necessary to build the model, as well as produce a model that can be deployed in a policy setting.

Yorek and Ugulu provide a qualitative definition of artificial neural networks and their applicability in the education space. Their definitions prove especially useful since they attempt to create an artificial neural network (ANN) to measure student attitudes and qualitative data. An artificial neural network is a mathematical model that takes after the biological function and physical structure of neurons in the brain. Yorek et al concisely express the advantages of using ANN because they can be "used to model complex relationship without using simplifying assumptions which are commonly used in linear approaches" [CITE]. Furthermore, ANN has the advantage of being able to represent linear and non-linear relationships that can be extrapolated directly from the data. Though York and Ugulu are using qualitative data, their outcome of interest is still a categorical variable. In the context of this research paper, the output is categorizing a student's "attitudes towards nature" using Kellert's typologies. The research approach in this paper provides a conceptual framework to predict the outcomes of school district performance on Oklahoma A-F Report Card.

While Yorek and Ugulu provide a valuable conceptual framework for which to understand and apply ANN, Kash Barker et al provide an excellent example of how to apply ANN as well as Support Vector Machines to institutional decision making. The research conducting by Barker et al fall directly under the "Decision Support Systems" that Bakhshinategh et al describe in their assessment of application of EDM and LA. Barker et al use demographic and survey data collected from the University of Oklahoma and use ANN and SVM (support vector machines) to predict a student's probability of graduating on time [?]

In addition to providing a useful model for this research paper, Barker et al's sampling method provides additional support and ideas for approach district level data. Barker et al run ANN and SVM (each data set is split into training and testing sets) on a combined data set where all student cohorts are randomized and split in ANN and SVM models, is run again on training and test data sets that are split between years (i.e. training on the

1995 cohort and testing on the 1996 cohort) and finally testing and training within years. In developing a predictive model for Oklahoma schools and school districts, training and testing the data in a similar way could provide additional useful insights. This is especially true due to the fact that the data sources to be used collect information on school districts annually. In sum, Barker et al provide a precedent for using ANN and SVM to make institutional decisions. The authors are a prime example of how combine proven socio-economic indicators of academic success with state-of-the-art procedures and methodologies.

References