12-2016

Using machine learning to predict student achievement on the state of Texas Assessment of Academic Readiness examination in charter schools

Christopher D. Gonzalez
The University of Texas Rio Grande Valley

USING MACHINE LEARNING TO PREDICT STUDENT ACHIEVEMENT ON THE STATE OF TEXAS ASSESSMENT OF ACADEMIC READINESS EXAMINATION IN CHARTER SCHOOLS

A Thesis

by

CHRISTOPHER D. GONZALEZ

Submitted to the Graduate College of
The University of Texas Rio Grande Valley
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2016

Major Subject: Computer Science

USING MACHINE LEARNING TO PREDICT STUDENT ACHIEVEMENT ON THE STATE OF TEXAS ASSESSMENT OF ACADEMIC READINESS EXAMINATION IN CHARTER SCHOOLS

A Thesis
by
CHRISTOPHER D. GONZALEZ

COMMITTEE MEMBERS

Dr. Emmett Tomai
Chair of Committee

Dr. Richard Fowler
Committee Member

Dr. Wendy Lawrence-Fowler
Committee Member

Dr. Andres Figueroa Lozano
Committee Member

December 2016

ABSTRACT

Gonzalez, Christopher D., Using Machine Learning to Predict Student Achievement on the State of Texas Assessment of Academic Readiness examination in Charter Schools. Master of Science (MS), December, 2016, 34 pp., 6 tables, 10 figures, references, 15 titles.

The purpose of this study was to research and develop a way to use machine learning algorithms (MLAs) to predict student achievement on the State of Texas Assessment of Academic Readiness (STAAR), specifically in the charter school setting. Charter schools have the disadvantage of a constant influx in students, so providing historical student data in order to analyze trends proves difficult. This study expands on previous research done on students in secondary and post-secondary school and determining features that indicate success in these settings. The data used is from the district of IDEA Public Schools who focuses on providing education to low income and minority populations. This study uses data that was readily available to IDEA Public Schools and MLAs provided by MATLAB to create models in order to predict if a student is going to meet the standard on the STAAR test at the end of the year.

DEDICATION

The completion of the master's program would not have been possible without the encouragement and never ending support of my parents Danny Gonzalez and Rosie Escamilla. I would like to give a special thank you to Lien Nguyen for being there for every high point and every low. I would not have made it this far without her. Thank you for your love and strength when mine had faded.

ACKNOWLEDGMENTS

I could not give enough thanks to Dr. Emmett Tomai, chair of my thesis committee, for all of his support not only during my graduate career but my undergraduate as well. He believed in me and gave me opportunities that have changed my life and am eternally grateful. My thanks also go to my thesis committee members for taking time out of their busy schedule for guidance and input.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Chapter I is structured as follows: an introduction to the field of study; an introduction to the State of Texas Assessments of Academic Readiness (STAAR) examination; a brief review of the current literature; the statement of the problem; purpose of the study. Chapter II will provide an expanded review of literature which includes predictor variable selection and early warning indicators for student success in post-secondary education, low income and demographic impact on student achievement, and machine learning algorithm comparison. Chapter III will detail the data used in this study, the procedure in developing models, and model finalization for the subjects of math, reading, writing, science, and social studies. Chapter IV will explain the results when applying the finalized models in each subject, a summary of the study and future work section.

Introduction to the Field of Study

There have been numerous studies conducted to determine if a student is ready for college. This may be because nearly 39% of college students feel like there was a gap in what they learned in high school and what is expected from them in college or work place (Peter D. Hart Research Associates/Public Opinion Strategies, 2005) or because the US had an overall decline in college graduation rates standing when compared to other countries (OECD, 2010).

Though the studies involving college readiness are plentiful, there seems to be a gap in primary and secondary school research. This study will focus its attention of the subsection of public schools known as charter schools. Charter schools have seen an increase in enrollment and an increase in serving the Hispanic population. From 2003 to 2013 there had been in jump from 1.6 percent of public school students attending charter schools to 5.1 percent (U.S. Department of Education, 2016). This means that these is a constant influx of student population and having historical data for each student can prove difficult. Schools in Texas are required to administer the STAAR examination to their students to prove that they are meeting the recommended standard according to the state. Using the STAAR examination as a benchmark for success in primary and secondary school, this study will focus on determining features of students that prove to predictive of STAAR achievement. The goal is to be able to determine student who will not meet the standard early in the academic school year so that interventions can be introduced to better provide the student with support.

Introduction to STAAR

The State of Texas Assessments of Academic Readiness (STAAR) examination is a standardized test that ranges the grades 3-12 and spans the subjects of math, reading, writing, science, and social studies. It was first introduced in the 2011-2012 school year in order to increase the rigor of its predecessor and whose content and performance standards are designed to align with career and college readiness. The test has three ratings, Level I, Level II, and Level III. Earning a rating of Level I indicates that a student did not meet the performance standard, earning Level II indicates meeting the recommended standard and earning a Level III indicates the student has mastered the material and exceeded the performance standard. Grades 3-8 take subject tests based on enrolled grade and grades 9-12 must take all end of course examinations at some point before graduation. Refer to Table 1 for view which tests are taken at each grade level.

Table 1. *Grade Level and Required STAAR Examinations*

Grade Level	**Examination**
3rd Grade	Math, Reading
4^{th} Grade	Math, Reading, Writing
5^{th} Grade	Math, Reading, Science
6^{th} Grade	Math, Reading
7^{th} Grade	Math, Reading, Writing
8^{th} Grade	Math, Reading, Science, Social Studies
9^{th}-12^{th} Grade	English I, English II, Algebra I, Biology, U.S History

Statement of the Problem

The measure for student success in Texas is measured by the STAAR examination. If a student meets or exceeds the standard on a given subject test of STAAR, they are considered mastering that subject. There exist many factors that cause a student to not be successful in school. Every school tries to identify students who are currently struggling or may struggle in the future in order to develop interventions. The process in which these students are identified vary from school to school and vary in accuracy. The problem there in lies how to select the students that need intervention so that resources are not wasted and the school has more students meeting or exceeding the STAAR examination.

Statement of the Purpose

The purpose of this study was to investigate student features that charter schools would have access to early in the school year. Using these features and models developed in this study, a school would be able to identify students that are predicted not to meet the STAAR standard at the end of the year for each subject in order for the school to provide additional support and intervention which would ultimately help the student meet the standard. To be clear, the purpose is not to accurately predict their STAAR level but rather capture as many students as possible who would not meet the standard. False positives for a student not meeting standard is a more desirable outcome than false negative, due to the fact that a student who would have needed additional support would not be identified.

CHAPTER II

REVIEW OF LITERATURE

Introduction

This chapter includes a review of literature involving the determination of features of student success, various machine learning algorithms used for classification and their comparison followed by studies that uses machine learning to predict student data.

Features of Student Success

In order to determine what features were going to be used in this study for predictor variables, a review of literature in impactful features of students for academic success was needed. Race and ethnicity in student education has been studied extensively and is even more important now that populations such as Hispanics continue to rapidly grow in schools across both charter and traditional public schools. The Hispanic population in schools doubled from 1990 to 2006 (Fry & Gonzales, 2008). The United States has had a history struggling and continues to struggle with achievement gaps between minority and nonminority students (Dee, 2005). Race and ethnicity has been proven to have an impact on student achievement and because of this, it was added into this study.

The College and Career Readiness and Success Center at American Institutes for Research conducted a study whose goal was to find predictors and indicators for postsecondary success. One of its many findings was that high school students with less than 10 percent of absences and a 3.0 high school GPA were more likely to succeed in postsecondary school. ACT also performed a similar study and focused on creating predictive models for postsecondary success. They used predictors such as gender, race, high school GPA, course work across multiple grades, and family income to create their models. It was found that work and performance done in the 8th grade had a significant impact in a student final grades of high school and early college (ACT, 2008).

There has been literature that shows a connection to poverty and reading proficiency to no graduating on time (Hernandez, 20012). This literature showed the 35% of children who had the factors of being poor, living in poverty stuck neighborhoods, and not reading proficiently had trouble graduating high school on time.

Machine Learning Algorithms

Selecting which classifiers to use would require a review in literature comparing different techniques, classifiers, and data composition. Once a review of these topics was done, this study could then use multiple classifiers that are found to work well with similar data used in this study.

Ensemble learning involves creating multiple models and combining them. This method often results in a better result than only creating a single model. Each model created casts a vote or an average is taken in order for the entire ensemble to come to a conclusion. Ensemble machine learning algorithms such as bagging and boosting trees, often times performance better than many classifiers (Dietterich, 2000). Boosted techniques were also studied to address class imbalance and were proven to more effective than other techniques such as data sampling (Seiffert, Khoshgoftaar & Hulse, 2009).

Nearest neighbor algorithms don't necessarily create models but rather uses the entire training set directly to make decisions. When a prediction needs to be made, it searches through the training set for a set amount of similar instances usually denoted by the variable K. K nearest neighbor algorithms usually struggle with feature significance and feature amount, but some studies have proposed using weighted approaches that ease this restriction (Zang, Qu, & Deng, 2015).

Classifiers are often used in the medical field for diagnosis. In the case of Parkinson's disease and Essential tremor, they share similar symptoms and support vector machines have been used to distinguish the two when diagnosing at a high accuracy rate (Surangsrirat, Decho & et al, 2016). Support vector machines can use a supervised or unsupervised approach, but if data is not labeled, then only an unsupervised approach is viable.

Machine Learning Algorithms Predicting Student Data

The process of using machine learning algorithms to predict student data is a field that has been explored in many studies. Shahiri et al. compiled a literature review of using machine learning on various predator variable sets (Shahiri et al., 2015). The variables mentioned in the review could be grouped in the following categories: internal assessments, external assessments, GPA, demographic, social interactions, financial and psychometric. The classifiers used were mainly Naïve Bayes, Support Vector Machines, Decision Trees, K-Nearest Neighbor, and Neural Networks. The purpose of the study was to establish a systematic literature review and a need for others to further study the use of machine learning on student data. This literature review proved to be correctly identify trends when exploring the field of machine learning on student data. There is an abundant amount of studies that use tree models as well as simple Naïve Bayes on predictive variables that are commonly used between studies. Studies usually differ themselves on the data being used or the variation of the classifier used.

Osmanbegović and Suljić used datamining techniques to form a twelve predictor variable model to predict student success in a college economics course (Osmanbegović & Suljić, 2012). Some predictor variables included GPA, entrance exam score, scholarships, and earnings. Naïve Bayes was found to perform the best when compared to other classifiers used. Most of the predictor variables used were quantitative and ranged in values rather than boolean value or psychological in nature.

Though there are many studies on predicting achievement in post-secondary school, Golino et al. used high school data and non-linear methods such as tree based models to predict their defined high and low student achievement (Golino, Gomes & Andrade, 2014). They claimed that using tree based models was better when using predictor variables that deal with physiological factors which are non-linear. The study showed an accuracy of 68.18% when using these models and hoped to improve upon this by increasing sample size and fine tuning the random forest parameters.

CHAPTER III

METHODOLOGY

Data Gathering

The data used in this study was from the charter school IDEA Public Schools. 11709 total students from all of the district's schools were used, ranging from grade 4 to 11. For each student the following data was gathered: student's state categorized race; current grade level; economically disadvantaged status; special education status; limited language proficient (LEP) status; previous year STAAR reading, writing, math, science and social studies level if applicable; on grade level for reading status; flag for missing more than 10% of the previous school year if applicable; current school year STAAR reading, writing, math, science and social studies level if applicable. These predictor variables were selected based on the literature review and what is available to charter schools. All pieces of data were available within the first week of the school year. This restriction was implemented with the goal in mind of being able to create interventions and action based on results early in a school year.

The first step in gathering data was to establish an understanding of the systems that house each desired variable and how to extract it. IDEA Public Schools uses Powerschool to hold student enrollment, demographic, attendance, and course grades. A query was then created to extract data for each student enrolled with IDEA Public Schools during the school year. This included an identifier that was used to connect all student data to a single student, state categorized race, economic disadvantaged status, special education status, LEP status, attendance and grade level. Race was categorized with an integer from 1 to 7, each of which corresponded to a state categorized race. IDEA Public Schools targets areas of low income and of minority to build schools so there exists a skewed portion of students who are Hispanic, economically disadvantaged and are LEP. This later proved to be a problem when developing models. Current grade level was extracted as an integer from 4-12. Grades below 4 were not extracted because they would not have previous year STAAR results. Economic disadvantaged, special education and LEP status were extracted as boolean values to represent if a student was flagged as being part of these sub groups. The attendance flag was a boolean value if the student missed more than 18 days of school which is 10% of the 180 days of instruction required by Texas. The data provided by Renaissance's STAR test administered at the beginning of a school year was extracted from an Excel file that was given to IDEA Public Schools on a daily basis. This test is administered multiple times throughout the school year to monitor student's growth in reading. The administration used in this study is the first one which is administered within the first few weeks of school. Some students who registered with IDEA Public Schools late did not have results for this test. To determine if a student was at or above grade level was the subtraction of the student's grade equivalency, which was a resulting measure from the examination, and current grade. If this was negative, the student was below grade level and a positive value showed a student above grade level. If the student was at or above grade level, a 1 was recorded. If the student was below grade level, a 0 was recorded and if the student did not take the assessment, a -1 was recorded. These values where then joined with the Powerschool data using the student's unique identifying number as a join key. Students' historical STAAR results were held in a system called DMAC. In order to extract students' STAAR results, the student must have tested or currently be enrolled with the district trying to extract the data. The STAAR results for previous and current school year was extracted from DMAC through downloadable Excel spread sheets for each subject, school, grade level, and school year. These spread sheets where then combined and for each subject and school year an integer from 0 to 3 was given. Receiving a 1 indicated that the student did not meet the standard for that subject test during that school year, receiving a 2 indicated that a student met the standard and receiving a 3 indicated that the student exceeded the standard. If a student did not take a subject test for the year, a 0 was recorded. Then for each school year and subject, the DMAC data was joined with the previous combined data set on the students' unique identifier. This completed the data gathering and transformation.

No matter the machine learning algorithm used, there had to be a division of learning, validation and testing data. In this study, 60% of the data was used for learning, 20% for data hold out validation, and 20% for testing. Performing a Fisher-Yates shuffle on the entire data set and pulling out the needed learning and hold out data provided the data necessary for creating models.

Model Creation

MATLAB, with its machine learning and classification extensions, was the environment used to create and run tests. For the subjects of math, reading, writing, science and social studies, a model needed to be created so that using the created model, we could predict the corresponding subject STAAR level for the student.

Each subject's testing and validation data was divided by using the student's state categorized race, current grade level, economically disadvantaged status; special education status, LEP status, previous year STAAR reading, writing, math, science and social studies level, on grade level for reading status, and attendance flag as the predictor variables with the current year results for the given subject as the response variable. The response variable ranged from 0-3. A class of 0 indicates a student who would not take that subject test this year, 1 means the student did not meet the standard, 2 means the student met the minimum passing standard and 3 means they exceeded the passing standard. 20% of the data was used for holdout validation.

A variety of classification learners where used on the data such as decision trees, nearest neighbor, and support vector machines. The student population was heavily skewed for certain predictor variables such as race, economic disadvantage status and even the previous school year levels since a majority of the students at least met the standard on the STAAR examination. RUS Boosted Decision Tree proved to be the best classifier to use. This was mainly due to the fact that RUS Boosted Decision Trees handle data that has heavily skewed predictor variables. The following sections will detail each subject's model development results.

Math Model Creation

Going through the permutations of prediction variables it was found that using all variables provided the best ROC curve and confusion matrix results for math. ROC curves are often sued to compare classifiers when looking at the area under the curve (AUC). AUC is a standard for performance assessment of classifiers. An AUC of 0.5 is simply no better than a coin toss while a AUC of 1.0 is the best case for any classifier. The ROC curve in Figure 1 below shows a 0.69 area under the curve (AUC) with respect to the response variable class 1, which is not meeting standard on the STAAR examination.

Figure 1. *RUS Boosted Decision Tree ROC Curve for Math Using All Prediction Variables with Respect to Class 1*

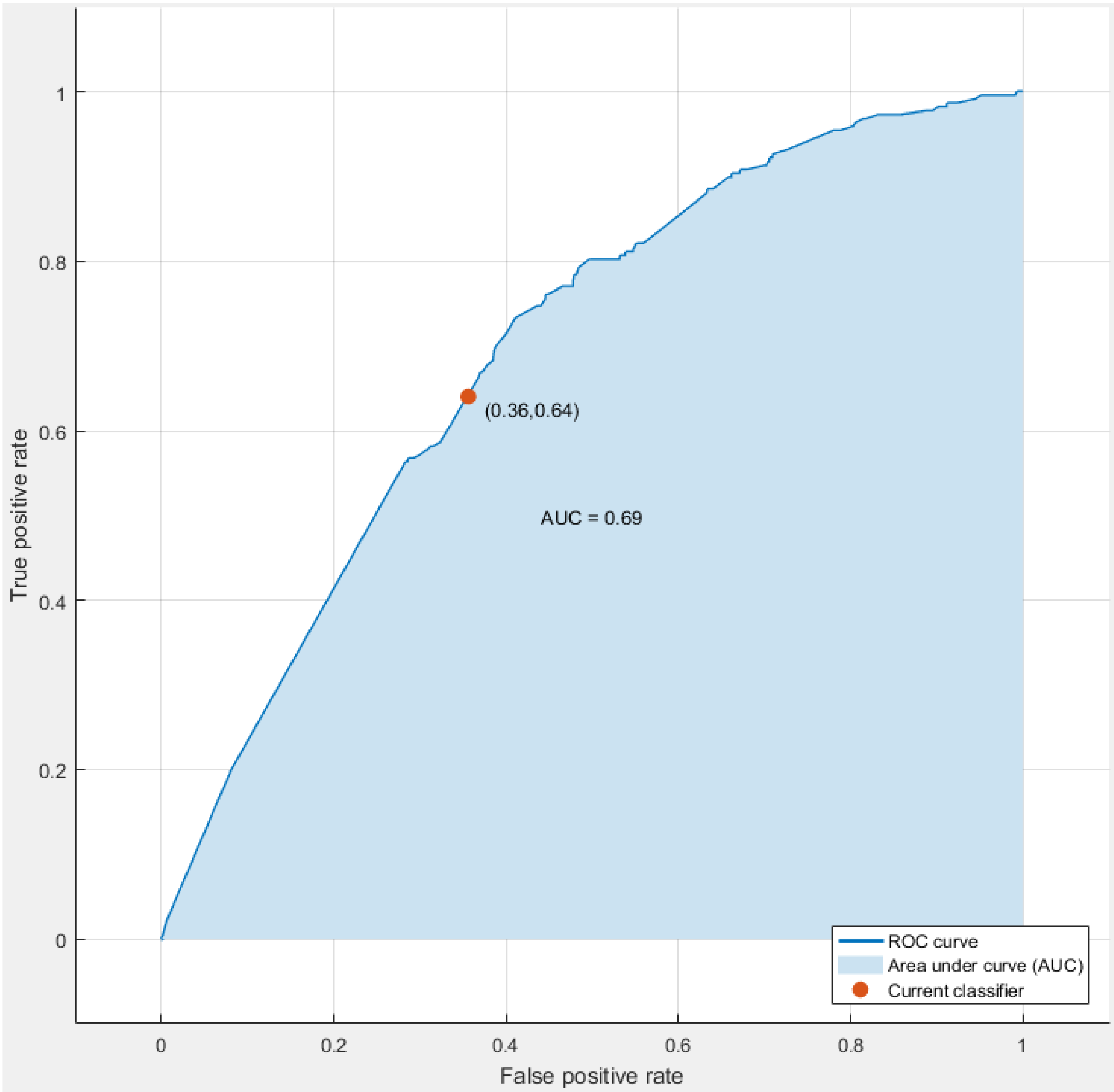
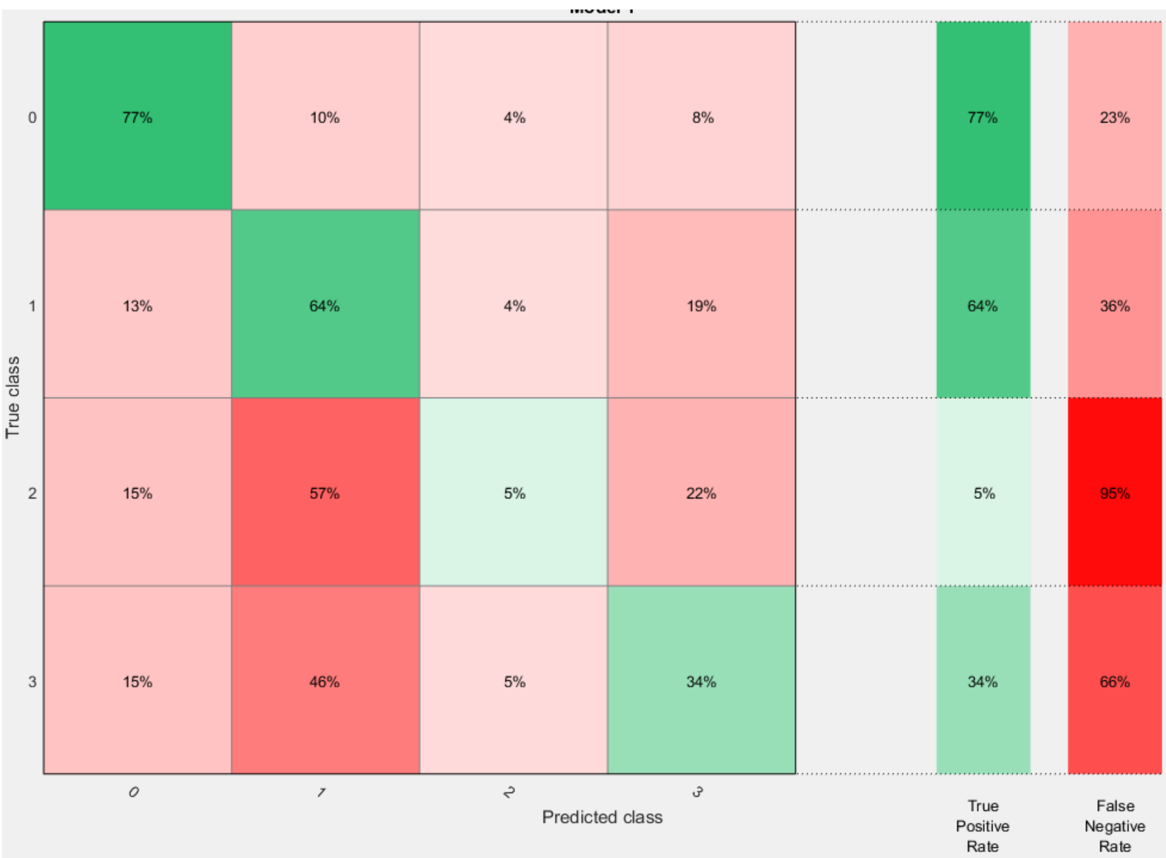


Figure 2 shows the confusion matrix using RUS Boosted Decision Trees when using all predictor variables. Remember the goal of this study is to not simply achieve the best classifier in terms of true positive results across the diagonal, but rather a focus on the achieving the highest possible true positive rate for the class of 1. Accomplishing this will allow a school to provide intervention and support to those students who need it. Figure 2 shows that for math, the classifier was able to accurately predict 64% of students who would not meet the passing standard on their math STAAR examination.

Figure 2. *RUS Boosted Decision Tree Confusion Matrix for Math Using All Prediction Variables*



Reading Model Creation

Reading was a subject that had best results when excluding some predictor variables. Removing current grade level, LEP status, attendance flag, and at reading grade level flag produced the ROC curve shown in Figure 3. The curve shows an AUC of 0.88, which is higher than that of math.

Figure 3. *RUS Boosted Decision Tree ROC Curve for Reading Using 8 out of 12 Prediction Variables with Respect to Class 1*

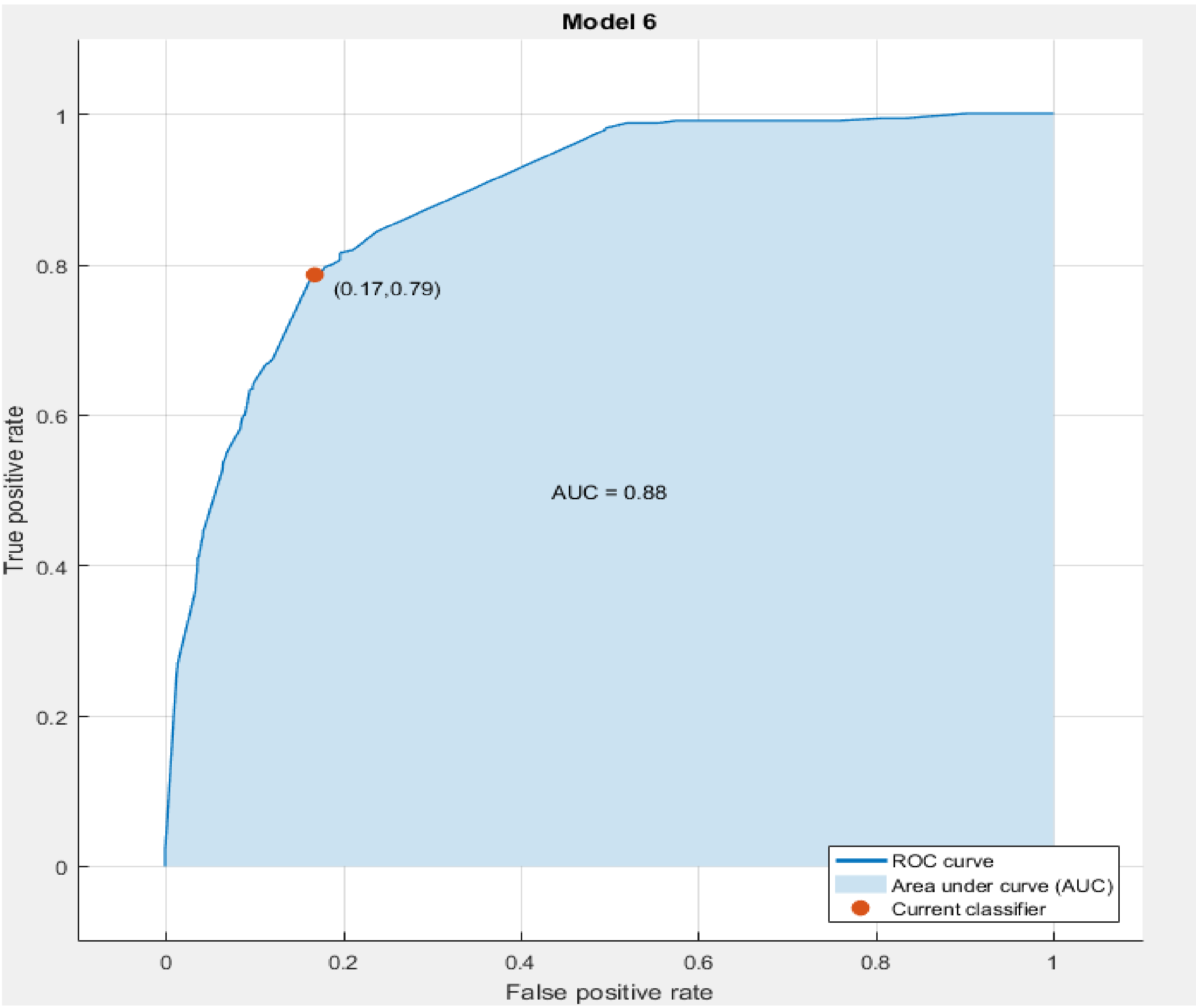
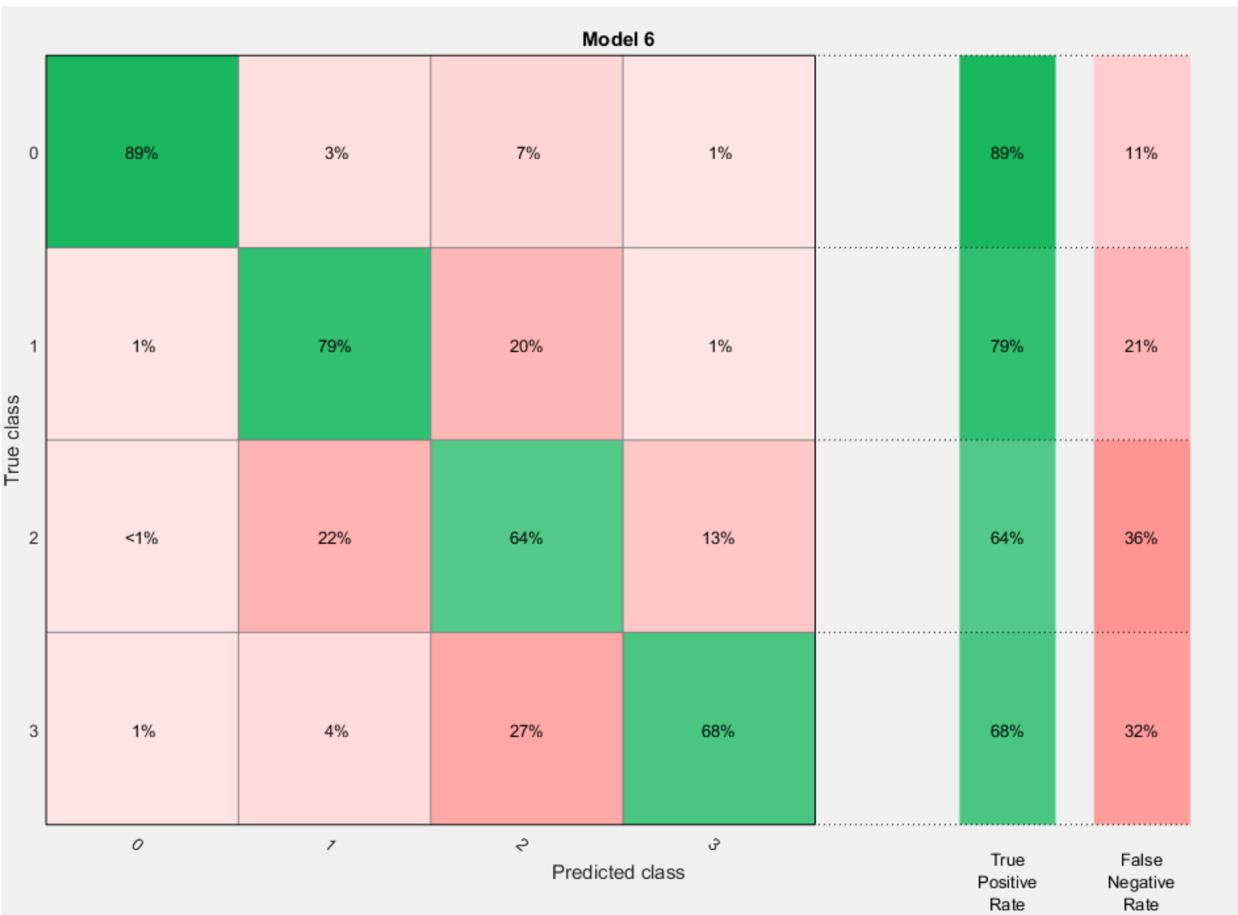


Figure 4 shows the confusion matrix using RUS Boosted Decision Trees when excluding the predictor variables mentioned above. It can be seen that 79% of the student who would not have met the standard were accurately predicted. What is interesting is not only was the true positive rate for class 1 good, but for all classes as well. The diagonal shows that this is a promising model for predicting all classes, though for this study we are only interested in class 1.

Figure 4. *RUS Boosted Decision Tree Confusion Matrix for Reading Using 8 out of 12 Prediction Variables*

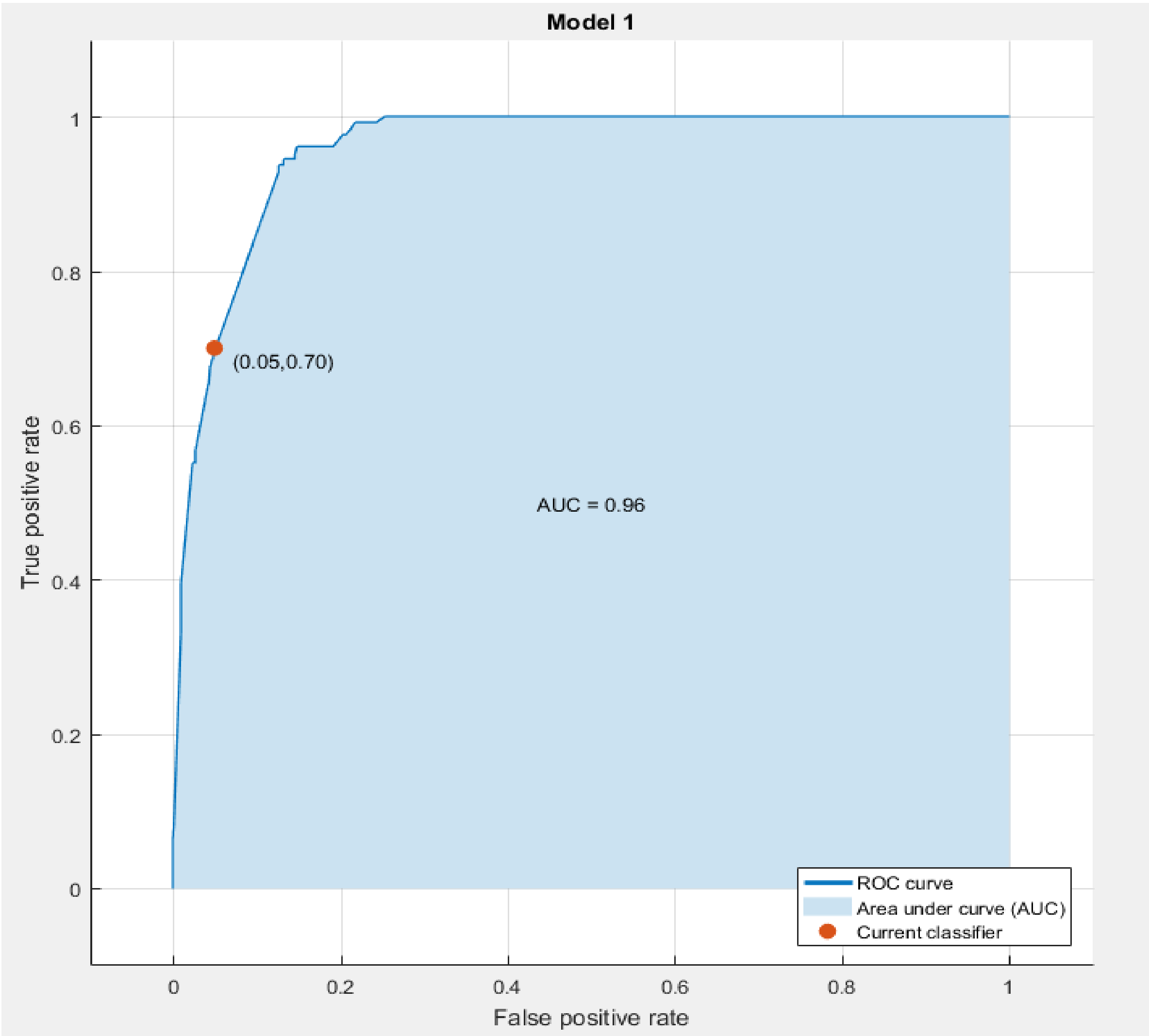


The Renaissance STAR test provided to IDEA Public Schools in order to determine if a student is on grade level for reading was excluded in this model. It can be said that this test is not aligned to STAAR reading as it lowered the accuracy of the model when including it. It also seems like it does not matter is a student misses more than 10% of school when compared to math.

Writing Model Creation

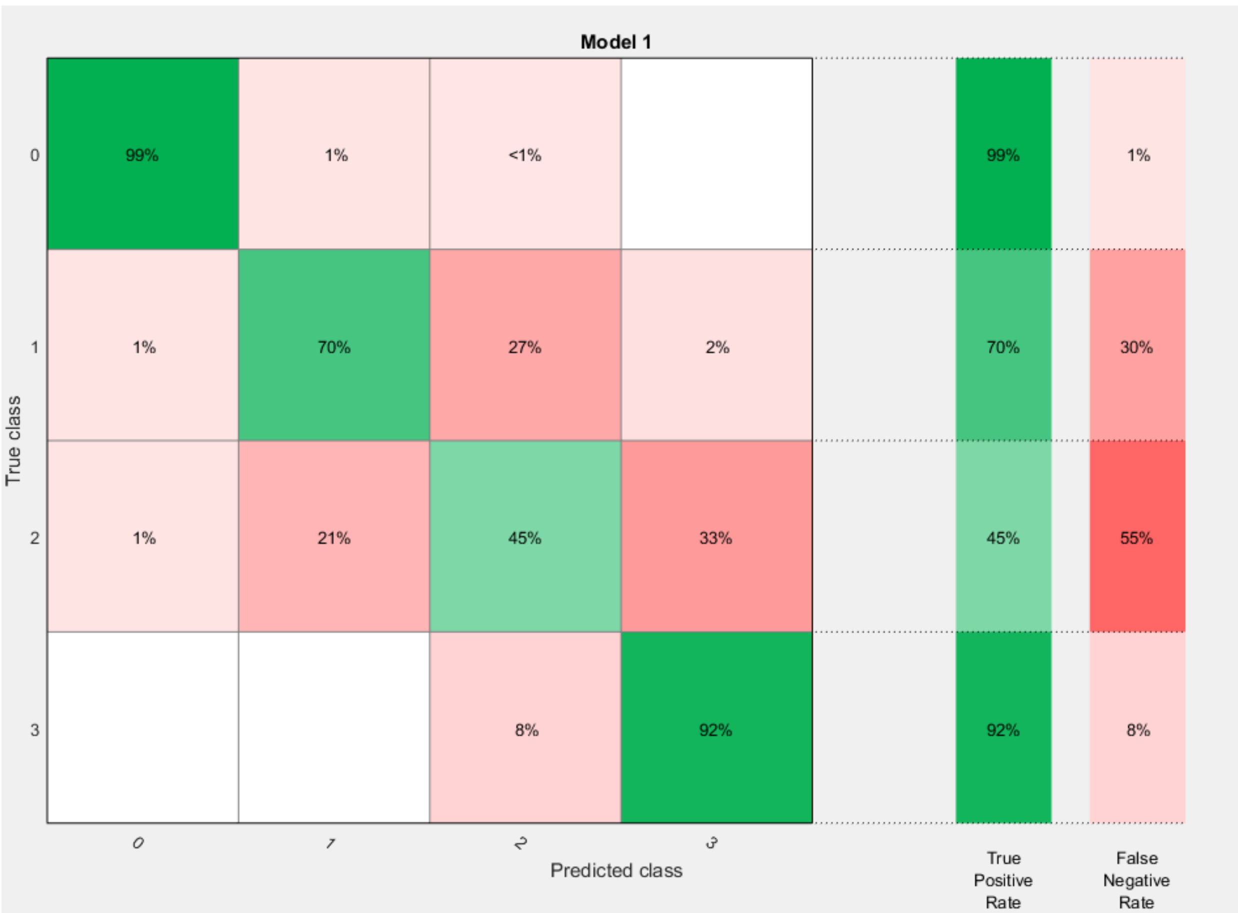
Similar to math, using all the predictor variables proved to provide the best results for the subject of writing. The ROC curve using RUS Boosted Decision Trees with respect to class 1 is shown in Figure 5. The AUC for this ROC curve was 0.96, which is considered excellent.

Figure 5. *RUS Boosted Decision Tree ROC Curve for Writing Using All Prediction Variables with Respect to Class 1*



Looking at the confusion matrix in Figure 6 which was produced using RUS Boosted Decision Trees and all predictor variables, it can be seen that 70% of the students who would not meet the recommended standard for the writing STAAR test were accurately captured. Unlike reading which excluded whether a student was reading on grade level, writing proved to have a better result when including it. This was unexpected after seeing the reading performance since writing and reading performance are usually highly dependent on a student's reading grade level and if the test used to determine this wasn't aligned with reading, it was expected to not be aligned with writing.

Figure 6. *RUS Boosted Decision Tree Confusion Matrix for Writing Using All Prediction Variables*



It is interesting to note that 92% of the students exceeding the recommend standard for writing STAAR were captured and would be worth investigation in future work.

Science Model Creation

For science, using all predictor variables provided the best ROC curve and confusion matrix. Figure 7 shows that an AUC of 0.93 was achieved with respect to class 1.

Figure 7. *RUS Boosted Decision Tree ROC Curve for Science Using All Prediction Variables with Respect to Class 1*

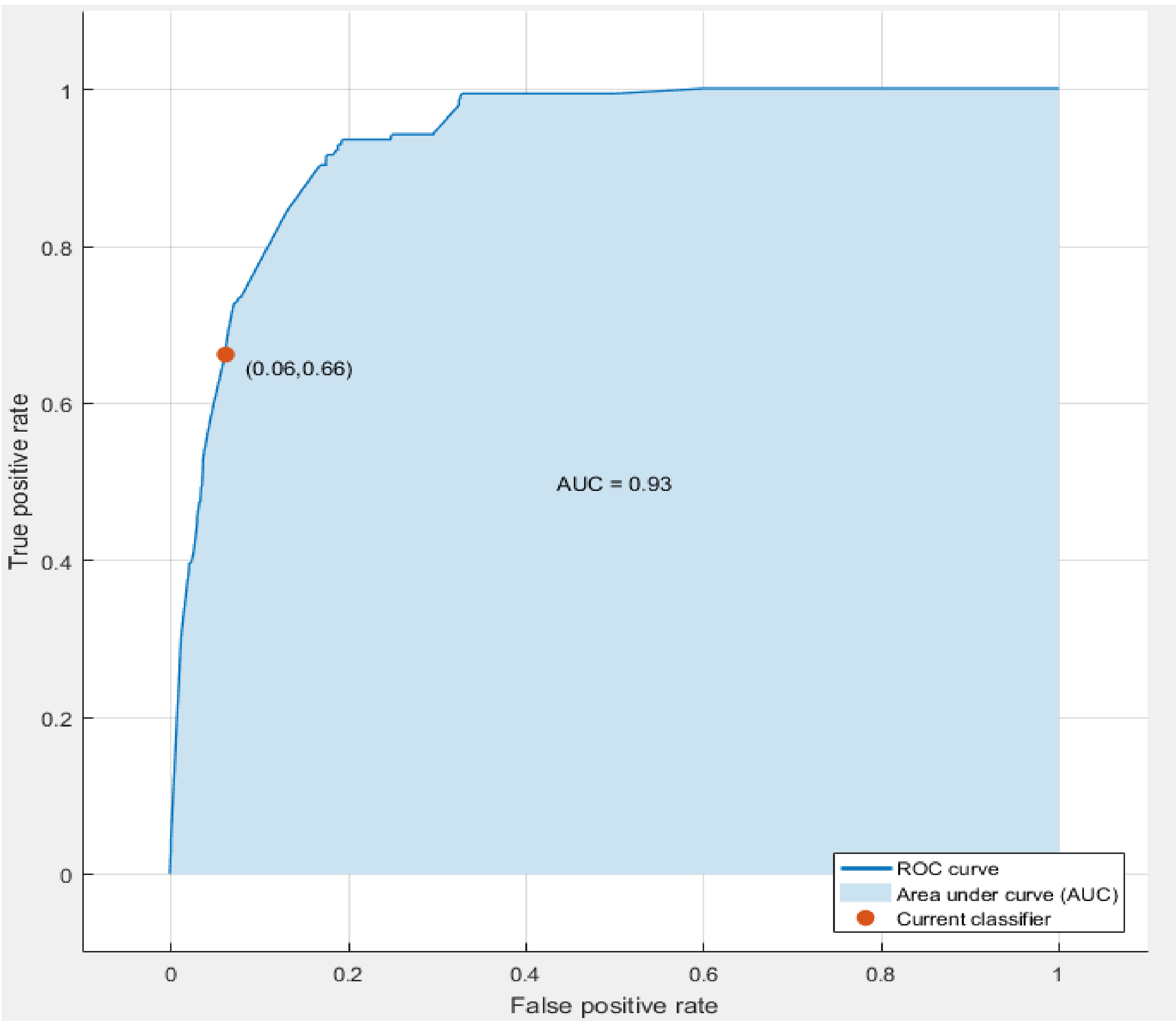
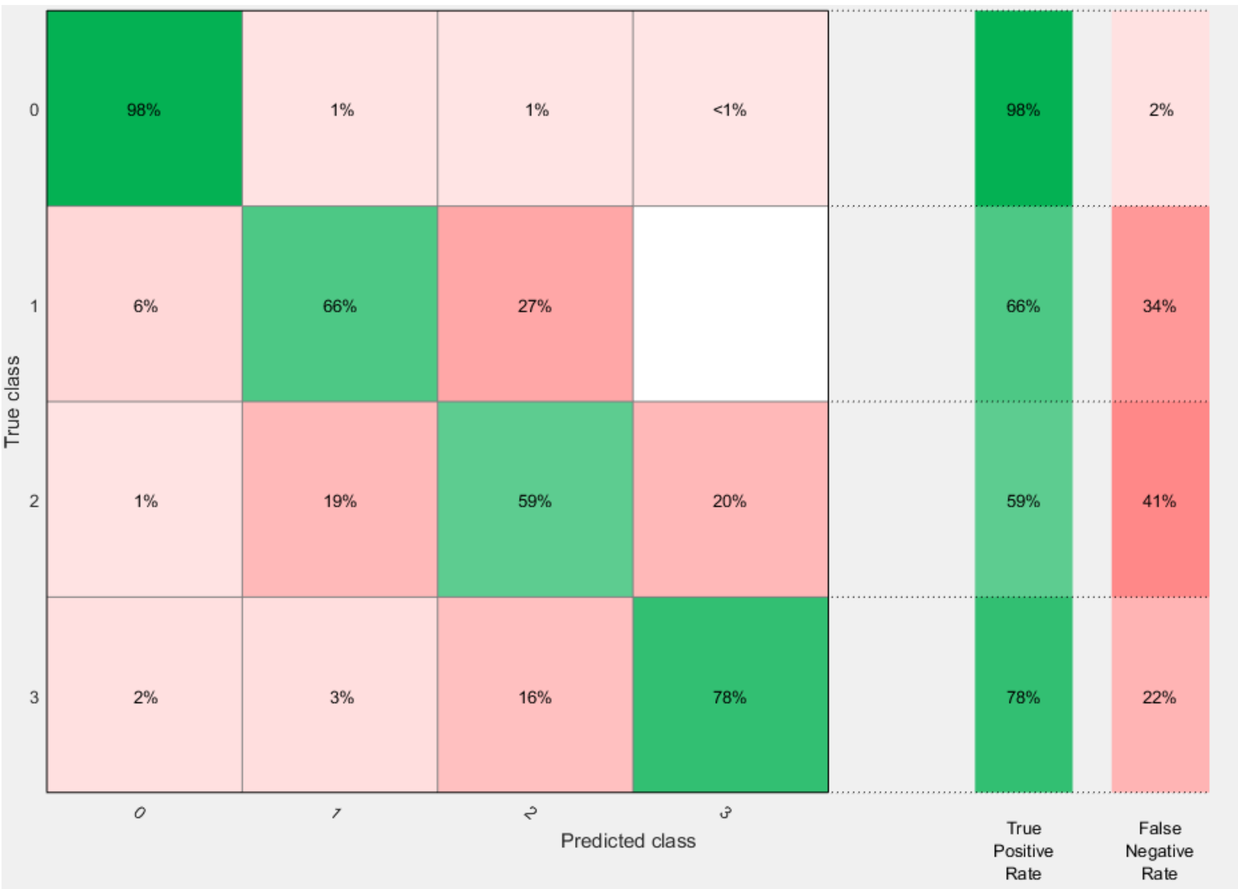


Figure 8 was the confusion matrix produced when using all response variables and RUS Boosted Trees. 66% of students not meeting the recommend standard were captured.

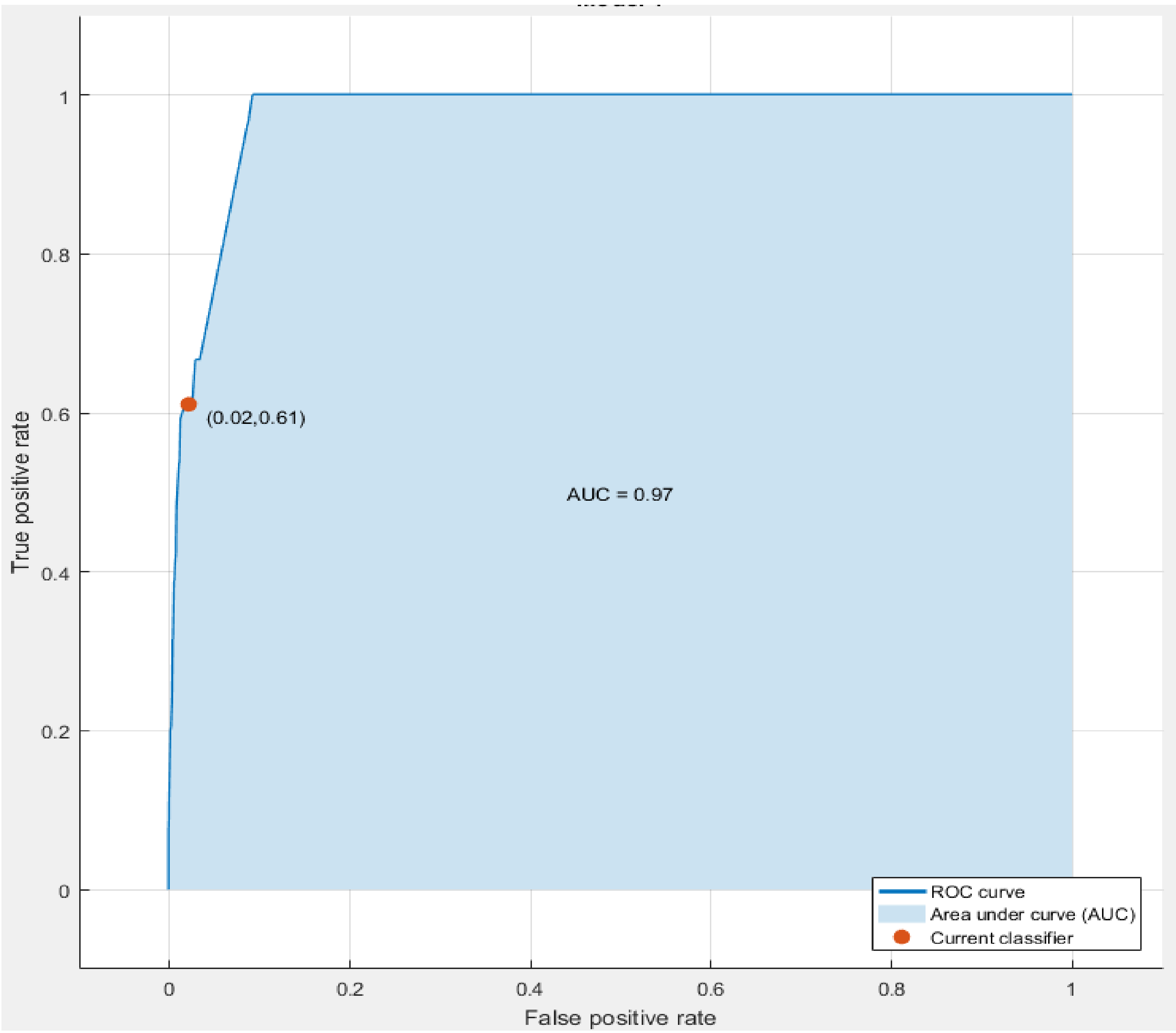
Figure 8. *RUS Boosted Decision Tree Confusion Matrix for Science Using All Prediction Variables*



Social Studies Model Creation

Social studies also provided the best results when using all predictor variables. Figure 9 shows an AUC of 0.97 which is the highest out of all the subjects.

Figure 9. *RUS Boosted Decision Tree ROC Curve for Social Studies Using All Prediction Variables with Respect to Class 1*



Though social studies had the highest AUC, it did not have the highest accuracy of detecting true positives of class 1. Figure 10 shows the confusion matrix produced when using RUS Boosted Decision Trees and all predictor variables. 61% of students not meeting the required passing standard were captured using this model.

Figure 10. *RUS Boosted Decision Tree Confusion Matrix for Social Studies Using All Prediction Variables*

Conclusion of Model Creation

Once all the models were finalized, the next step was to export the models and run them on the remaining 20% of data to analyze the performance on data that had not been seen yet. This would be the final test to see the fitness of the models and if they could be used to accurately detect students who were not going to meet the recommended standard for a given subject. Looking at the confusion matrix and ROC curve of all the subjects it was assumed that reading was going to have the best results with social studies having the worst. Chapter IV goes over the results of the final tests for each subject.

CHAPTER IV

SUMMARY AND CONCLUSION

Introduction

Once all the subject models were finalized, they were exported in order to be used for the remaining set of data that was set aside for testing. Each model was then ran on its subject data whose results are shown in the following sections with confusion matrices. Following those sections will be a discussion of the results and future work.

Math Results

It can be seen in Table 2 that 57.38% of students not meeting standard were captured when the math model was used on data it had not yet seen.

Table 2. *Confusion Matrix for Math Using Math Model*

		Predicted Class			
		0	**1**	**2**	**3**
True Class	**0**	80.88%	8.87%	3.69%	6.57%
	1	22.95%	57.38%	5.33%	14.34%
	2	15.58%	57.68%	5.26%	21.47%
	3	13.62%	51.61%	3.94%	30.82%

Reading Results

The reading model was able to accurately predict 71.28% of the students who would not meet standard. This confusion matrix is shown in Table 3.

Table 3. *Confusion Matrix for Reading Using Reading Model*

		Predicted Class			
		0	1	2	3
True Class	0	95.35%	2.79%	1.40%	0.47%
	1	1.83%	71.28%	25.59%	1.31%
	2	0.72%	21.28%	61.25%	16.76%
	3	1.15%	4.03%	29.11%	65.71%

Writing Results

Upon using the writing model on the remain testing data, it was found that it was able to predict 68.92% of students who would not meet the writing standard on STAAR. The confusion matrix is shown in Table 4.

Table 4. *Confusion Matrix for Writing Using Writing Model*

		Predicted Class			
		0	1	2	3
True Class	0	99.12%	0.77%	0.06%	0.06%
	1	2.03%	68.92%	26.35%	2.70%
	2	0.88%	20.35%	44.03%	34.73%
	3	0.00%	2.33%	20.93%	76.74%

Science Results

The science model predicted 72.97% of student not meeting the recommended standard for the science STAAR examination. Table 5 shows the confusion matrix for this model's results.

Table 5. *Confusion Matrix for Science Using Science Model*

		Predicted Class			
		0	1	2	3
True Class	0	99.15%	0.78%	0.07%	0.00%
	1	4.86%	72.97%	21.62%	0.54%
	2	1.72%	20.31%	60.07%	17.90%
	3	1.85%	0.62%	27.16%	70.37%

Social Studies Results

The social studies model achieved 67.61% accuracy when considering the students who would not meet the recommended standard on their social studies STAAR examination. Table 6 shows this models results.

Table 6. *Confusion Matrix for Social Studies Using Social Studies Model*

		Predicted Class			
		0	1	2	3
True Class	0	99.44%	0.21%	0.05%	0.31%
	1	4.23%	67.61%	28.17%	0.00%
	2	0.00%	18.18%	49.78%	32.03%
	3	0.00%	3.33%	22.22%	74.44%

Discussion

This study's focus was to use various features available to a charter school early in the school year to predict which students would not meet the standard on a given STAAR subject examination. To summarize the results of the models created in this study with respect to this focus, math had a 57.38% accuracy, reading a 71.28%, writing a 68.92%, science a 72.97%, and social studies a 67.61%.

Math was a model that used all predictor variables but still struggled to get great performance. There were a total of 244 students who did not meet the recommended standard for math and only 140 were predicted using this model. When compared to the other subject model, these results were the worst. The math STAAR examination has gone through many changes and continues to be refined. This may lead to inconsistencies from student to student when trying to prepare them for the exam. These inconsistencies between schools and teachers, may have been a cause for this models performance. The same provider that identifies if students are on grade level for reading is working on providing that data for all grade levels of math. This may be a worth adding into the predictor variables and could not only impact math but science as well.

The reading model only used 8 out of the 12 predictor variables and was able to identify 273 of the 383 students who did not meet the standard for reading. The reading STAAR exam is one that has gone through the least amount of changes and thus this model could be used year to year and produce similar results. The exclusion of the on grade level predictor could be something to investigate and add back in as the provider continues to refine its methods. This model not only performed well in predicting students who would not meet the standard but also those who would meet or exceed the standard. This model could be used in ways other than providing intervention to students such as creating groups of students with high performance.

The writing model performed better than math but still could use some further investigation. It too used all the predictors and even the inclusion of on grade level for reading which the reading model excluded. Writing is a subject that is only taken in 4^{th} and 7^{th} grade so it does not have the bulk of data that math and reading have. Using more data or historical data could increase the performance of the model when applied to this study's focus.

The science model performed the best out of all the models. Like that of reading, the science model had favorable results for all class categories which could result in it being used in way other than what is focused in this study.

The social studies model performed very similar to that of writing when comparing its ability to predict students not meeting/meeting/exceeding standard. This subject is only tested in 8^{th} grade and once in high school with the U.S History examination. This model could make use of more data whether it be current enrolled student data or modifying it to be more historical.

Future Research

Though IDEA Public Schools is one of the largest charter schools in Texas, they are not the only one. Gathering data from multiple charter schools and adding a predictor variable to indicate the district would be worth a study. Different charter schools have different cultures they enforce at each school. This cultural difference has been shown to have a significant impact in students (Balfanz, 2009). This study used 12 predictor variables, while these were selected because they were available early in the school year, there exist predictors that occur toward the middle of the school year that have impact on student achievement. Some of these could include behavior incidents, midyear attendance records, and midyear GPA. Though a charter school does not have access to a student's entire school history, they do have the ability to gather the historical STAAR results for a given student. Using the entire history of a student could prove meaningful and there could be trends that lead to better model accuracy. Research has also been done on more qualitative variables. The study performed in Chicago schools measured relationships such as teacher-student trust and computer availability (Allensworth & Easton, 2007). Adding these types of variables, could provide stronger models.

Summary

The purpose of this study was to develop models that could accurately predict students who would not meet the standard for the STAAR examination. The models would use data readily available to charter schools early in the year in order for them to set up interventions and support programs for the students predicted not to pass. This study developed models for each tested subject and had a variety of accuracy between the subject models.

The results of this study showed that the predictor variables used worked for most subjects but math. This subject needs further investigation and the possibility of different predictor variables to bring it in line with the other subjects. Future work could include midyear predictor variables and the addition of historical STAAR achievement.

Little research has been done on predicting student achievement in primary and secondary school as well as little research on charter schools when compared to the amount of literature for postsecondary success and college completion. As charter schools continue to grow, there exists a need to study what makes them different and how students perform in this setting. Research from this study will contribute to the literature on machine learning on charter school student data and identifying success for these students.

REFERENCES

ACT. (2008). “The forgotten middle: Ensuring that all students are on target for college and career readiness before high school.” Iowa City, IA: ACT.

ACT. (2012). “The condition of college & career readiness.” Iowa City, IA: ACT.

Allensworth, Elaine M., and John Q. Easton. "What Matters for Staying On-Track and Graduating in Chicago Public Highs Schools: A Close Look at Course Grades, Failures, and Attendance in the Freshman Year. Research Report." *Consortium on Chicago School Research* (2007).

Dee, Thomas S. "A teacher like me: Does race, ethnicity, or gender matter?." *The American economic review* 95.2 (2005): 158-165.

Dietterich, Thomas G. "Ensemble methods in machine learning."*International workshop on multiple classifier systems*. Springer Berlin Heidelberg, 2000.

Fry, Richard, and Felisa Gonzales. "One-in-Five and Growing Fast: A Profile of Hispanic Public School Students." *Pew Hispanic Center* (2008).

Golino, Hudson F., Cristiano Mauro Assis Gomes, and Diego Andrade. "Predicting Academic Achievement of High-School Students Using Machine Learning." *Psychology* 5.18 (2014): 2046.

Hart, P. "Rising to the challenge: Are high school graduates prepared for college and work." *Study conductedforAchieve, a nonprofitorganization for education reform, Washington, DC. Retrievedfrom http://www. achieve. org/risingtothechallenge* (2005).

Hernandez, Donald J. "Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation." *Annie E. Casey Foundation* (2011).

National Center for Educational Statistics. "The Condition of Education - Participation in Education - Elementary/Secondary - Charter School Enrollment - Indicator April (2016)." *The Condition of Education - Participation in Education - Elementary/Secondary - Charter School Enrollment - Indicator April (2016)*. N.p., n.d. Web. 1 Oct. 2016.

Osmanbegović, Edin, and Mirza Suljić. "Data mining approach for predicting student performance." *Economic Review* 10.1 (2012).

Seiffert, Chris, Taghi M. Khoshgoftaar, and Jason Van Hulse. "Improving software-quality predictions with data sampling and boosting." *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 39.6 (2009): 1283-1294.

Shahiri, Amirah Mohamed, and Wahidah Husain. "A Review on Predicting Student's Performance Using Data Mining Techniques." *Procedia Computer Science* 72 (2015): 414-422.

Surangsrirat, Decho, et al. "Support vector machine classification of Parkinson's disease and essential tremor subjects based on temporal fluctuation." *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. IEEE, 2016.

Zhang, Nana, Yanpeng Qu, and Ansheng Deng. "Evolutionary Extreme Learning Machine Based Weighted Nearest-Neighbor Equality Classification." *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2015 7th International Conference on*. Vol. 2. IEEE, 2015.

BIOGRAPHICAL SKETCH

Christopher D. Gonzalez received his Bachelor of Science in Computer Science in the spring of 2014 at University of Texas Pan American. He then furthered his education by earning a Master of Science in Computer Science degree in the fall of 2016. Christopher worked for University of Texas Pan American under the HHMI grant developing a web application to support the management of student research for student mentors from 2012 to 2014. He then moved on to work for IDEA Public Schools as a data analyst in 2014. His responsibilities included developing reports using student and operational data in order for the organization to be able to make informed decisions. He was then promoted to data architect in 2015 which expanded his responsibilities to developing structures to organize and store data as well as informing stakeholders of important aspects of the data. In 2016 he was then promoted to business intelligence project manager which he then planed, guided, developed and maintained software systems that were used by stakeholders in the organization to digest data and inform decisions. His permanent mailing address is P.O Box 881 Weslaco Texas, 78596.