

One of the most important pieces of education policy to emerge over the past several years is how school performance is evaluated. This is often a source of contention in the education community as opponents claim that it causes harm to schools that need the most help, while proponents of school evaluations claim it acts as a motivational tool for teachers to perform better. In Oklahoma, the State Department of Education uses their own A-F report card to assess the progress of a school's impact on student learning. While proponents claim the methodology behind it shields it from bias, the opponents to the evaluation system claim its narrow focus drive does the opposite and fails to account for the hardships of minority and low income students.

The contentious nature of the evaluation framework lends credence to the idea that if it is possible to predict "failing" schools an efficiency in terms of resources and labor could be met in order to prevent that school from failing. As such, predicting the outcome of a school's performance becomes all the more important. Neural Networks and other machine learning techniques hold massive potential for predicting educational outcomes. Neural Networks in the context of artificial intelligence and machine learning hold promise because they capture non-linear relationships, which often times is not captured by classical program evaluation and regression framework. Since there are many complex and non-linear systems in education, neural networks can be a powerful tool in education and education policy.

In the following sections we discuss the collection of different data sources that pertain to characteristics of educational institutions throughout the state of Oklahoma, where that data is sourced from and basic summary statistics of the data that will become the inputs to the neural network predictive framework. Then, the initial results of neural network model are discussed and finally next steps are outlined.

1) Data:

a. Data Sources:

The data used in this analysis was collected primarily from the Oklahoma State Department of Education (OKSDE). The data consists of three sources: The A-F Report Card, The *Oklahoma Educational Indicators Program* reports as published by the Oklahoma State Department of Education's Office of Educational Quality and Accountability (OEQA) and published reports on OKSDE's "Transparency Index" which contains published data files for different educational topics. Each of the data source contains records for public schools in Oklahoma that is organized by a static identification code that is used to join these data sets to create a comprehensive data file.

For the following analysis, the data is restricted to only public schools in Oklahoma and excludes charter and private schools. At current count there are little more than 30 Charter Schools and Charter Schools primarily exist in only two of the over 500 different school districts throughout the state. Because Charter Schools do not always operate on the same parameters as traditional public schools, excluding them is an attempt to control for those unobservable or immeasurable qualities that separate them from traditional public schools. Furthermore, private schools are not included in OKSDE A-F Evaluations as well as report to another governing body separate from the State Department of Education. It should be noted that included in the analysis are organizations typically referred to as 'centers'. The Centers are additional organizations within a school district or school that are specific subsets of grades. For example, an ECE Center may be an early learning facility for young children inside of an elementary school. Regardless, Centers are evaluated on the same A-F report card as the other schools in state and district.

i. A-F Report Card:

The A-F Report Card was passed into law by the 2011 Oklahoma Legislature and used by the OKSDE in 2012. According to the 2013 documentation by the State Department of Education, the A-F Report Card “is designed to incentivize schools to strive for and reach high levels of college and career readiness”. The A-F report card is a evaluation tool that assess school quality in three main categories: overall student grade-level performance on academic standards, grade-level performance of ‘low perform students’ on academic standards, and school level (i.e. High School, Middle School, Elementary School) characteristics. The latter of these areas can be metrics like graduation rate and college entrance exam participation for high schools, or attendance for Middle or Elementary Schools.

Similar to traditional assessments, schools are given a score, referred to as the Index Score and a latter grade to judge the school’s performance on several different variables in each of the subareas as previously described. The models and methods applied to the analysis are primarily interested in the letter grade and the index score. The table below provides a selection of variables, the construct(s) the variable illustrates in the analysis, as well as the variable name used in the python code for the analysis

Variable	Construct	Variable Code
Index Score	Ranging from 0 to 100, is the sum of all points earned on the A-F Report card.	IndexScore
Letter Grade	A categorical representation of the Index Score which follows the typical grading structure (i.e. Index Score \geq 95, A+, Index Score between 89 and 85 is a B+)	LetterGrade
Full Code	A unique composite key used to identify a given school across multiple data sources	SchoolCode
Exam County	The number of students taking a given subject test (the Exam Count applies to math and reading for all grade 3 through 8, and science, history and writing from grades 5 and 8)	ELA_ExamCounty

<Subject> Performance Index	A 0 through 100 score representing the share of students passing the given subject tests	ELA_IndexScore
<Subject> Letter Grade	A categorical representation of the Performance Index representing the number of students passing the given grade level subject test	
Attendance Rate	The school's overall attendance rate	AttendanceRate
Attendance Rate Bonus	The bonus points awarded for different levels of attendance rate	
Graduation Rate	The proportion of graduated students to the number of enrolled students. This is an example of bonus points that are only applied to High Schools. Elementary Schools and Middle Schools are automatically credited these points to correct for differences	

For a full record of variables and constructs please see the tables in the appendix. The information is freely available to the public through the Oklahoma State Department of Education's Transparency Index. For complete source and citation please see the works cited list.

ii. Office of Educational Quality and Assessment

The Office of Educational Quality and Assessment is a reporting organization housed within the Oklahoma State Department of Education the publishes yearly reports that combine the previous year's subject grade level subject test along with other important socio-economic data from the American Community Survey, as well as variables that describe the workforce in a given school. It should be noted that this report is separate and has no impact on the A-F Report Card performance. Below are a selection of variables and their constructs sourced from the OEQA reports.

Variable	Construct	Variable Code
School Code	A unique composite key used to identify a given school across multiple data sources	SchoolCode
Average Income	The average income for the census tract based on the address of the school. Sourced from the American Community survey	AvgIncome

Poverty Rate	The proportion of households considered “in poverty” in the school’s census tract	PovertyRate
Average Teacher Salary	The average teacher salary for the given school. Each school typically follows the recommended salary schedule provided by OKSDE	AvgSalary
Average Years Experience	The Average number of years of experience each teacher has in each school	AvgExperience

The OEQA data can be acquired through two different channels. The first, more accessible channel, is through the pdf reports that are published by the OEQA on their website. The second channel is by requesting the flat file for a given year. The flat file contains the same information as each school’s pdf report just in a tabular format that is more readily accessible by computational tools.

iii. Oklahoma State Department of Education Transparency Index

The Oklahoma State Department of Education has published a “Transparency Index”, a repository of education related data that was used to collect demographic information about the enrolled students in Oklahoma public schools. As part of the ‘Student Count’ section of the transparency index, OKSDE has published several years of student demographics on the ‘Enrollment of Oklahoma State Public Schools’ section. The data is provided in a .csv format, wherein the document provides the count of students by demographic and gender for each school and grade. The table below is a selection of variables and fields collected from this data source:

Variable	Construct	Variable Code
School Code	A unique composite key used to identify a given school across multiple data sources	SchoolCode
Grade	The grade level for the given school	Grade
Hispanic Female	The count of Hispanic females in a grade for a given school in a given year	Hispanic_Female

Hispanic Male	The count of Hispanic males in a grade for a given school in a given year	Hispanic_male
Total	The total number of enrolled students for the given year	Total

b. Data Collection & Cleaning Procedures

The data described in the previous section was collected and loaded into a MySQL database maintained by the Oklahoma Public School Resource Center (along with several other data sources). To collect the necessary data, MySQL queries were written that aggregated these different data sources and written into csv files. The data sources were compiled into three different files based on their sources. In other words, the A-F data was queried and written in one file, the OEQA file in another and the demographic information and the final.

For the A-F Report Card data, the schools were separated into three distinct groups: Elementary, Middle and High School. The stratification was created in order to control for unobservable characteristics between groups. Furthermore, because each grade level and subject have their own variable, not splitting the schools by level would cause a large missing data problem, one that where imputing the missing data would make little sense if the school does not test for that grade in the first place. For example, a high school that serves grades nine through twelve would have a null value for the grade 3 reading performance index and since that high school does not serve that grade, it makes little sense to impute that mean. Another clarification for the A-F Report Card data set is that several variables are top coded at 95% and bottom coded at 5%. For example, attendance rates can be top-coded at 95% and AP Participation rate can be bottom coded at 5%.

The demographics data as sourced from the OKSDE Transparency Index was aggregated in MySQL as row wise percentages of the total number of students in each school. In other

words, the query total the number of each demographic for male and female, divided by the total number of students in the school (sum of Total column for each grade) for each year 2014 through 2016. The data was aggregated as a proportion in order to avoid issues of scale due to the wide range of total student population in schools across Oklahoma.

For several school districts with extremely small number of students, data was redacted from public record. This is a rare occurrence and due to the small number of schools where this occurred, those records were dropped from the analytical file.

c. Summary Statistics for Selection of Key Variables:

The table below outlines key statistics for selected variables:

i. OEQA Data Summary Statistics (Selected Variables)

	Average Income	Average Property Value /ADM	Unemployment Rate	Poverty Rate	% Free & Reduced Lunch
Min	25,047	0	0.00	0.0200	0.00
1 st Quartile	50,849	28,507	0.05	0.12	0.51
Median	56,810	41,924	0.07	0.17	0.67
Mean	59,681	52,973	0.069	0.17	0.65
3 rd Quartile	65,339	56,464	0.09	0.23	0.80
Max	219,858	824,541	0.28	0.56	1.42

	FTE Non-Special Education Teachers	Average Teacher Salary	% with Advanced Degrees	Average Years Experience
Min	0.00	0	0.0000	0.00
1 st Quartile	10.76	42,318	0.1520	10.55

Median	17.96	43,886	0.2300	12.75
Mean	21.13	44,161	0.2438	12.80
3 rd Quartile	26.99	45,751	0.3180	14.92
Max	171.87	64,204	1.0000	27.46

	FTE Special Education Teachers	FTE Counselors	FTE Other Staff	FTE Administrators
Min	0.0000	0.0000	0.00	0.000
1 st Quartile	0.9325	0.3200	0.43	0.930
Median	1.6900	0.6700	1.13	1.000
Mean	2.4788	0.8944	1.62	1.449
3 rd Quartile	3.0000	1.0000	2.25	2.000
Max	27.5200	12.4900	20.63	22.010

ii. A-F Data Summary Statistics (Selected Variables)

	Index Score	Reading Growth Index	Math Growth Index	Overall Growth
Min	2.00	0.00	0.00	0.00
1 st Quartile	68.00	65.00	61.00	62.00
Median	76.00	75.00	72.00	72.00
Mean	75.46	72.73	69.54	69.58
3 rd Quartile	85.00	83.00	81.00	80.00
Max	108.00	100.00	100.00	100.00

iii. Enrollment Data Summary Statistics

	% Hispanic	% Native American	% Asian	% African American	% Pacific Islander	% White	% Two or More Races	Total
Min	0.0000	0.0000	0.0000	0.0000	0.000000	0.0000	0.0000	1.0
1 st Quartile	0.0423	0.0449	0.0000	0.0065	0.000000	0.4064	0.0181	166.0
Median	0.0798	0.1198	0.0043	0.0236	0.000000	0.5510	0.0595	315.0
Mean	0.5633	0.2894	0.1230	0.3296	0.004298	1.2958	0.1943	380.3
3 rd Quartile	0.1478	0.2836	0.0140	0.0675	0.002000	0.6712	0.1084	499.0
Max	0.9748	0.991	0.2515	0.9767	0.3043	0.9652	0.748	:3778

2) Models & Methodology:

a. Methodology

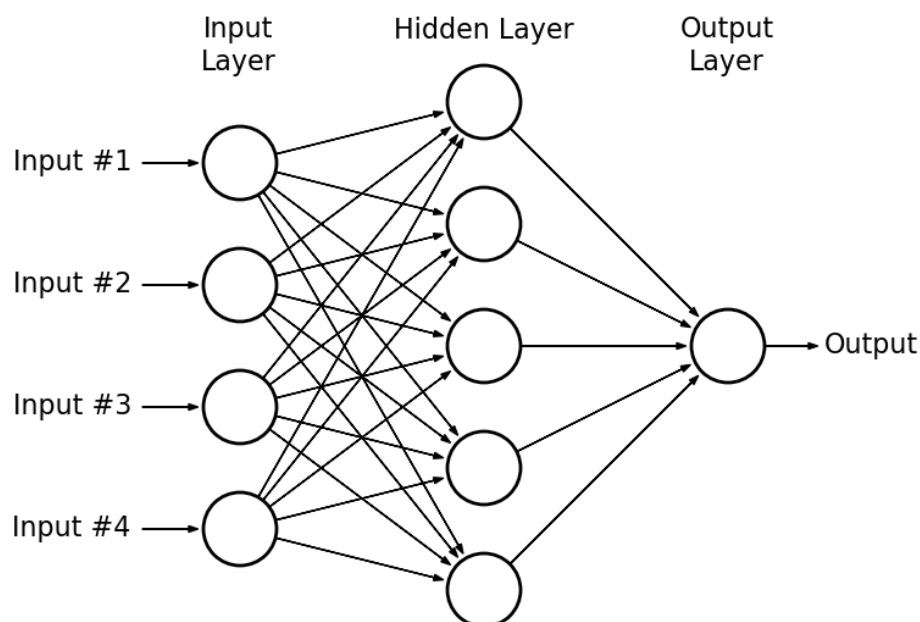
Barker et al describe a Neural Network as “ a data processing system consisting of a large number of simple, highly interconnected processing elements (artificial neurons) in an architecture inspired by the structure of the cerebral cortex of the brain” (cited from Tsoukalas and Uhrig (1997)).

At its core, the neural network is a handy way of capturing interactive effects of input variables. A neural network can provide a more reasonable means to predicting categorical outcomes in high dimension data sets. Depending on the which type of school we are predicting the outcome for (letter grade) we can have anywhere between 85 to 115 variables. In this case, especially with a high dimension data set, we cannot take the ‘kitchen sink’ regression approach. The kitchen sink approach is simply throwing all 85 to 115 different variables into a regression predicting some outcome, either categorical or continuous. This is approach is undesirable is it

leads to overfitting and bad predictive models but also is computational and time expensive. Furthermore, performing stepwise regression model selection is extremely time consuming. Therefore, neural networks provide a great advantage that not only compute through the complexity of model selection but also capture non-linear relationships between variables that step wise regression model selection does not always yield.

The diagram below provides an effective summary of what is happening in the neural network for our analysis. Using the middle school data as example, the first layer is simply the features in the data set. The output in our case would be the multiple letter grades possible. For simplification, we can call these ‘A’, ‘B’, ‘C’, ‘D’ and ‘F’, where each letter includes the plus and minus variations (except for those schools that are F that does not have the plus or minus variation).

Figure 1: *Diagram of Neural Network¹*



¹ Neural Network Diagram: http://www.astroml.org/book_figures/appendix/fig_neural_network.html

The hidden layer, as Barker et al describe, “consists of several predetermined hidden nodes connected to the input layer with a set of weights”. This hidden layer is again meant to capture and learn from possible non-linear relationships where each connection is weighted and then output a likelihood of being one of the five grades possible.

b. Methods:

Categorical Neural Network Model Specifications:

Using the Neural Network architecture, we then used a train, test and validation set to assess model performance. First, to create the network, we started with 1 Dense layer with 10 output units. Using this single dense layer we then produced a “softmax” max layer in which to produce the corresponding grade categories (A, B, C, D, and F). These layers used the ‘relu’ activation and the compiler method was optimized using ‘adam’, employed the categorical cross entropy loss function and produced the accuracy metric. This neural network model was trained using the training data set and the keras classifier that employed 200 epochs and a batch size of 15. In order to ensure that the model was free of bias, the analysis employed a kfold estimator. In the case of our initial analysis, the kfold estimator used 10 splits to validate the neural networks predictions and avoid bias.

Continuous Output Neural Network Model Specifications:

After implementing the (above) multinomial neural network, we implemented another neural network model for a continuous output which is referred to the Index Score in the data documentation. Similar to the multinomial approach, we split the middle school data set into test data, training data and validation data. However, we differed our approach slightly in that we

used sklearn's preprocessing library to scale the data. We decided to scale the data because of the range of values each column and averaging each column to mean zero would potentially give better results.

Like the categorical model we created a sequential neural network using Keras. The first layer was 13 input units, with a 'relu' activation. The loss function for the neural net was the mean squared error using the adam optimizer. The Keras Regressor used the aforementioned model with 100 epochs and batch size of 5.

3) Results:

a. Neural Network for Letter Grade

School Level	% Miss Classified	% Classified Correctly
Middle School	94.16%	5.84%
High School	83.91%	16.09%

The results from the categorical neural network show that our model performed extremely poor when it came to predicting the letter grade. The poor performance of the model (even with cross validation) could be due to many reason. One such reason is that the data that was used to build the neural network was not normalized (i.e. scaled) as to make all columns have mean zero. In doing so, we could hope to see a better performance for the categorical neural network.

b. Neural Network for Index Score

School Level	Mean Squared Error	Standard Deviation of MSE
Middle School	.049	.13

It is crucial to remember that in these results, they have been scaled to mean zero in order to account for wide range of values across the different columns in the data set. Our value of .044 seems promising that the neural network will perform better on scaled data sets, given that the mean squared error is relatively small. However, we can see that the standard deviation for the mean squared error of the predicted values in the test set is relatively large, suggesting increased variability in the predictions.

4) Next Steps:

To further developing model, we will produce a scaled data frame and re-run the categorical variables model as described in the previous sections. Furthermore, to create a model that has more predictive power for the categorical model, we will add more layers and 'deepen' the model to seek better predictive accuracy.

Works Cited:

1. K. Barker, T. Trafalis, and T. R. Rhoads. Learning from student data. In Proceedings of the 2004 IEEE Systems and Information Engineering Design Symposium, 2004., 2004.