SMU Data Science Review

Volume 2 | Number 2 Article 10

2019

Teacher Retention: Identifying Areas to Improve Teacher Recruitment and Retention in North Carolina Public Schools

Bin Yu
biny@smu.edu

Nuoya Rezsonya
Southern Methodist University, nrezsonya@smu.edu

Jacob Drew
Southern Methodist University, jdrew@smu.edu

Teacher Retention: Identifying Areas to Improve Teacher Recruitment and Retention in North Carolina Public Schools

Bin Yu, Nuoya Rezsonya, Jacob Drew

Master of Science in Data Science
Southern Methodist University
6425 Boaz Lane Dallas, Texas 75025
{byu,nrezsonya,jdrew}@smu.edu

Abstract. In this paper, we established a regression model to identify factors that significantly influenced both positively and negatively school-level yearly teacher attrition rate within North Carolina public schools using Belk Endowment Educational Attainment Data Repository for North Carolina Public Schools [1], district-level North Carolina IRS income tax data [2], and county-level North Carolina crime data [3]. Frequent teacher turnover has been negatively affecting teacher retention, student academic achievement, and educational attainment. School administrators are struggling to reduce teacher attrition rate since teacher retention is important to students' academic achievements. Our regression model utilized an eXtreme Gradient Boosting (XGBoost) Regression model to predict school-level teacher attrition rates using controllable school features that school administrators can affect, tax revenue features and public safety features. Out of 252 features reviewed, we found that school-level features impacted the most on predicting school-level yearly teacher attrition rate. The most important features were the school-level percentage of teachers with 0-3 years experiences, school-level short term suspensions per 100 students, school-level percentage of teachers that have reached proficient standard 1, school-level percentage of students who have economical disadvantages and school-level percentage of teachers who are licensed. Our findings provide opportunities for educational administrations to recognize key factors that can influence teacher attrition rates in North Carolina public schools.

1 Introduction

According to the Report to North Carolina General Assembly [4], there were 94,792 teachers employed in North Carolina between March 2016 and March 2017. Of these teachers, 8,249 were no longer employed in NC public schools (including those not teaching in public charter schools) by the end of 2016 to 2017 school year due to either career change or leaving for another state to teach. There are 5 categories of teacher attrition reasons: left the LEA but remained in education, left the LEA for personal reasons, terminated by the LEA, left the

LEA for reasons beyond the LEAs control, left the LEA for other reasons not listed above [5]. If teachers are satisfied with their current status, resignation due to a career change is less likely to happen. North Carolina educational administrators have tried to reduce the teacher attrition [6] in the past but it did not show any obvious improvement.

Finding the teacher attrition reasons, providing better solutions to improve teacher retention rate and teacher recruitment processes are always challenging tasks for North Carolina education administrations. In this paper, we tried to provide a framework to help understanding the reasons behind teachers' leaving from their public schools in North Carolina, we used North Carolina Public Schools data, local tax revenue data, and county crime rate to present a method to find features that could be controllable within school administrators' capacity to reduce school-level teacher attrition rate.

Teacher attrition is hurting North Carolina teacher retention, student academic achievement, and educational attainment. High rate of yearly attrition rate causes low teacher retention rate. As schools are funded mainly from local state by using tax revenue and bond measures, school funding is not evenly distributed meaning poorer areas receive less funding than wealthier areas. Teacher turnover has worse impacts on high-poverty schools because they already have fewer resources than other schools. When high yearly turnover occurs, students from high-poverty areas are more likely to have newly graduated teachers. Teacher attrition is highly disruptive for students' learning experience. Other than teaching, teachers play a very important role in students' character development together with their parents and guardians. If a teacher leaves in the middle of the school year, it damages the bond between students, parents and guardians, interrupts students' academic system, and takes more time for students to adapt into the study environment with a new teacher. This is also even worse in poorer areas which already have less resources to start with. When teacher attrition is frequent, it puts burden on school administrators who are with the goal of keeping the stable learning climate for students. It also limits administrators' capacity to fill vacant positions with teachers with similar qualifications. If leaving teachers have high qualifications and more experiences, it takes administrators more time and money to get the replacement.

In order to identify important features and understand how features are influencing school yearly teacher attrition, we established an XGBoost Regression model to reveal factors that contributed to teacher attrition rate. Through our findings, administrators can introduce new policies to improve the current attrition situations and improve the teacher recruitment processes.

To start the processing of building our regression model, we selected features that could be used in model. We examined all available features that we collected from Belk Endowment Educational Attainment Data Repository for North Carolina Public Schools [1], district-level North Carolina IRS income tax data [2] and county-level North Carolina crime data [3].Then we selected features through a mathematical process and a manual process. Selected features had three categories from our three different resources. Because not all factors

are controllable by school administrators, features like students demographic information are excluded from the analysis.

After selecting desirable features, We have utilized selected features to build six different regression models including eXtreme Gradient Boosting (XGBoost) Regression, Linear Regression, Lasso Regression, Ridge Regression, Elastic Net Regression, and Support Vector Regression (SVR) to predict school-level yearly teacher attrition rate at the school level in North Carolina. Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE) were used as the metrics to evaluate each model's prediction error. Among the six different models, eXtreme Gradient Boosting (XGBoost) Regression has outperformed the rest because this model enervated the least prediction errors.

XGBoost Regression was our top performing model because the model output indicates the expected attrition rate of new data could be 0.0517 points away from the actual attrition rate on average. XGBoost Regression is a regression algorithm that is infused with ensemble method. Using XGBoost, we have found the school-level percentage of teachers with 0-3 years experiences, school-level short term suspensions per 100 students, school-level percentage of teachers that have reached proficient standard 1, school-level percentage of students that have economic disadvantages and school level percentage of teachers are the top 5 important features to predict the attrition rate.

2 Data Preparation

North Carolina public school data set is originally from North Carolina Public Schools website [7] and was downloaded from Drew J., The Belk Endowment Educational Attainment Data Repository for North Carolina Public Schools, (2017), GitHub repository [1]. The data from North Carolina Public Schools website are highly comprehensive. It provides data from all aspects of schools including school performance, environment, funding and expenditure, teacher qualifications, class size, attrition rate and salary, student demographics, test scores, graduation rates, college enrollment and growth.

We collected North Carolina IRS income tax data at zip code level from the IRS website [2]. It included the selected income and tax items classified by State Zip Code and size of adjusted gross income. The download method is from Drew J., The Belk Endowment Educational Attainment Data Repository for North Carolina Public Schools, (2017), GitHub repository as well [1]. IRS data are based on individual income tax returns filed with the IRS are available for Tax years 2013 through 2016. We used the year 2014 to 2017 based on the North Carolina public school data set we selected. We had to align the tax year with school year properly when we did the analysis. And we needed to match the tax data LEA name with public school data LEA name as well.

North Carolina crime data at county level is from the Uniform Crime Reporting Program of the North Carolina State Bureau of Investigation [3]. North Carolina Uniform Crime Reporting Program is part of a nationwide, cooperative statistical effort administered by the North Carolina Federal Bureau of Investi-

gation. We downloaded the crime rate data from 2014 to 2017 for each county. It included the information like index rate, criminal homicide rate, forcible rape rate, robbery rate, aggravated assault rate, burglary rate, larceny rate, motor vehicle theft rate, and arson rate.

North Carolina teacher attrition rate by reason data set is from the yearly North Carolina state teacher professional report [4]. The data is in the appendix of the report which including the teacher turnover rate by reasons including left LEA but remain in education, personal reasons, initiated by LEA, beyond the control of LEA, and all other reasons. We used this data set to regroup the teacher attrition rate. North Carolina teacher attrition rate is the point of interest for this project. The definition of the measurement of the attrition rate is described in Section 3.2 above.

In order to consolidate data from 4 different sources, we have rearranged them to make sure that they are all at school campus level. Crime rate data set is at county level, we merge it back to the public school data on county column of the school data. This process has made sure that all of the school in the same county would have the same crime rate for this county. Income tax data is at zip code level, we have founded out the zip code for each school campus and then have joined them together on zip code. Teacher attrition reason data set has joined with public-school data set by LEA name.

There are also data cleaning processing of sparse fields. There are features being removed due to the fact that they have either more than 60% of NA value or more than 25% of unique value. Imputation with zero has also been performed on features that have less than 60% of NA values. Because numeric columns have different scales and they need to be at the same scale, we have standardized the features to bring them to the same scale.

After joining district IRS income tax data and county crime rate data to the public schools data for 2017, the initial data set for this project has been established. It has a total of 407 features from all features at school level across all public- schools ranging from elementary to high schools along with district IRS information and county crime data. Appendix Table 7. Data Dictionary provides a detailed data dictionary of the data that has been used in this project.

3 Factors Related to North Carolina Teacher Attrition

3.1 Teacher Compensation and Benefits

All school expenditures including teacher compensation and benefits come from school funding. Schools have the right to decided how to distribute the funding. School funding has four resources: state, federal, and local tax revenue and bond measures. Based on the federal role in school funding, federal contributions to elementary and secondary education is around 8%. This means that the other 92% comes from either state or local tax revenue. State and local funding are primary resources for K-12 education [8]. Education is primarily a state and local responsibility in the United States. School funding is not equally distributed to

schools either [9]. Since funding is allocated based on the needs, circumstances, and grade level of each student [10] [11] [12], wealthier districts have more funding available for their public schools.

Historically, teacher compensation and benefits have not kept up with inflation. During 2016-2017, the nationwide average annual salary for teachers was $58,950, which is only 1% higher than the prior year [13]. Between 2000 and 2017, annual salary increases for teachers have not even kept up with inflation based on salary percentage changes. The nationwide average annual teacher's salary has decreased by $974 (1.6%). In North Carolina, this decline is much more significant, with a $6,643 (11.8%) decrease during the same period while accounting for inflation [13]. In North Carolina this difference is much more significant than nationwide when considering teacher salaries between 1999-2000 and 2016-2017 school years.

Table 1. Estimated average annual salary of teachers in public elementary and secondary schools 1969-70 through 2016-17

	Current Dollars							Constant 2016-2017 Dollars			
State	1969 -70	1979 -80	1989 -90	1999 -2000	2009 -10	2015 -16	2016 -17	1969 -70	1979 -80	1989 -90	1999 -2000
1	2	3	4	5	6	7	8	9	10	11	12
United States	$8,626	$15,970	$31,367	$41,807	$55,202	$58,353	$58,950	$55,411	$49,917	$59,944	$59,924
North Carolina	$7,494	$14,117	$27,883	$39,404	$46,850	$47,941	$49,837	$48,139	$44,125	$53,286	$56,480

State	2015 -10	2009 -16	2016 -17	Pct Change
	13	14	15	16
United States	$61,804	$59,426	$58,950	-1.6
North Carolina	52,453	48,823	49,837	-11.8

3.1.1 LEA Salary Expenditure

In North Carolina, a school district is referred to as a local education agency (LEA). The LEA salary expenditure comes from the result of LEA total expense amount multiplied by LEA salary expense percentage, which these two categories of data have both been collected from North Carolina Public Schools website. Salary expenditure at the LEA level indicates the amount of dollars spent on salary only. Figure 1. shows no evidence of obvious linear relationship between these two features. However, the points are highly concentrated with salary expenditure between $4,200 and $6,100 while the one-year attrition rate between 5% and 25%. In the meantime. there are also scattered points located in the high attrition rates zone, as in more than 35% of attrition rate. Moreover, we can tell that as the LEA salary expenditure increases, the points become sparser. This tells us the greater LEA expenditure on salary, the less likely the teachers are to leave.

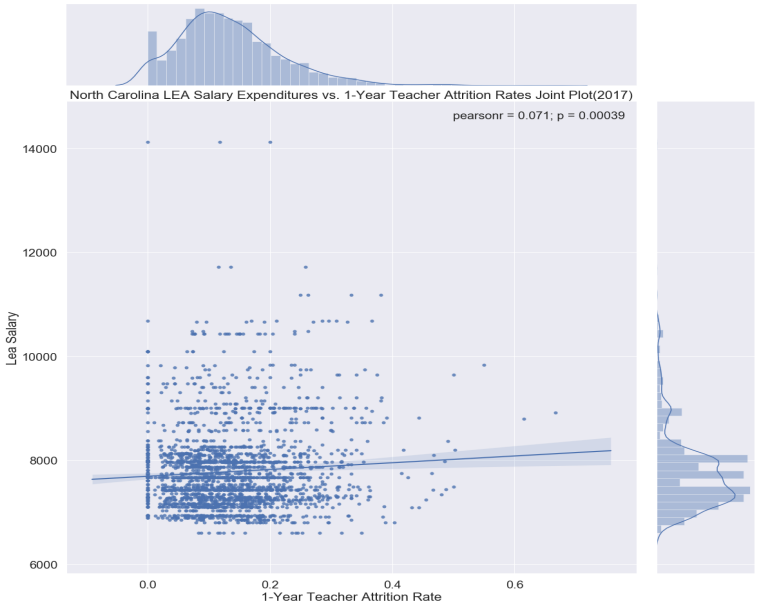

Fig. 1. North Carolina Public School 1-Year Teacher Attrition Rates vs. LEA Salary Expenditure Pct (2017). The Pearson correlation is -0.21 which indicates two features have negative relationship. This relationship is significant since the p-value is way less than 0.05.

3.1.2 LEA Total School Spend

The LEA total school expenditure comes from the LEA total expense amount that have collected from North Carolina public schools. Figure 2 shows the relationship between LEA total school expenditure and the one-year turnover rate of teachers. The relationship seems similar to the relationship between LEA salary expenditure and the one-year turnover rate of teachers.

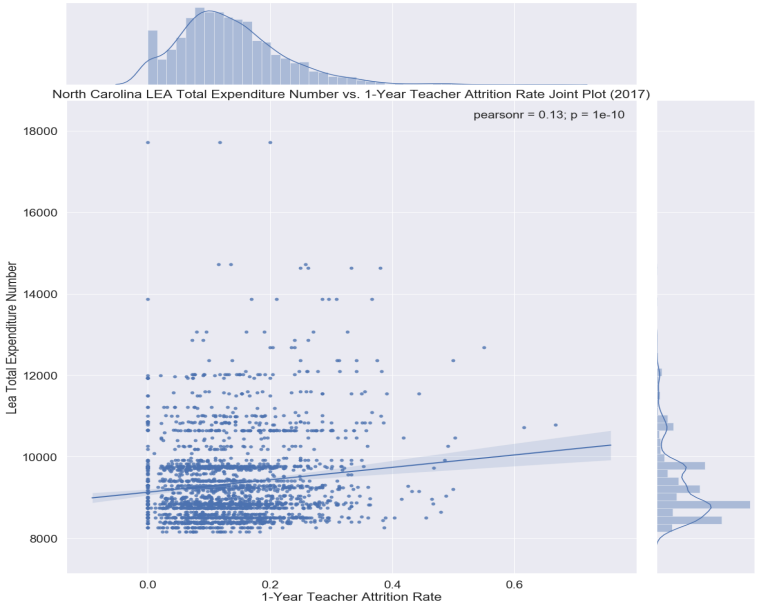

Fig. 2. NC Public School 1-Year Teacher Attrition Rates vs. LEA Total School Expenditure (2017). The Pearson correlation is 0.13 which indicates two features have positive relationship. This relationship is significant since the p-value is way less than 0.05.

Figure 2 shows no evidence of obvious linear relationship between those two and all points form a wider fan shape. Points are more highly concentrated with total expenditure between $8,000 and $10,000 while the one-year attrition rate between 0.05% and 0.22%. There are less points located in the high attrition rates zone, as in more than 0.35% of attrition rate, and they are in the total expenditure range between $8,000 and $12,000. We can still tell that as the total spend increases, attrition is less likely to happen.

3.2 Organization Location and Working Environment

Researches show that school socioeconomic location and environment affects teachers and their working lives [14]. Improving the school environment can

affect teacher-student relationships, increasing teacher job satisfaction and productivity. The geographic location of the school has a direct relationship with teacher attrition.

Darling-Hammond (2004) states that 40% to 50% of teachers in high poverty schools left their school within the first 5 years [15]. Several reasons for the teachers leaving the high poverty schools are frustrations associated with the working conditions, dealing with more difficult students, poor administrative leadership, a lack of collaborations, and an inadequate amount of discipline. Teachers in high poverty territory schools would be more likely to face big challenges and experience stress in their career.

3.3 District Income Tax

As we have discussed, education is primarily a State and local responsibility in the United States. It is States and communities, as well as public and private organizations of all kinds, that establish schools and colleges, develop curricula, and determine requirements for enrollment and graduation [8]. Federal funding could add up to an additional $5,652 per elementary student, depending on program eligibility.

Based on the nature of public-school funding, local state then has the major role of distributing the funding among schools. For example, during the 2017-18 school year, small, low-wealth school systems received $12,551 in state funds for each special-needs elementary school student with limited English proficiency from a low-income family [10]. State of North Carolina spent around 40% of the tax revenue on the K-12 education every year since 2000. A higher tax revenue district has the advantage to spend more money on the teachers salary and improve the school environment which in return help teachers to stay. We have investigated the relationship between district income tax and teacher attrition. Figure 3 shows that there is a trend of attrition rate decreases while Number of farm returns 100K-200K increases.

3.4 Public Safety

Both district crime rates and the crimes committed against teachers could also affect teacher attrition. A high district crime rate would affect teachers lives outside of school, while crimes against teachers would affect the teachers daily activities and productivity inside school. A safe and peaceful environment is greatly needed so that teachers can concentrate on teaching and are willing to dedicate their work to school activities.

A NCES report regarding the school crime and safety during 2017 [16] indicates that students are not the only victims of intimidation or violence in schools. Teachers are also subjected to threats and physical attacks, which are sometime committed by students. During the 2011-2012 school year, 9% of school teachers reported being threatened with injury by a student from their own school. The percentage of elementary teachers who reported being physically attacked by a

Fig. 3. NC Public School 1-Year Teacher Attrition Rates vs. Number of farm returns 100K-200K (2017). The Pearson correlation is -0.1 which indicates two features have negative relationship. This relationship is significant since the p-value is way less than 0.05.

student (8%) was higher than the percentage of secondary teachers (3%). In addition, more teachers reported being threatened with injury or being physically attacked in public-schools (10%) than in private schools (3%).

3.5 County Crime Rate

In 2017 calendar year, North Carolina state has an index crime rate at 3,061.5 per 100,000 people [3], with a decrease of 3.2% from 2016, property crime rate at 2,677.8 per 100,000 people with a 3.9% decrease compared with 2016, however the violent crime increased 2.3% at 383.7 per 100,000 people. Whereas teacher attrition rate in school year of 2017-2018 was 8.1% which decreased 0.6% from 2015-2016 school year (8.7%) (Table 2) . We suspect there is a positive correlation between district crime rate and teacher attrition.

Table 2. Statewide Crime Index Rate per 100,000 Persons Percent Change from 2016 to 2017

	Index Crime Rate	Violent Crime Rate	Property Crime Rate	Violent crime Rates				Property Crime Rates			
				Murder	Rape	Robbery	Agg. Assault	Burglary	Larceny	MVT	Arson
2016 Rate	3,161.8	375.0	2,786.8	7.0	21.2	95.8	251.1	729.1	1,904.9	152.8	14.8
2017 Rate	3,061.5	383.7	2,677.8	6.5	20.6	95.8	260.8	673.5	1,843.7	160.7	15.1
Percent Change	-3.2%	+2.3%	-3.9%	-7.5%	-2.6%	-0.0%	+3.9%	-7.6%	-3.2%	+5.2%	+2.0%

4 Feature Selection and School-level Yearly Teacher Attrition Rate

4.1 Feature Selection

A two-step feature selection method applied, and this method involved both mathematical process and manual process. Mathematical process involved Voting Scheme Feature Selection technique while the manual one was to validate the coverage of the feature selected Vote Scheme Feature Selection.

There was a total of 407 features after we finished the data cleaning. However, there were features that was beyond the education organization's control. For example, a school cannot decide and control how much of funding it receives at times because it is funded by the tax revenue. A school also cannot control student demographics because it cannot decide on who go to schools and who do not. We believe feature selection should be coming from statistical perspective and also from practical knowledge. If an administration cannot control those factors, we believe the need for keeping them are low. That is why the manual feature selection process is included to eliminate non-controllable features. We manually removed 11 features which related with the races of the public-school students from the data set.

Voting Scheme Feature Selection [17] comes from a combination of 8 algorithms including Pearson, Linear Regression, Ridge Regression, Lasso Regression, Elastic Net, Recursive Feature Elimination, Random Forest (RF), Correlation, and Maximal Information Correlation (MIC), meaning that each of the algorithms would be used to select features independently and then return a score (ranking or coefficient, see Table 3). Then scale (MinMaxScaler) the scores from each algorithm. In the end, we got the mean of the scores for each feature (Mean column). We used the (Mean - STD/2) as the threshold to eliminate the features. In this step, we were able to eliminate 141 features from the final model.

Table 3. Voting Scheme Feature Selection Results

Features	Linear reg	Ridge	Lasso	ElasticNet	RFE	RF	Corr.	MIC	Mean
tchyrs_0thru3_pct	0.76	1.0	1.0	1.0	0.91	1.0	1.0	1.0	0.96
lateral_teach_pct	0.76	0.39	0.7	0.63	0.85	0.07	0.74	0.63	0.6
Proficient_TCHR_Standard 1_Pct	0.76	0.53	0.52	0.54	0.86	0.11	0.48	0.49	0.54
tchyrs_11plus_pct	0.76	0.77	0.3	0.31	0.86	0.06	0.45	0.52	0.5
flicensed_teach_pct	0.76	0.23	0.33	0.22	0.85	0.19	0.66	0.69	0.49
Developing_TCHR_Standard 4_Pct	0.76	0.48	0.6	0.6	0.61	0.02	0.32	0.31	0.46
Crime_Rate_Murder_Rate	0.76	0.63	0.55	0.63	0.77	0.01	0.15	0.14	0.46
EOG/EOCSubjects_CACR_AIG	0.76	0.69	0.38	0.5	0.8	0.02	0.08	0.3	0.44
lea_services_expense_pct	0.72	0.55	0.41	0.43	1.0	0.03	0.18	0.17	0.44
Proficient_TCHR_Standard 2_Pct	0.76	0.45	0.33	0.31	0.89	0.03	0.33	0.41	0.44

After completing this two-step methods of feature selection, we arrived at a total of 252 features to feed into the modeling step. This data set of final features has been then used as the independent variables in the prediction of the teacher attrition rate. Table 3 provides the top 10 selected features through feature selection process. There are 6 out of 10 features that fall into the school level teacher personnel category. Three of these features are within teacher qualification category including percentage of teachers with experience 0 to three years, percentage of teachers with experience more than 11 years and the percentage of licensed teachers. There are two features coming from the development category which are percentage of teachers who reached Proficient Standard 1 and Developing Standard 4. State average student number, the percentage of other reasons for attrition and murder rate form up the rest of the top 10 features Table 3. has illustrated the relationship between percentage of teachers with experience 0 to three years with 1-year teacher attrition rate. Although there is no obvious linear relationship, it has a fan shape opening toward the right top corner which is very obvious that higher percentage of teachers with 0-3 years experiences return higher attrition rate. It is actually a negative influence trend on attrition rate.

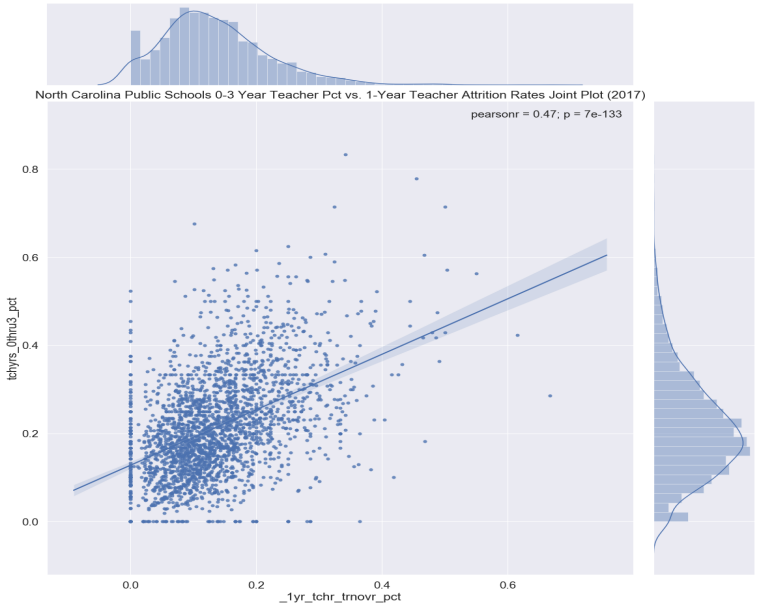

Fig. 4. North Carolina 0-3 Year Teacher Pct vs. 1-Year Teacher Attrition Rates Joint Plot (2017). The Pearson correlation is 0.47 which indicates two features have positive relationship. This relationship is significant since the p-value is way less than 0.05.

4.2 School-level Yearly Teacher Attrition Rate

The teacher attrition rates are defined as the total number of teachers leaving the educational services divided by the total number of teachers in North Carolina public K-12 schools. The teacher retention rates can be calculated as one minus the teacher attrition rate. Teacher attrition rates information can be retrieved from the Personnel file in North Carolina Public Schools data sets. It should be noted that the reasons leading to teacher attrition can be categorized into five groups (see Appendix Table 6.).

The North Carolina General Assembly Report for Teaching Professionals provided a detailed breakdown on each category as: teachers who left the Local Education Agency (LEA) but remained in education, teachers who left for personal reasons, teachers who were terminated by the LEA, teachers who left for reasons beyond LEA control, and teachers who left for other reasons[4].

School-level yearly teacher attrition rate was the numeric variable we were to predict in this project. All the final features we selected were used in the modeling process to predict this turnover rate. Based on the data sources and information guide from North Carolina Public Schools website, all classroom teachers employed in a school during April of a year but not employed as a classroom teacher in the same school system during April of the following year were included in the school's turnover statistics [18]. Figure 6 shows the distribution of the North Carolina teacher attrition rate by year 2017. Attrition

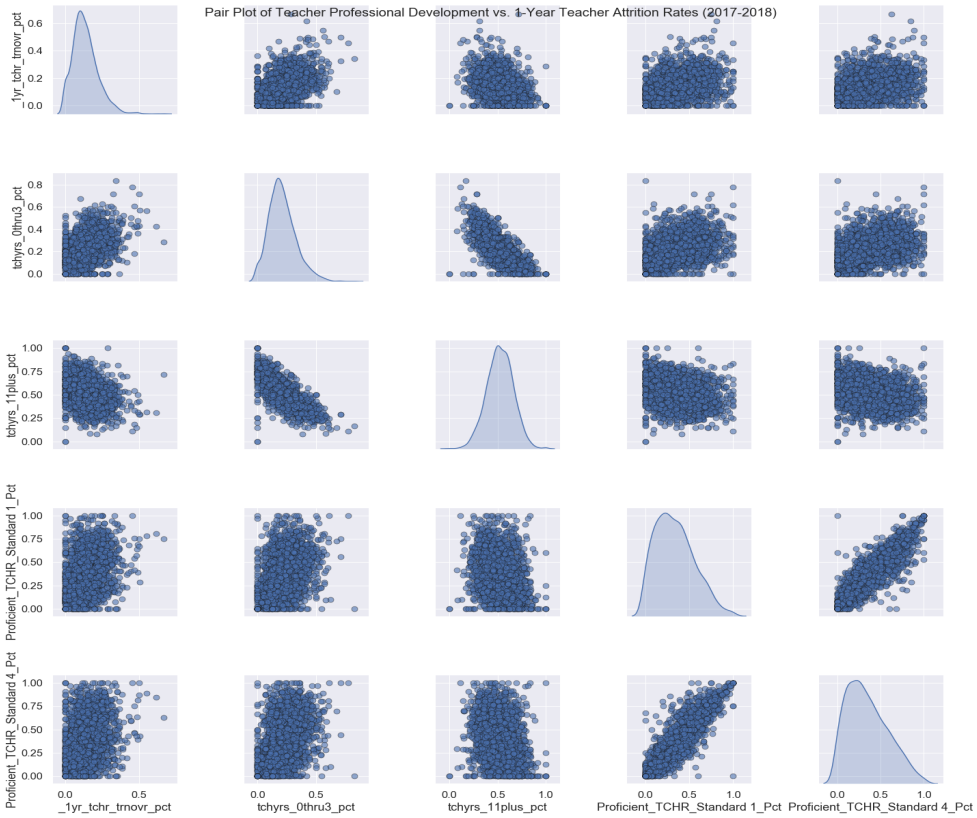

Fig. 5. North Carolina Teacher Professional Development vs. 1-Year Teacher Attrition Rates Pair Plot (2017). The pair plot indicates the relationship between these professional development features. Some of them have positive relationships and some of them have negative relationships.

rates were concentrated at the bottom part which most points are in the range between 0.1 to 0.2. Points are very sparse in the top area.

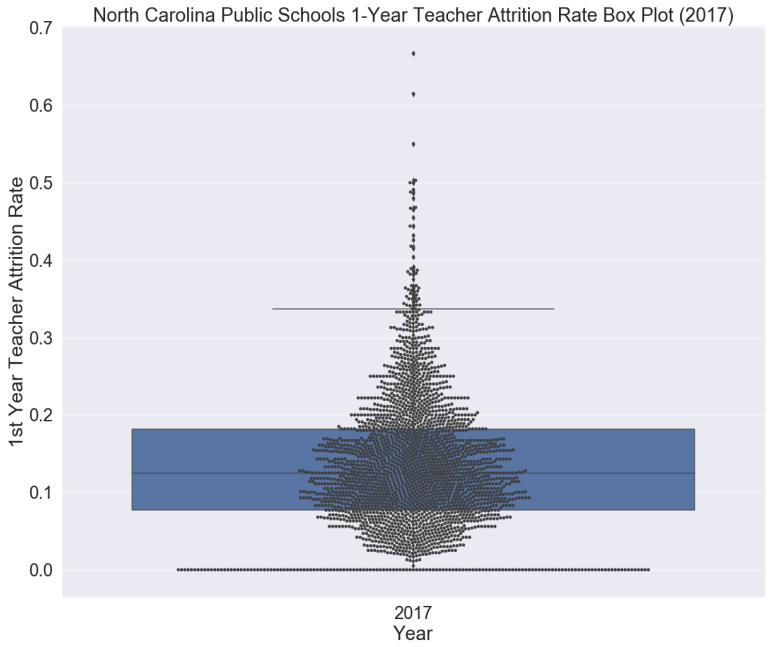

Fig. 6. North Carolina Public-School One-Year Teacher Attrition Rate Distribution (2017). The box plot shows the distribution of the 1-year teacher attrition rate of North Carolina in 2016-2017 school year.

5 Model Construction

We also established 6 different regression models by using regressive algorithms including XGBoost Regression, Lasso Regression, Elastic Net, Ridge Regression, Support Vector Regression, Multiple Linear Regression to give us wider choice range so that we can go with the one algorithm gives the best regression performance with the lowest Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE). XGBoost Regression has outperformed the rest with the lowest values of MAE, MSE and RMSE. Ten-fold validation has been used across all 6 regression models with shuffle split of the whole data so that 80% of the whole data goes into training set and the rest is

the test set. Grid search technique has all been performed across all models to give the best sets of tuned hyper-parameters so that each model is at their best performance to predict the response variables.

5.1 Prediction Performance Evaluation Metrics

In order to qualify how well a regression model performs in predicting the response variable, we included three evaluation metrics to evaluate the model we constructed. They are Mean Absolute Error, Mean Squared Error, Root Mean Square Error. We included R Squared as a part of the evaluation results, but we were not using R Squared as a reference metric for the performance evaluation because R Squared is commonly considered as a biased estimator to evaluate model performance. Mean Absolute Error, Mean Squared Error and Root Mean Square Error all have value range from zero to infinity and were all preferred lower values.

Mean Absolute Error (MAE) is the average of the absolute difference between the predicted values and real value we have. It is a linear score meaning that all the individual differences are weighted equally in the average. For example, the difference between 10 and 0 would be twice the difference between 5 and 0 [19]. The mean squared error is similar to MAE, it is taking the average of the squared difference between the predicted value to the real value. This distance is considered as the prediction error. The squaring process can remove negative values if any. Unlike MAE, all the individual differences are not equally weighted due to the squaring step meaning the larger differences have larger weight. Root Mean Square Error (RMSE) represents the sample standard deviation of the differences between the predicted values and the real values. This standard deviation is the prediction errors(residuals) and measures how spread out the prediction errors are [20]. Both MAE and RMSE represents the average prediction errors. Due to the nature of the calculation, RMSE penalizes bigger differences more than MAE which lead RMSE to have bigger or equal value as MAE.

5.2 XGBoost Regression Algorithm

Before we introduce the XGBoost Regression, let us start with a brief explanation of Decision Tree and Ensemble methods. Decision Tree is supervised learning algorithm which makes quantitative or qualitative predictions based on various conditions or questions. Due to the nature of how decisions are made at each split, it eventually forms a tree-like procedure. Here comes the most important two questions. The first question is that how we make decisions on features. The second is that if at one point a yes or no answer is needed to go to the next step, what is the threshold for getting a yes or no answer?

If we rely on only one single decision tree, we are going to assume we make the correct decision at each split which is a big assumption to make. That is why we need the Ensemble Methods. It is a machine learning technique that combines multiple models to produce a stronger regression model. For example,

instead of just using a single decision tree and assume that a correct decision has been made at each split, Ensemble can help on making a group of Decision Trees and get the features being used at each split and then a final Decision Tree is created based this several Decision Trees.

The goal of Ensemble methods is always to improve the model performance by decreasing the variance and bias in terms of minimizing the difference between the predictions and the real values. There are three main kinds of Ensemble techniques including: Bagging, Boosting and Stacking. Boosting replies on the patterns of difference between predictions and real values. Later models would be specifically focusing on those features that are hard to fit the data from the previous models and then get them right. At the end, a combination of multiple models is done by giving different weights to each single model. The method XGBoost Regression we use in this project is from the Boosting family.

Instead of using just one single Decision Tree, we used multiple trees with Boosting. For example, there are two features in the data, one was the school performance indication SPG score and the other one was the total expenditure at LEA level. SPG score is a categorical feature which has letters A, B, C, D, E, F indicating 6 different levels of school performances. LEA total expenditure is numerical which we could make them into categories as in high, median and low. Assume we are going to use these two variables to predict the one-year attrition rate. Among the multiple trees, you can imagine that one tree is using the SPG score splitting into 6 nodes and in each node splitting again into 3 nodes of the LEA total expenditure categories. The second tree may use LEA total expenditure first to split into 3 nodes then each node split into the 6 nodes of the SPG score. These two trees may have different model performance or similar. With different features and different orders to feed the features into the tree, different trees are made. They may have very different prediction performances or very similar prediction performances. At the end, final model is constructed by giving different weights to each model.

XGBoost stands for eXtreme Gradient Boosting. It is a gradient boosted Decision Tree designed for computing speed and model performance. As we mentioned above, the goal of XGBoost is not different from the rest of Ensemble family. It is always to improve the model performance by minimizing the difference between predictions and actual values. Comparing to other methods under the Boosting family, XGBoost gives more of a regularized model formalization to control the overfitting hence gives a more accurate prediction [21]. Always providing an outstanding predictive accuracy is the most favorite advantage that XGBoost offers. In the meantime, it provides flexibility on loss functions and options of tuning parameters. It also has the ability to handle missing data which other regression algorithms has limitation on this.

6 Results and Interpretation

6.1 XGBoost Model Performance

Based on the performance evaluation from all 6 models, it was very clear that XGBoost Regression outperformed the rest of models we constructed. Table 4. has demonstrated the evaluation outputs from models. You can see that while other models had very good performance, XGBoost Regression was still the best one with the lowest MAE, MSE, and RMSE.

Table 4. Model Cross-Validation Results

Model	MAE	MSE	RMSE
XGBoost	0.0517	0.0043	0.0658
Lasso Regression	0.0537	0.0051	0.0713
Ridge Regression	0.0546	0.0053	0.0722
ElasticNet Regression	0.0549	0.0052	0.0722
Linear Regression	0.0554	0.0054	0.0734
Support Vector Regression	0.0564	0.0055	0.0739

Table 4. demonstrated the performance of each model we constructed. The outperformed XGBoost Regression model had a Mean Absolute Error value of 0.0517. This was a low MAE value which indicates the expected attrition rate based new data could be 0.0517 points away from the actual attrition rate on average. It is not possible to create a model that has 100% correct prediction capability but with a very low value of MAE, MSE and RMSE XGBoost Regression has managed to give us confidence to say that XGBoost Regression gives us prediction outcomes that are very close to the reality. Through comparing the predicted values and actual values against each other, we can see how well the model is predicting. Figure 7. shows the relationship between the predicted values and the actual values. You can see that points are scattering around the reference line in the middle. While the actual value goes up, the predicted value follows the same trend. There are zero values being predicted with non-zero values and very few outliers on the right top corner. Most of plotted points are highly close to each other.

6.2 Important Features Selected in XGBoost Regression Model

Another benefit that XGBoost Regression provides is feature importance which gives us a very straightforward visibility of how important of each attribute. A score is used to represent the level of usefulness a feature in the process of constructing the trees within the model. A higher score of a feature means this feature has been used more to make key decisions with the decision trees. Importance scores are calculated for each individual attribute so that rankings

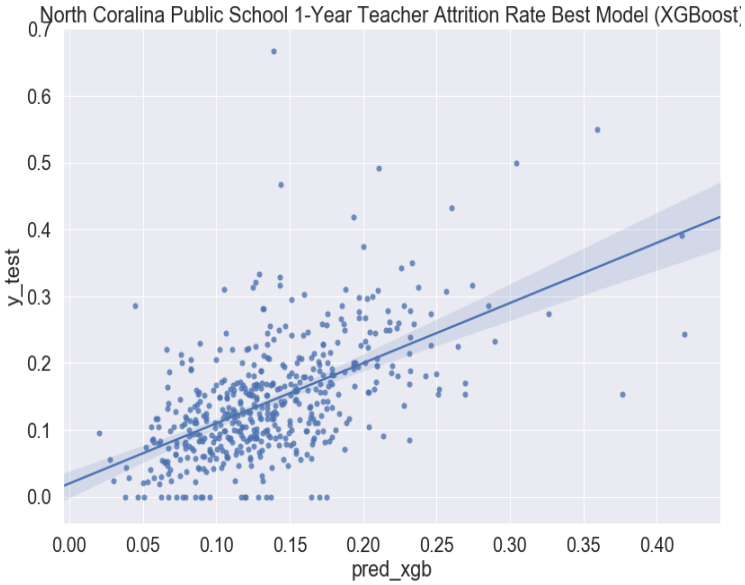

Fig. 7. North Carolina Public-School 1-Year Teacher Attrition Rates Best Model from XGBoost Regression. It shows the relationship between the predicted values and the actual values. You can see that points are scattering around their reference line in the middle. While the actual value goes up, the predicted value follows the same trend.

are sorted in descending order and visualizations of feature importance can be created. Importance was calculated for a single decision tree by the amount that each attribute split point improves the performance measure, weighted by the number of observations the node is responsible for. Then all feature importance is averaged across all the decision trees within the model.

After feeding into 252 features into XGBoost Regression, it selected 51 features that were significant valuable during the processing of establishing decision trees within the model. Table 5 displayed the top 10 important features XGBoost Regression selected. The top 5 features were school-level percentage of the teachers with 0-3 years experiences, school-level short term suspensions per 100 students, school-level percentage of teachers that have reached proficient standard 1, school-level percentage of students that have economic disadvantages, school-level percentage of teachers that have been licensed. Within these top 5 features, three features that was related with teacher qualification and experience level, while the rest of these 5 features were related to student discipline and student financial background. The school-level percentage of the teachers with 0-3 years experiences appears again as the most important feature in constructing XGBoost Regression model. It was also listed as the most important feature during the feature selection process. The rest of these top 5 important features were also in the top features in the feature selection process through Voting Scheme Feature Selection. All top 5 features have importance more than 100, scores fallen below 100 after the sixth feature. Figure 8. shows the top 50 important features sorting by scores.

Table 5. XGBoost Regression Feature Importance (top 10)

Feature	Score
tchyrs_0thru3_pct	251
short_susp_per_c_num	159
Proficient_TCHR_Standard 1_Pct	149
pct_eds	113
flicensed_teach_pct	101
lea_sat_avg_score_num	98
Attrition_Reason_OtherReasonsPerct	97
lea_services_expense_pct	96
Avg Personal Property Taxes	93
student_num	93

Figure 4. shows 1-year teacher attrition rates and the school level percentage of the teachers with 0-3 years experiences. We already found that higher percentage of teachers with 0-3 years negatively influence the attrition rate. When comparing attrition rate against teacher with 11 years plus experiences, we found the opposite. Figure 9. shows the relationship between attrition rate and the school level percentage of the teachers with 11 years plus experiences. Having

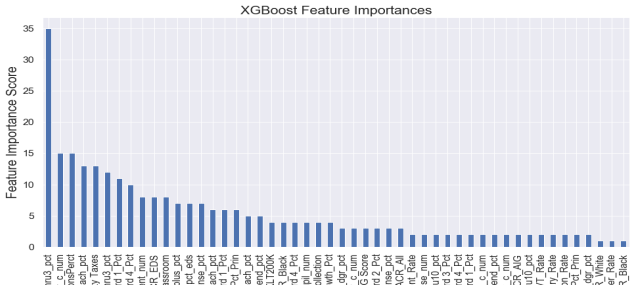

Fig. 8. North Carolina Public-School Teacher Attrition XGBoost Regression Model Feature Importance. It shows top 50 important features sorting by scores.

higher percentage of highly experienced teachers has the trend of lowering the one-year attrition rate.

Through the model selected features, percentage of students who have economical disadvantages at school level is one of the important features. Figure 10. shows the relationship between percentage of students who have economical disadvantages at school level and the attrition rate. It has a fan shape opening towards the lower right corner. It shows that with higher attrition rate, the percentage of student who have economical disadvantages also concentrates at the right top corner indicating higher percentage of students having economical disadvantages.

7 Ethics

According to the definition in Oxford Dictionary, ethics is moral principles that govern a person's behavior or the conducting of an activity [22]. In working with educational data, it is imperative that moral principles should be strictly followed in the research process that provide guidance on what can or cannot be done. There are three perspectives on ethics that all of the educational researcher should be aware of [23]. First, originated from Aristotle's notion of virtues, ethics are proposed as the appropriate manner in terms of how a person possesses justice, generosity, and honesty. In our data analysis, we strove to treat data with integrity, honesty, and fairness. Particular attention was made to ensure data was not accidentally fabricated or manipulated. Second, ethics focus on the duty humans have to act and perform in ways that show respect to other human beings. In analyzing our data, we included income tax data to provide a comprehensive picture on the factors that could predict teacher attrition rate in North Carolina. Last but not the least, ethics relates to guiding principles for ethical conduct that should benefit a majority of people from the utilitarian perspective. Our data analysis was done in the hope of benefiting all stakeholders of school districts, including, students, teachers, administrators, and more impor-

Fig. 9. North Carolina Public-School 11 Years Teacher Pct vs. 1-Year Teacher Attrition Rates Joint Plot (2017). The Pearson correlation is -0.33 which indicates two features have negative relationship. This relationship is significant since the p-value is way less than 0.05.

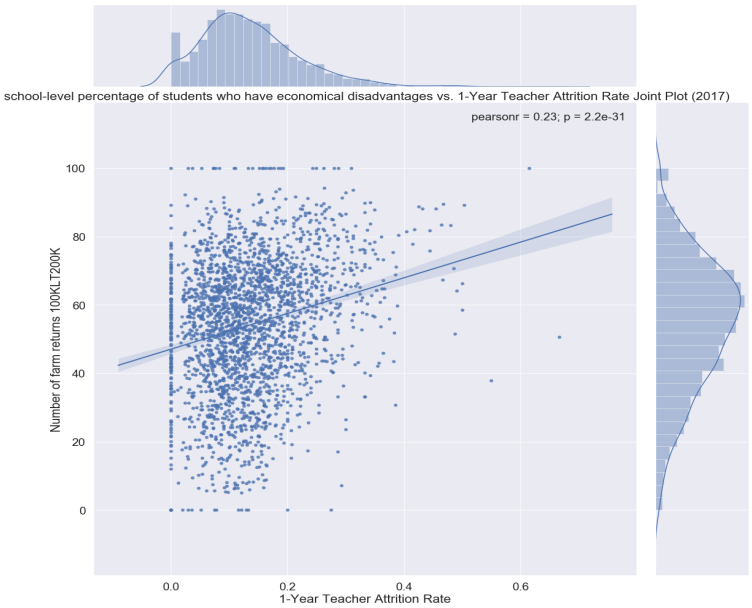

Fig. 10. School-level Percentage of Students Who have Economical Disadvantages vs. 1-Year Teacher Attrition Rates Joint Plot (2017). It shows that higher attrition rate occurs when the percentage of student who have economical disadvantages increases.

tantly, the local community. All of these three perspectives provided guidance for us to follow during our research process.

Overall, the target of this analysis was to identify factors that are significantly related to improving teacher retention rate in North Carolina. Our finding and suggestions would directly affect the educational attainment, student academic achievement, and teachers' daily activities in the state of North Carolina. When we collected the data, we carefully chose the most appropriate data sets, strove to make the right suggestions for the feature selection, and be honest when we interpret the results.

Both schools and families are important to education. As the factors that could possibly affect students' future achievements and educational attainment, teachers are equally influential to students' well beings and academic achievements. A high rate of teacher attrition could hurt the educational attainment, student academic achievement, and district budgets. Therefore, administrations need a stable establishment of teacher resources and local education organizations need to make well-informed decisions to ensure that students learn with high-quality teachers.

8 Conclusions

Among the teacher attrition factors we analyzed in this paper, school funding, compensation level, student demographics were not easily controllable by the school district. Thus, such factors were removed from the models in our analysis. Although we have only utilized the school level data, tax data, and crime data, we believe our findings could shed lights on the important factors related to teachers' attrition. Our analysis found school-level features predicted school-level yearly teacher attrition rate. The most important features included the school-level percentage of teachers with 0-3 years experiences, school-level short term suspensions per 100 students, school-level percentage of teachers that have reached proficient standard 1, school-level percentage of students who have economical disadvantages and school-level percentage of teachers who are licensed. Our findings provide opportunities for educational administrations to recognize key factors that could improve teacher attrition rates in North Carolina public schools. For example, administrations could consider hiring more seasoned teachers, especially the ones with Standard 1 category. Also, they could consider working on school to help the students who have economical disadvantages.

Since we only used school-level aggregated data, this analysis was limited by the extent to which data source could provide to our analysis. Future work could utilize more detailed data, such as grade-level data, student-level data, and more importantly, the teacher-level data. In that case, we could dig deeper into the factors at teachers' personal level so that we can identify the factors that significantly impact teacher attrition rate. Additionally, this analysis established a framework that could be replicated at another geographical location and similar data sources can be used to address the similar problems in other states.

References

1. J. Drew, "The belk endowment educational attainment data repository for north carolina public schools." https://github.com/jakemdrew/EducationDataNC, 2018.
2. IRS, "Soi tax stats - individual income tax statistics - zip code data (soi)." https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi, 2017.
3. N. C. S. B. of Investigation, "Uniform crime reporting program." http://crimereporting.ncsbi.gov/Introduction.aspx, 2017.
4. N. D. O. P. INSTRUCTION, "Report to the north carolina general assembly, 2016-2017 state of the teaching profession in north carolina general statute 115c-12(22)." http://www.ncpublicschools.org/docs/district-humanresources/surveys/leaving/2016-17-state-teaching-profession.pdf, 2018.
5. J. Stanford, "Teacher turnover in north carolina." https://demography.cpc.unc.edu/2019/01/24/teacher-turnover-in-north-carolina/, 2019.
6. M. Osborne, "Districts address teacher turnover by providing affordable teacher housing." https://nccppr.org/districts-address-teacher-turnover-providing-affordable-teacher-housing/, 2018.

7. P. S. of North Caroline, "Public schools of north caroline." http://www.ncpublicschools.org/, 2018.
8. U. D. of Education, "The federal role in education." https://www2.ed.gov/about/overview/fed/role.html, 2017.
9. D. of School Business North Carolina Department of Public Instruction, "Highlights of the north carolina public school budget february 2018." http://www.dpi.state.nc.us/docs/fbs/resources/data/highlights/2018highlights.pdf, 2018.
10. L. SCHLEMMER and J. DEBRUYN, "10 questions to understand school funding in nc." https://www.wunc.org/post/10-questions-understand-school-funding-nc, 2018.
11. N. C. C. for County Research, "Basics of county financing for public schools." http://www.ncacc.org/DocumentCenter/View/2852/Basics-of-County-Financing-of-Public-Schools?bidId=, 2018.
12. T. Stoops, "Public school finance." https://www.johnlocke.org/policy-position/education-spending/, 2018.
13. N. C. for Education Statistics, "Table 211.60.: Estimated average annual salary of teachers in public elementary and secondary schools." https://nces.ed.gov/programs/digest/d17/tables/dt17_211.60.asp, 2017.
14. Q. Gu and C. Day, "Challenges to teacher resilience: Conditions count," *British Educational Research Journal*, vol. 39, no. 1, pp. 22–24, 2014.
15. T. Marston, "Factors that contribute to teacher retention in high-poverty middle schools," *Electronic Theses and Dissertations*, p. 2374, 2014.
16. I. N. C. for Education Statistics, "Indicators of school crime and safety." https://nces.ed.gov/pubs2017/2017064.pdf, 2016.
17. A. Saabas, "Selecting good features part iv: stability selection, rfe and everything side by side." http://blog.datadive.net/selecting-good-features-part-iv-stability-selection-rfe-and-everything-side-by-side/, 2014.
18. N. C. S. B. of Education, "Nc school report card: Personnel." http://www.ncpublicschools.org/src/guide/personnel/, 2016.
19. A. Swalin, "Choosing the right metric for evaluating machine learning models." https://medium.com/usf-msds/choosing-the-right-metric-for-machine-learning-models-part-1-a99d7d7414e4, 2018.
20. S. Glen, "Rmse: Root mean square error." https://www.statisticshowto.datasciencecentral.com/rmse/, 2016.
21. L. Labs, "What is the difference between the r gbm (gradient boosting machine) and xgboost (extreme gradient boosting)." https://www.quora.com/What-is-the-difference-between-the-R-gbm-gradient-boosting-machine-and-xgboost-extreme-gradient-boosting, 2016.
22. Lexico, "Ethics." https://www.lexico.com/en/definition/ethics, 2019.
23. L. Shawa, *Ethics in educational research*, p. 433. ResearchGate, 01 2017.

Appendix

Table 6. North Carolina Self-Reported Reasons for Leaving

Code	Description
1) Teachers who left the LEA but remained in education (Includes individuals resigning to teach in another NC LEA or charter school and individuals who moved to non-teaching positions in education)	
58	**Moved to a non-teaching position in education in another LEA or Agency** Teachers moved to counselor, media coordinator, or non-teaching duties in another LEA or Agency Teachers moved to administrative positions (school-based) in another LEA or Agency Teachers moved to supervisory, director, or coordinator positions in another LEA or Agency Teachers accepted non-teaching support or administrative positions in another LEA or Agency
59	**Moved to a non-teaching position in education in another LEA or Agency** Teachers moved to counselor, media coordinator, or non-teaching duties in another LEA or Agency Teachers moved to administrative positions (school-based) in another LEA or Agency Teachers moved to supervisory, director, or coordinator positions in another LEA or Agency Teachers accepted non-teaching support or administrative positions in another LEA or Agency
70	**Resigned to teach in a NC charter school** Teachers leaving LEA to accept a teaching position in a NC Charter School Teachers obtaining another teaching job on their own initiative (as opposed to spouse relocation)
71	**Resigned to teach in a NC non-public/private school** Teachers leaving LEA to accept a teaching position in a NC non-public/private school Teachers obtaining another teaching job on their own initiative (as opposed to spouse relocation)
75	**Moved to a non-teaching position in the LEA** Teachers moved to counselor, media coordinator, or non-teaching duties in current LEA of employment Teachers moved to administrative positions (school-based) in current LEA of employment Teachers moved to supervisory, director, or coordinator positions in current LEA of employment Teachers accepted non-teaching support or administrative positions in current LEA of employment
2) Teachers who left for personal reasons (Includes individuals retiring with reduced benefits, individuals resigning to teach in a non-public school in NC, individuals resigning to teach in another state, individuals dissatisfied with teaching, change) non-teaching positions in education)	
57	**Resigned Family responsibility/Child care** Teachers resigning for maternity/family leave Teachers resigning to care for ill parents or members of the immediate family Teachers resigning to care for family business or personal needs
60	**Resigned To continue education/Take a sabbatical** Teachers resigning to return to school Teachers resigning to pursue an educational leave of absence
61	**Resigned Family relocation** Teachers resigning due to spouses relocation Teachers resigning as a result of marriage and relocation Teachers resigning due to family relocation
62	**Resigned To teach in another state** Teachers leaving NC to teach in a public school in another state Teachers leaving NC to teach in a private school in another state

63	**Resigned Dissatisfied with teaching** Teachers resigning due to dissatisfaction with teaching
64	**Retired with reduced benefits** Teachers resigning due to personal disability or health related issues
68	**Retired with reduced benefits** Teachers retiring after age 50 with reduced benefits Teachers retiring with less than full benefits
72	**Resigned Career Change** Teachers resigning to pursue another employment opportunity Teachers resigning to pursue interests outside teaching
73	**Re-employed Retired Teacher Resigned** Teacher who had retired, was re-employed and subsequently resigns
3)Teachers whose departure was initiated by the LEA (Includes individuals who were non-renewed, dismissed, or resigned in lieu of dismissal)	
50	**Dismissed** Teachers demoted or dismissed under GS 115C-325(h) Probationary teachers dismissed during the school year under GS 115C-325(m) Teachers dismissed under GS 115C-325 (Below standard ratings) Teachers reported to the dismissed teacher list Teachers dismissed and the ruling upheld by case manager
53	**Non-Renewed Probationary Contract Ended** Probationary teachers whose contract is not renewed after the end of the year
54	**Interim Contract Not Rehired (Report only for interim contracts of 6 months or more** Interim teachers not rehired under retirement cap Teachers not rehired under a term contract with specific employment dates Teachers not rehired due to return of a permanent teacher from a leave of absence
55	**Resigned in lieu of dismissal** Teachers resigned to avoid placement on dismissed teacher list Teachers resigned rather than go through full dismissal hearing Teachers resigned during an active investigation regarding performance/behavior as a professional educator
56	**Did not obtain or maintain license** Teachers not renewed due to failure to fulfill lateral entry requirements Teachers not renewed due to failure to earn 15 renewal credits Teachers failed to meet Praxis or provisional license requirements Teachers let license expire Teachers license was revoked
78	**Resigned In Lieu of Non-Renewal**
4)Teachers who left for reasons beyond LEA control (Includes individuals who retired with full benefits, individuals who resigned for health reasons, individuals who resigned due to family responsibilities and/or childcare, death, and individuals who resigned due to family relocation)	
51	**Reduction in Force** Teachers not rehired due to loss of enrollment, funding, or programming Teachers covered under local RIF policies

66	**Retired with full benefits** Teachers age 60 with 25 years of creditable service Teachers with 30 years of creditable service Teachers age 65 with at least 5 years of creditable service Teachers retiring with fullunreduced retirement benefits
67	**Deceased** Teachers who die while in active service in a NC public school
74 eligible to be employed in North Carolina	**Resigned End of Visiting International Faculty (VIF) Term** Teachers whose cultural visas have expired and are no longer
76	**Resigned Moving Due to Military Orders** Teachers resigning due to being moved under military orders
77	**Resigned End of Teach for America (TFA) Term**
5)Teachers who left for other reasons (Includes teachers resigning or leaving teaching for reasons not listed or those who resigned for unknown and other reasons)	
65	**Resigned Other reasons** Teachers resigning or leaving teaching for reasons not listed on the survey
69	**Resigned Reasons unknown**

Table 7. North Carolina Teacher Attrition Data Analysis Data Dictionary (Public School)

Feature Name	Description
Unit_code	Code to identify School LEA State
Student_num	Number of students at school level *schoolsize*
Student_num	Number of students at school level *schoolsize*
School_Name	School Name
Lea_Name	LEA Name
State_Name	State Name
Category_cd	Category code E, M, H, I, T, or A
Street_ad	Street address
Scity_ad	City name
State_ad	State
Szip_ad	ZIP code
Vphone_ad	Phone Number
Type_cd	Charter or public school
Closed_ind	1=School is closed, 0=School is currently open
New_ind	1=School is new this year, 0=School is not new *If a school is new this year, no data is available
Super_nm	Superintendents Name
Url_ad	URL Address of the school or district
Grade_range_cd	Range of grades offered
Calendar_type_txt	Description of school calendar and school type
School_type_txt	Description of school type
Calendar_only_txt	Description of Calendar
Title1_type_cd	Title I Status
Esea_status	ESEA status
Grad_project_status	Required Graduation Project, Y=Yes or N=No
Student_num	Number of students at school level (school size)
Lea_avg_student_num	Average school size within the LEA
St_avg_student_num	Average school size within the State
Welcome_ltr_url	URL for LEA/Charter school welcome letter
Stem	STEM (Science, Technology, Engineering, Math) status
County	County the school is physically located in
CLP_IND	Consistently Low Performing indicator (used on SRC Snapshot Only)
Focus_CLP_IND	Focus school designation due to CLP status (used on SRC Snapshot Only)
Summer_program_ind	Summer school program was included in Read to Achieve results
ASM_No_SPG_IND	Indicator for no SPG due to alternative school status
No_data_spg_ind	Indicator for no SPG due to lack of data
Level	Grade Level (K-8) or Course (English II, Math I, Biology)
Size	Average class/course size for school
Lea_size	Average class/course size for district
total_expense_num	Total expense (Dollars Spent) at school level
Lea_ total_expense_num	Total expense (Dollars Spent) at LEA level
St_ total_expense_num	Total expense (Dollars Spent) at state level
Salary_expense_pct	Percent of expense spent on Salaries at school level

Table 8. North Carolina Teacher Attrition Data Analysis Data Dictionary (Public School)

Lea_salary_expense_pct	Percent of expense spent on Salaries at LEA level
st_ salary_expense_pct	Percent of expense spent on Salaries at state level
Benefits_expense_pct	Percent of expense spent on Benefits at school level
Lea_ benefits_expense_pct	Percent of expense spent on Benefits at LEA level
St_ benefits_expense_pct	Percent of expense spent on Benefits at state level
Services_expense_pct	Percent of expense spent on Services at school level
Lea_ Services_expense_pct	Percent of expense spent on Services at LEA level
St_ Services_expense_pct	Percent of expense spent on Services at state level
Supplies_expense_pct	Percent of expense spent on Supplies at school level
Lea_supplies_expense_pct	Percent of expense spent on Supplies at LEA level
St_supplies_expense_pct	Percent of expense spent on Supplies at state level
Instruct_equip_exp_pct	Percent of expense spent on Instructional Equipment at school level
Lea_ instruct_equip_exp_pct	Percent of expense spent on Instructional Equipment at LEA level
St_ instruct_equip_exp_pct	Percent of expense spent on Instructional Equipment at state level
Feature Name	Description
other_expense_pct	Percent of expense spent on Other Expenses at school level
Lea_ other_expense_pct	Percent of expense spent on Other Expenses at LEA level
St_ other_expense_pct	Percent of expense spent on Other Expenses at state level
Lea_ building_expense_pct	For future use (not available for 2013-14)
St_ building_expense_pct	For future use (not available for 2013-14)
federal_perpupil_num	Federal expense per pupil at school level
Lea_ federal_perpupil_num	Federal expense per pupil at LEA level
St_ federal_perpupil_num	Federal expense per pupil at state level
Local_perpupil_num	Local expense per pupil at school level
Lea_ Local_perpupil_num	Local expense per pupil at LEA level
St_ local_perpupil_num	Local expense per pupil at state level
state_perpupil_num	State expense per pupil at school level
Lea_ state_perpupil_num	State expense per pupil at LEA level
St_ state_perpupil_num	State expense per pupil at state level
Passed_bog_and_or_eog	Students who passed 3rd grade BOG/EOG. Valid values areyes or no
Exemption	LEP, EC, Already Retained, Passed RTA Test, Portfolio, etc.
Outcome	Promoted 4th grade, Retained reading, or retained (3rd grade)
Num_students	Number of students (by outcome) at school level
Lea_num_students	Number of students (by outcome) at LEA level
St_num_students	Number of students (by outcome) at state level
Pct_students	Percentage of students (by outcome) at school level
Lea_pct_students	Percentage of students (by outcome) at LEA level
St_pct_students	Percentage of students (by outcome) at state level
Sat_avg_score_num	Average SAT Score (Critical Reading plus Math) at the School Level
Lea_ sat_avg_score_num	Average SAT Score (Critical Reading + Math) at the LEA Level
St_ sat_avg_score_num	Average SAT Score (Critical Reading + Math) at the State Level

Table 9. North Carolina Teacher Attrition Data Analysis Data Dictionary (Public School)

Nat_ sat_avg_score_num	Average SAT Score (Critical Reading + Math) at the National Level
Sat_participation_pct	Percentage of High School Seniors taking the SAT at the School Level
lea_ sat_participation_pct	Percentage of High School Seniors taking the SAT at the LEA Level
st_ sat_participation_pct	Percentage of High School Seniors taking the SAT at the State Level
nat_ sat_participation_pct	Percentage of High School Seniors taking the SAT at the National Level
Esea_attendance	Valid values: met, did not meet
Lea_esea_attendance	Valid values: met, did not meet
ap_participation_pct	Percentage of High School Students taking an AP exam at the School Level
Lea_ ap_participation_pct	Percentage of High School Students taking an AP exam at the LEA Level
St_ ap_participation_pct	Percentage of High School Students taking an AP exam at the State Level
ap_pct_3_or_above	Percentage of AP Exams with Scores of 3 or Above at the School Level
Lea_ ap_pct_3_or_above	Percentage of AP Exams with Scores of 3 or Above at the LEA Level
St_ ap_pct_3_or_above	Percentage of AP Exams with Scores of 3 or Above at the State Level
ib_participation_pct	Percentage of High School Students taking an IB Exam at the School Level
Lea_ ib_participation_pct	Percentage of High School Students taking an IB Exam at the LEA Level
Prin_advance_dgr_pct	Percent of principals with advanced degrees at LEA level
St_ Prin_advance_dgr_pct	Percent of principals with advanced degrees at state level
_1yr_prin_trnovr_pct	One year principal turnover rate at LEA level
St_1yr_prin_trnovr_pct	One year principal turnover rate at state level
Prin_male_pct	Percent of male principals at LEA level
St_ Prin_male_pct	Percent of male principals at state level
Prin_female_pct	Percent of female principals at LEA level
St_prin_female_pct	Percent of female principals at state level
Prin_black_pct	Percent of black principals at LEA level
St_ Prin_black_pct	Percent of black principals at state level
Prin_white_pct	Percent of white principals at LEA level
St_ Prin_white_pct	Percent of white principals at state level
Prin_other_pct	Percent of principals of other races including Indian, Asian,Hispanic and Pacific-islander at LEA level
St_ Prin_other_pct	Percent of principals of other races including Indian,Asian,Hispanic and Pacific-islander at state level
Flicensed_teach_pct	Percent of teachers that meet NC fully licensed definition at school level
Lea_ flicensed_teach_pct	Average Percent of Teachers that meet NC fully licensed definition at LEA level
St_ flicensed_teach_pct	Average Percent of Teachers that meet NC fully licensed definition at state level
Class_teach_num	Average number of classroom teachers at School Level
Lea_class_teach_num	Average number of classroom teachers at LEA Level
St_class_teach_num	Average number of classroom teachers at State Level
nbpts_num	Number of National Board Certified Staff at school level
Lea_ nbpts_num	Average number of National Board Certified staff at LEA level
St_ nbpts_num	Average number of National Board Certified staff at state level
advance_dgr_pct	Percent of teachers with masters or higher degree at school level
Lea_ advance_dgr_pct	Average percent of teachers with masters or higher degree at LEA level
St_ advance_dgr_pct	Average percent of teachers with masters or higher degree at state level
_1yr_tchr_trnovr_pct	One Year Teacher turnover percentage at school level
Lea_1yr_tchr_trnovr_pct	One Year Teacher turnover percentage at LEA level
St_1yr_tchr_trnovr_pct	One Year Teacher turnover percentage at state level
Lateral_teach_pct	lateral entry teacher percentage at school level

Table 10. North Carolina Teacher Attrition Data Analysis Data Dictionary

Lea_lateral_teach_pct	Average lateral entry teacher percentage at LEA level
St_lateral_teach_pct	Average lateral entry teacher percentage at state level
Highqual_class_pct	Percent of highly qualified classes at school level
Lea_highqual_class_pct	Average highly qualified classes at LEA level
St_highqual_class_pct	Average highly qualified classes at state level
Lea_highqual_class_hp_pct	Percent of classes taught by highly qualified teachers in high poverty schools at LEA level
St_ highqual_class_hp_pct	Percent of classes taught by highly qualified teachers in high poverty schools at state level
Lea_highqual_class_lp_pct	Percent of classes taught by highly qualified teachers in low poverty schools at LEA level
Lea_highqual_class_all_pct	Percent of classes taught by highly qualified teachers in all schools at LEA level
St_ highqual_class_all_pct	Percent of classes taught by highly qualified teachers in all schools at state level
Lea_not_highqual_class_hp_pct	Percent of classes taught by not highly qualified teachers in high poverty schools at LEA level
St_ not_highqual_class_hp_pct	Percent of classes taught by not highly qualified teachers in high poverty schools at LEA level
Lea_not_highqual_class_lp_pct	Percent of classes taught by not highly qualified teachers in low poverty schools at LEA level
St_not_highqual_class_lp_pct	Percent of classes taught by not highly qualified teachers in low poverty schools at state level
Lea_not_highqual_class_all_pct	Percent of classes taught by not highly qualified teachers in all schools at LEA level
St_ not_highqual_class_all_pct	Percent of classes taught by not highly qualified teachers in all schools at state level
Total_class_tch_num	Number of classroom teachers
Total_nbpts_num	Number of National Board Certified teachers
St_ highqual_class_lp_pct	Percent of classes taught by highly qualified teachers in low

Table 11. North Carolina Teacher Attrition Data Analysis Data Dictionary (District Income Tax)

(TCE) prepared returns Ct 100KLT200K
(TCE) prepared returns Ct 25KLT50K
(TCE) prepared returns Ct 75KLT100K
(VITA) prepared returns Ct 100KLT200K
(VITA) prepared returns Ct 25KLT50K
(VITA) prepared returns Ct 50KLT75K
Add child tax credit Amt 25KLT50K
Add child tax credit Amt 75KLT100K
Advance premium tax credit Amt GE200K
Alternative minimum tax Amt 25KLT50K
Alternative minimum tax Amt 50KLT75K
Alternative minimum tax Amt 75KLT100K
Alternative minimum tax Ct 50KLT75K
Alternative minimum tax Ct 75KLT100K
Deductible points Amt LT25K
Deductible points Ct LT25K
Domestic production activities deduction Amt 100KLT200K
Domestic production activities deduction Amt 25KLT50K
Domestic production activities deduction Amt 50KLT75K
Domestic production activities deduction Amt 75KLT100K
Domestic production activities deduction Amt LT25K
Domestic production activities deduction Ct 25KLT50K
Domestic production activities deduction Ct 50KLT75K
Domestic production activities deduction Ct 75KLT100K
Domestic production activities deduction Ct LT25K
Earned income credit Amt 50KLT75K
Excess earned income credit (refundable) Amt 50KLT75K
Health care individual responsibility payment Amt GE200K
Home mortgage interest paid from personal seller Amt LT25K
Investment interest paid Amt 25KLT50K
Investment interest paid Amt LT25K
Investment interest paid Ct LT25K

Table 12. North Carolina Teacher Attrition Data Analysis Data Dictionary (District Income Tax)

Net investment income tax Amt 100KLT200K
Number of farm returns 100KLT200K
Number of farm returns 25KLT50K
Number of farm returns 50KLT75K
Number of farm returns 75KLT100K
Number of farm returns All
Number of farm returns GE200K
Number of farm returns LT25K
Number of refund anticipation loan returns 50KLT75K
Number of volunteer prepared returns Tot 50KLT75K
Other non-limited misc deduction Amt 50KLT75K
Other non-limited misc deduction Amt 75KLT100K
Other non-limited misc deduction Amt LT25K
Other non-limited misc deduction Ct 75KLT100K
Other non-limited misc deduction Ct LT25K
Partnership/S-corp net income (less loss) Amt LT25K
Residential energy tax credit Amt LT25K
Residential energy tax credit Ct LT25K
Self-employed (Keogh) retirement plans Amt LT25K
Self-employed (Keogh) retirement plans Ct LT25K
Tot Add Medicare tax Amt 25KLT50K
Tuition and fees deduction Amt 75KLT100K

Table 13. North Carolina Teacher Attrition Data Analysis Data Dictionary (County Crime Rate)

Crime_Rate_Index_Rate	County Level Crime Index Rate
Crime_Rate_Violent_Rate	County Level Crime ViolentRate
Crime_Rate_Murder_Rate	County Level Crime Murder Rate
Crime_Rate_Rape_Rate	County Level Crime Rape Rate
Crime_Rate_Robbery_Rate	County Level Crime Robbery Rate
Crime_Rate_Burglary_Rate	County Level Crime Burglary Rate
Crime_Rate_MVT_Rate	County Level Crime MVT Rate
Crime_Rate_Arson_Rate	County Level Crime Arson Rate