Using Deep Q-Learning to Predict Student Success

Sam Myers
Computer Science
Truman State University
Kirksville, Missouri, USA
(636)-255-4662
Sjm3253@truman.edu

ABSTRACT
This study had two main goals: to see what degree of accuracy an artificial intelligence system (AI) could predict student scores in a course well as to see how the amount of data given impacts this accuracy. To test both, an AI utilizing deep q-learning was fed data made up of information relevant to student success, including, but not limited to, parental civil status, average time spent studying and weekly alcohol consumption. Using this data, the AI is able to make a prediction of the score which is subsequently compared to the student's actual score. The AI then is able to use this difference in new calculations, using algorithms to determine which variables are more relevant and which are not. This process repeats through all data given until it reaches the end, at which point the percent error is given to the user. The AI used in this study was given three different sets of data, each containing a different multiple of the same data. Sample 1 consisted of one set of the data, not duplicated. Sample 2 consisted of four sets of the data, while Sample 3 consisted of sixteen sets. After being fed each dataset, the AI performed noticeably better as the amount of data increased, decreasing in percent error with each increase in data. Sample 1 averaged a percent error of around 27.768%, while Sample 2 averaged a percent error of around 6.118%. Sample 3 averaged a percent error of around .989%.

CS Concepts
• Computing methodologies → Machine Learning → Learning Settings • Computing methodologies → Machine learning •Computing methodologies → Machine learning → Machine learning approaches → Neural networks

Keywords
Artificial Intelligence; Deep Q-Learning; Dataset

1. INTRODUCTION
In all stages of life, one is judged on the actions they take. Parents warn their children about the perils of underage drinking. Teachers encourage their students to study, saying it will no doubt lead to success. Counselors worry about familial relationships and the amount of exercise one gets. Cruel as it may seem, there is a reason for this.

Over time humans have learned to predict what leads to triumph and what leads to failure. A student who never misses a day of school will, in all likelihood, perform better than one that misses one day out of three. A teen whose parents make 100,000 dollars a year will make better grades than one whose family lives in poverty. A child whose parents are separated is more likely to struggle in school than one whose parents live together. By recognizing these discrepancies, one is able to address the problem, giving those who may be struggling the adequate help they need so that they may succeed as well. However, one has to wonder if this is a purely human thing? Does one need to empathize, connect with another being in order to predict where in life they are headed? Or can a person be stripped down to their base characteristics and be handed to a robot to get the same results? By using Deep Q-Learning to analyze data from hundreds of students, this paper aims to analyze the effectiveness of AI in predicting student success, as well as to analyze the accuracy of the AI when given differing amounts of data. In order to do so, the AI will make a prediction based on given data, such as parents' education level, distance from school and participation in extracurricular activities. The AI's predictions will then be compared to known success rates and graded on the margin of error before being fed back into the program to be processed and improved upon. This process will repeat until the list of students is exhausted, at which the AI will present the accuracy of its prediction. This entire process will once again be repeated, instead using differing amounts of data in order to find a correlation between amount of data and accuracy of the AI. Once this process is completed, the difference in percent error will be compared and analyzed. With both results (Accuracy of predictions, difference in percent error), the study will be able to evaluate the AI's performance differs across datasets allowing for a better idea of how the AI functions as a whole.

2. LITERATURE REVIEW
This study is not the first to use AIs to predict student success. Over the years a variety of studies have been published with similar goals and similar methodologies, one example an in depth look at various studies performed by Rastrollo-Guerrero et al.. This study had the goal of presenting a complete look at different techniques and algorithms used to predict student success. To do so the study compiled 64 different articles in this field, excluding articles lacking in quality and contribution. This study first highlights issues in other studies, such as the high propensity that the study will focus on university age students, with such studies making up 87.5% of total studies performed [5]. Rastrollo-Guerrero et al. go on to divide the AIs used into four types, these being supervised machine learning, unsupervised machine learning, artificial neural networks and collaborative filtering. From this, it is concluded that supervised learning was the most widely used out of these types, being used in half of all studies [5]. As a result of this, Rastrollo-Guerrero et al. warns that results gleaned may favor supervised learning simply due to the immense amount of such studies. The studies are then further split into four different purposes; these being student dropout rates, student performance, recommender activities and resources and student knowledge [5]. Similar to supervised learning, student performance makes up a majority of these studies, making up 70% of all studies performed [5]. The conclusions reached by this paper expand on these numbers, stating that supervised learning is the most popular due to its accuracy and likelihood of providing accurate results [5]. Recommender systems follow up supervised learning in popularity, yet for a different reason. Recommender systems, Rastrollo-Guerrero et al. explains, find success mainly in recommending resources for students, a marked difference from supervised learning. [5]. Nonetheless, these results are promising and may potentially open the door for

more studies to utilize this method. Neural networks are less used as well, yet have been shown to obtain a “great precision in predicting the student’s performance” [5]. The study further conclude that neural networks might find greater use in other areas of education, such as dropout predictions or activity recommendations. Finally, the study concludes that while unsupervised learning is used less due to its low accuracy, this fault may provide an incentive for those to improve upon it, bettering its use in the field as a whole [5].

One can take a closer look at a study like the ones analyzed by Rastrollo-Guerrero et al. by looking at Yao et al.’s 2019 study titled “Predicting Academic Performance for College Students: A Campus Behavior Perspective”. This study is built around a data set spanning around 3 years that details the location of over 6,000 students. This data gives a variety of information, such as student’s sleep schedule, time spent studying and time spent with other students of the same discipline [6]. These figures are gathered from the use of a campus smart card that students use in order to gain access to facilities, a system found to be more complete and more accurate than surveys in which the one giving the response is not pressured into honesty [6]. This data also provides a look into factors not before considered, such as who students spend their time with, a fact that may at first seem useless but has been proven to strongly correlate to how a student will perform [6]. With this data the study identifies three challenges to deal with, these being the difficulty in modeling changes in prediction across major, the variance in student population across major, and a lack of sufficient behavioral data for some students [6]. As a solution to these challenges, the study utilized a novel Multi-Task Learning-To-Rank Academic Performance Prediction framework (MLTR-APP) [6]. This program uses a sequential smoothing regularization in order to model correlation between semesters. It also recognizes that while different strengths in students have different impacts by major, there is still commonality in how these strengths impact performance. To counter the lack of data for some students, the program incorporates student similarity, taking in three different characteristics (Diligence, orderliness and sleep pattern) and uses these characteristics to predict how a student would behave with the data the program does have. Finally, the multi-task learning model is integrated with student similarity and trained across two datasets, the first made up of 3,352 students from 18 majors across 5 semesters, and the second made up of 3,245 students from 17 majors across 5 semesters. This shows the effectiveness of the model as well as that each characteristic listed earlier is effective for predicting student success [6]. With this model and data, the study can effectively find its results, overcoming the challenges identified before. The first result is the effect of diligence on student success. Yao et al. found that while correlation between diligence and success is significantly lower in the first semester (likely due to the fact that the first semester nearly entirely relies on what is taught in high school), in semesters after, diligence has a strong correlation with success. For orderliness, results are similar, with students who consistently are more orderly receiving better grades when compared to those who are not. The study also found correlation between student’s sleep schedule and success, with students with earlier wake-up times and earlier bedtimes performing better than those with later wake-up and bedtimes. However, no significant effect on correlation is found after this is integrated with diligence and orderliness, potentially due to the estimates needed in order to find a student’s wake-up and bedtimes [6]. A strong correlation was also found between frequently co-occurring students, a correlation that could be used to predict one student’s success when other’s success is available. With this data, the study is able to accurately figure what factors play an important role in student success, and with expansion, could in all likelihood accurately predict a student’s grade. The study this paper focuses on aims to have similar data, albeit with less thorough information and more categories that play a role in student success.

While it is undoubtably important to review similar studies, it is also important to review how studies similar to Rastrollo-Guerrero et al. and Yao et al. have played a role in increasing student success. One such result is covered in the article titled “Using Artificial Intelligence with Human Intelligence for Student Success” by Thomas Miller and Melissa Irvin [2]. This article covers the results of an AI introduced at the University of Southern Florida with the purpose of increasing first year retention. Using data such as information from the admissions process as well as a survey sent to all students, the program was able to build a model that used logistic regression in order to predict students who are at risk of dropping out [2]. While not always 100% accurate, the program has displayed a reasonable degree of accuracy, helping the college to better identify a group they can more effectively serve. The University of Southern Florida also uses a program in order to predict student grades, using the same data the first program does for a different purpose. After predicting the student’s first semester grade, the program compares it to the grade they received, with cases where these grades differ in a negative factor being noted. Such cases can be used to indicate a student may need help and that a visit may be beneficial. The University of Southern Florida is not the only school to utilize artificial intelligence either. According to an article written by Matthew Lynch [1], Southern Connecticut State University began working with IBM on increasing student retention. By integrating IBM’s Watson Analytics, the university aimed to improve student success and retention. Watson was able to help the university gain insight to what factors most influenced how a student performed, leading to the university opening the Academic Success Center, a building that provides help for students in their academic pursuits [1]. Georgia State had similar results after implementing a similar system. Using predictive analytics, the university was able to identify that nursing students who performed poorly in Introductory Math were less likely to succeed in their course, with only 10% of students who received a C or below graduating. As a result of this, Georgia State increased the amount of student advisors, leading to a five percent increase in its four-year graduation rate, and a six percent increase in its six-year graduation rate [1]. Finally, The University of Arizona has begun using machine learning in a similar fashion to how Yao. et al. did, accessing data from student ID cards in order to best identify what actions cause student retention [1]. Overall, attempts to use AI to increase student success have been largely successful, and paint a bright future for the still growing field.

3. METHODOLOGY

This study is based on a dataset from Kaggle.com, published by “UCI Machine Learning” in 2016. The dataset is made up of three files, one being the relevant data and final grades of students in a math class, the next being the relevant data and final grades of students in a Portuguese language class, and the final being a merger of the two. This data is from two schools offering the same courses, the schools in question being “Gabriel Pereira” and “Mousinho da Silveira”. This study utilizes the data from the file detailing the math class. This data, originally presented in a mix of text and numeric format, has been edited in order to be better read by the program. This editing takes the form of numeric replacement. For example, one of the datapoints is the datapoint “Romantic”. This point, which details is if the student is in a romantic relationship or not, is answered with a “Yes” or “No”. These text-based answers are replaced with numbers, with “No” being replaced by “0” and “Yes” with “1”. Similar strategy is employed in other datapoints, such as parental education.

Chart 1. Datapoints Utilized by Program

Datapoint	Original Scale
School	Binary
Sex	Binary
Address	Numeric (15-22)
Family Size	Binary
Parent Cohabitation Status	Binary
Mother’s Education	Numeric (0-4)
Father’s Education	Numeric (0-4)
Mother’s Job	Nominal (5 options given)
Father’s Job	Nominal (5 options given)
Reason for Attending School	Nominal (3 options)
Student’s Guardian	Nominal (3 options)
Travel Time from Home to School	Numeric (1 - 60)
Time Spent Studying per Week	Numeric (1-10)
Number of past Class Failures	Numeric (0-4)
Extra Educational Support	Binary
Family Educational Support	Binary
Extra Paid Classes in Relevant Subject	Binary
Participation in Extra-Curricular Activities	Binary
Attended Nursery School	Binary
Wants to Take Higher Education	Binary
Internet Access at Home	Binary
Within a Romantic Relationship	Binary
Quality of Family Relationships	Numeric (1-5)
Free Time After School	Numeric (1-5)
Amount of Time Spent with Friends	Numeric (1-5)
Workday Alcohol Consumption	Numeric (1-5)
Weekend Alcohol Consumption	Numeric (1-5)
Current Health Status	Numeric (1-6)
Number of Absences	Numeric (1-93)

In order to gauge the impact the amount of data has on the accuracy of the program, three different data sets are used. The first is made up entirely of the original data, edited in order to be read easier by the program. This dataset is made up of 356 rows of data, each with 29 individual datapoints and one solution. The second dataset will be made up of the same set of data, duplicated four times in order to provide a larger collection of data for the program to work with. Finally, the third data set will be made up of the original data set duplicated sixteen times to provide, once again, a larger collection of data for the program to work with.

This data is fed into a Deep Q-Learning program, a program that is nearly entirely based on the Deep Q-Learning program found in “AI Crash Course” by Hadelin de Ponteves [4]. This code can be found at https://github.com/PacktPublishing/AI-Crash-Course. The choice to use Deep Q-Learning was based primarily on the ease in which data could be fed into the program, with further considerations towards effectiveness with low amounts of data as well as speed of execution playing a smaller role in this decision. The Deep Q-Learning program this study’s program is based on is capable of taking in information in the form of an Excel file, a near necessity given the form the data takes. The fact that this program could take in data in such an Excel file made the program itself a true standout, especially when presented with the alternatives Because of this, the effectiveness in other areas being seen as less important and had little impact on which form of AI was used. While other programs and datatypes could have undoubtably been used, given the schedule of the study, Deep Q-Learning was selected.

One of the most important parts in the processing of data is its scaling. Due to the variance in the scale of datapoints, it is necessary to scale each point into a specific range. If one were not to take this step, certain datapoints would have a larger effect on the prediction. For example, the “Absences” datapoint is scaled on a range from 0 to 93. However, the datapoint internet (A binary value representing if the student has Internet access at home) only has the range 0-1. As such, each datapoint is scaled to a range of 0 to 1. The same process is applied to the solution as well.

Deep Q-Learning is based on the idea of artificial neurons, which when combined, form a neural network. Each individual neuron takes in input signals and then releases an output signal, similar to how a biological neuron would. These output signals are known as Q-values [4]. However, the most important part of this procedure is not the inputs or outputs, but the weights assigned to the inputs. These weights tell the program how large or small of a roll the input plays, with these weights being updated by the program as it runs [4]. These weighted inputs feed into the activation function, the function that returns the outputs. The activation function can come in a variety of forms, such as the threshold function, which releases a signal only if a parameter is met, or the sigmoid function, which releases an increasingly strong output as parameter increases [4]. Deep Q-Learning programs do not exclusively use one of these functions, instead using them in different layers of the network, with artificial neurons receiving input from neurons before them, and releasing output to the neurons after. Once this data reaches the end it is compared to the correct solution, in this studies case being the final grade of the student. The difference between the two is calculated and, in a process called back-propagation, is fed back into the program to update the weights that effect the input [4]. The program first performs through what is essentially guesswork, noticing how changing the weights effects the results error. When the program notices its error decreasing it increasingly focuses on the results that have bettered its estimate, while also still trying, to a lesser extent, random weightings in order to ensure no possibilities for betterment are missed. This process is achieved by breaking the data into batches, each of which is handled a slightly separate way, once again to make sure all possible improvements take place. This process is known as “Mini-batch gradient descent” [4]. The final piece of Deep Q-Learning is called Exploration and is how the program chooses how to handle the data. Instead of having a prearranged list of ways the data is calculated, the Q-values (the outputs the program thinks may be best) are weighted based on the success the program thinks it may bring. A value is selected from this, ensuring random actions that may be beneficial are still taken while also focusing on what the program currently believes is the best solution [4]. For example, a program designed to pick the correct box may have three options, Box One, Box Two and Box Three. The program knows that picking Box One is the best current option based on past results, and gives it a Q-value of fifty. Box Two is given a thirty-five and Box Three a fifteen. It is entirely probable that Box Three may be the "right" box the highest percent of times and the program has simply been unlucky. As such, it is beneficial for the program to choose it on the possibility that it may be better than the current favorite box, even if current numbers do not favor it. To ensure this happens, the program would pick Box One fifty percent of the time, Box Two thirty-five of the time, and Box Three fifteen percent of the time. If Box Three begins to be the "right" box the most often, the program will shift to choosing it more. This way the program manages to focus on what appears to be the best solution while leaving the door open for other options. By using this process across three sets of data, the program used in this study aims to predict the grade of a student, returning the average percent error when all predictions have been completed. Each dataset will be used three times and averaged to provide a comprehensive look at how the program functions.

4. RESULTS

After the program was ran, the following results were output. Each result represents the average percent error of the dataset and does not represent the predicted score. Percent error has been rounded to the sixth decimal place.

Table 2. Results of Program

	Dataset One	Dataset Two	Dataset Three
Trial One	30.655871%	3.896241%	0.607785%
Trial Two	27.187150%	9.813397%	1.442665%
Trial Three	25.430312%	4.746702%	0.916779%
Average	27.757777%	6.118780%	0.989076%

Dataset One, made up of 356 students, returned the following results. Trial One returned a percent error of 30.655871. Trial Two returned a percent error of 27.187150. Trial Three returned a percent error of 25.430312. The average of these results was a percent error of 27.757777.

Dataset Two, made up of dataset One duplicated four times, returned these results. Trial One returned a percent error of 3.896241. Trial Two returned a percent error of 9.813397. Trial Three returned a percent error of 4.746702. The average of these results was a percent error of 6.118780.

Dataset Two, made up of Dataset One duplicated sixteen times, returned these results. Trial One returned a percent error of 0.607785. Trial Two returned a percent error of1.442665. Trial Three returned a percent error of 0.916779. The average of these results was a percent error of 0.989076.

Table 3. Difference Between Average Percent Error

Difference Between Dataset One and Two	Difference Between Dataset Two and Three	Difference Between Dataset One and Three
21.638998%	5.129703%	26.768701%

The difference between Dataset One's average percent error and Dataset Two's average percent error is 21.638998. The difference between Dataset Two's average percent error and Dataset Three's average percent error is 5.129703. The difference between Dataset One's average percent error and Dataset Three's average percent error is 26.768701.

Table 4. Ratio Between Average Percent Error

Dataset One: Dataset Two	Dataset Two: Dataset Three	Dataset One: Dataset Three
4.536489:1	6.186357:1	28.064343:1

The ratio between Dataset One's average percent error and Dataset Two's average percent error is 4.536489:1. The ratio between Dataset One's average percent error and Dataset Two's average percent error is 6.186357:1. The ratio between Dataset One's average percent error and Dataset Three's average percent error is 28.064343:1.

5. DISCUSSION

While at first glance the data paints a clear picture, by looking further one can begin to draw conclusions about how well the program functioned and how the varying amount of data may have impacted the results.

Firstly, one can see the accuracy of the program. With a limited amount of data (Dataset One), the program was still able to predict the student grade within an average accuracy of thirty percent. This result shows the possibility that a similar program could be put to use with a limited dataset, provided that consideration is given to the likelihood of error. Dataset One also demonstrated a surprisingly high level of consistency in its accuracy. In the three trials performed, no percent error was separated by more than five percent. This shows that, utilizing a limited dataset, the program is capable of reaching a relatively consistent level of accuracy. The accuracy in Dataset Two tells a similar story. Using a moderate, though still limited when compared to other programs, amount of data, the program was able to predict within an average accuracy of seven percent. However, unlike Dataset One, the difference in between the predictions, when compared with the scale of the data, is much larger. The largest difference in Dataset One was between Trial One and Trial Three and was 5.225559, 18.82557 percent of the average percent error in Dataset One. The largest difference in Dataset Two is 5.917156, 96.704833 percent of the average percent error in Dataset Two. This difference is likely caused by what appears to be an outlier, Trial Two, and the fact that any difference is magnified due to the decreasing size of the percent error. Further testing would be required to identify the source of this difference with confidence. Finally, when using a larger dataset, the program was able to predict within an average accuracy of one percent, with the largest difference between two trials being .525886 (between Trial One and Trial Two), 36.525989 percent of the average percent error in Dataset Three. This level of accuracy shows the use a similar program may have when used in education. For example, a school using a program such as the one utilized in this study would be able to predict how students would perform based on student answers to a 29-question survey. The results of this survey would allow the school to focus greater efforts on students who are predicted to earn a lower final grade and, in some cases, intervene in order to rectify situations which may lead to a lower grade.

Asides from the percent error itself, one may also draw conclusions based on the ratios between datasets. Between Dataset One and Dataset two, the average percent error experienced a decrease by a factor of 4.536489. This figure roughly lines up with a linear progression as the amount of data increases. In this case, for every four instances of data, the accuracy increases four times (an average increase of 1.134122). This would lead one to believe that Dataset Three would lead to a further increase in accuracy by a factor of twelve. However, when one looks at the ratio between Dataset One and Dataset Three, there is an increase by a factor of 28.064343. In this dataset, each instance of data leads to an average increase in accuracy by a factor of 1.754021. This means that for each new instance of data in Dataset Three, the accuracy increased by a factor of 1.960654. This increase could be a result of multiple factors. Firstly, it is possible that the program's average percent error is lower due to simply having more possible guesses. It is entirely possible that the program figures out which datapoints have the largest impact and begins to guess accordingly within only four datasets, yet the average margin of error is unable to change greatly due to the number of opportunities to guess running out. In this case, having more opportunities to guess lets the program perform more accurate guesses for a greater margin of its total attempts, unfairly weighting the later guesses. Secondly, the program could more accurately figure out the datapoints which play a greatest role in the final twelve datasets, causing these guesses to be more accurate and thus appear to increase in accuracy faster than the first twelve datasets. Finally, there is a chance that the duplication of datasets plays a nonnegligible role in the program's accuracy. In this case, the program would learn to recognize individual patterns that repeat and guess based on those patterns instead of an overall pattern. This case is believed to be less likely due to the relatively small amount of duplication. Given the fact data is only repeated sixteen times, the program would be hard pressed to identify specific patterns, however there still remains a possibility it could be able to.

Another point to consider in the results is how the program reaches the level of accuracy it does. In the case of Dataset One, with only 246 individual opportunities to guess, the program is able to achieve an average percent error of 27.757777. When compared to accuracy with greater data, this number seems unimpressive, however when one considers the amount of data the program is able to use, it becomes clear that a percent error in the range of the thirties is no small feat. This accuracy can once again be explained in a number of ways. Firstly, is the number of high impact datapoints. These datapoints have an immediately obvious effect on the result and thus are likely to be quickly identified as important by the program and quickly taken into account. Examples of such datapoints include "studytime" (the time weekly time allotted to studying), "schoolsup" (whether or not the student has extra education support), and "paid" (whether or not the student attends extra paid classes in the course subject). Other similar studies have encountered this effect as well. Take for instance the Greece University of Petras study noted in the article "Can Educational Data Mining Predict Student Performance and Enhance Personalized Learning?" [3]. The university's program performed similarly to the one detailed in this study, reaching between seventy and eighty percent accuracy even with exceedingly limited data. This level of accuracy is not common, with the article detailing how a similar study done with the goal of predicting repeat offenses in criminal offenders only was able to achieve a sixty-five percent accuracy even when given as many as 127 attributes [3]. The conclusion was the university reached was similar to the one made by this study: the type of variables incorporated matters a lot, perhaps more than the amount of data used.

Figure 1. Impact of Study Time on Final Grade

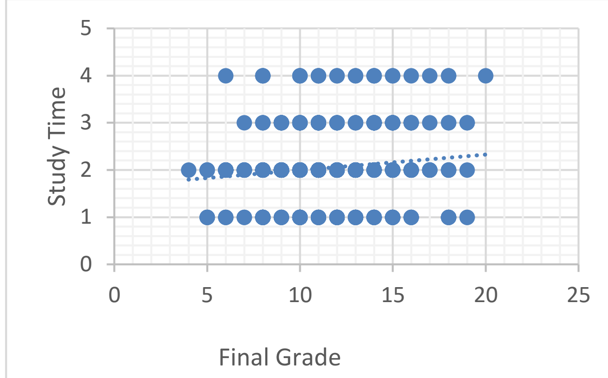


Figure 2. Impact of Educational Support on Final Grade

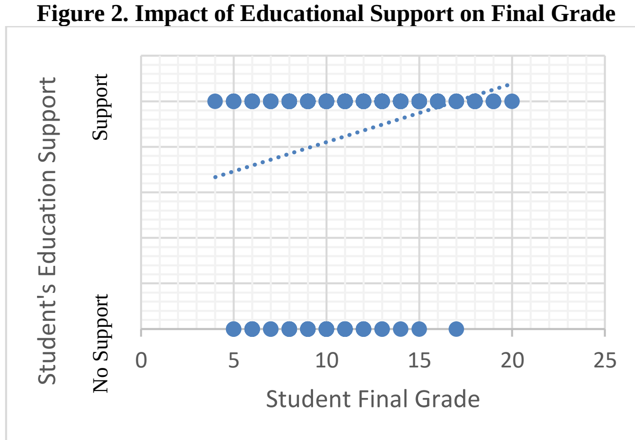
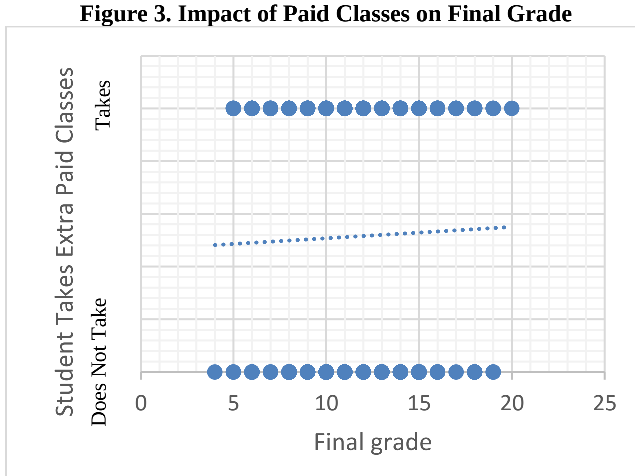


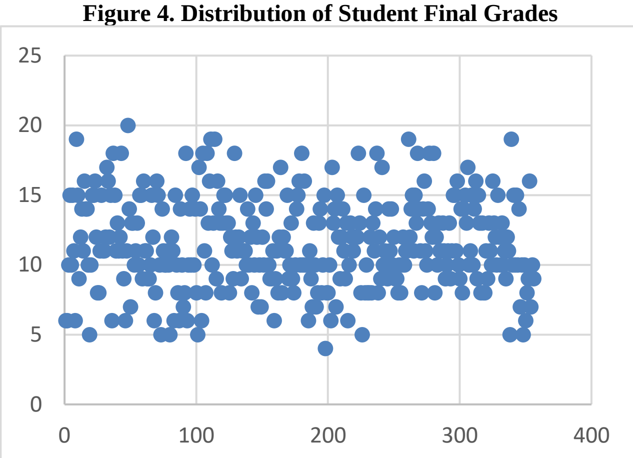
Figure 3. Impact of Paid Classes on Final Grade



Secondly is the high number of binary datapoints. Twelve out of the twenty-nine datapoints only have two answers; represented as either a "0" or "1" in the data. These exact numbers are likely easier for the program to understand and take into account. For instance, the numeric datapoint "Medu" represents the mother's education. In this case there are four possible values for this variable, ranging from uneducated to higher education. The computer must figure out not only how relevant this point is, but how each value would impact the student's grade. Is a highly educated mother more likely to raise a child who receives higher grades? Is an uneducated mother more likely to want their children to become educated? Compare these questions to the ones the program must figure out about binary datapoints. First it must figure out how relevant the datapoint is. Secondly it asks what value the variable is. While this example is vastly simplified, it makes clear that the presence of many binary datapoints is a factor in the accuracy the program displays.

Finally, one must consider the grades themselves. As a result of how grades are assigned, most grades fall into a common range. In this dataset, no student receives a grade lower than a 4, with most receiving a grade in the range of ten to fifteen.

Figure 4. Distribution of Student Final Grades



This arrangement of grades makes it easy for the program to, to a degree of accuracy, predict a student's grade even if no relevant data exists. This fact may play the largest role in the accuracy of the program, simply due to the fact that the program can virtually eliminate seventy-five percent of all guesses unless there are incredibly strong signs to the contrary. This fact is also specifically useful in Datapoint One. While other, larger datasets have enough data so that the program can figure out the datapoints that play the largest roles, the lack of data in Dataset One does not allow for this. However, instead of floundering and guessing at random, the program is able to make a guess between ten and fifteen and achieve a fairly low percent error, all while it improves it consideration of other variables. In other datasets this doesn't matter as much due to the fact that the program is able to figure which datapoints matters and do not need to rely on guesswork as much.

The low amount of unduplicated data is one of the main limiting factors in this study. Given a greater dataset, a more complete conclusion would be able to be gathered and likely paint a slightly different picture than the one painted in this study. A second limiting factor is the low number of trials. A greater number of trials would likely help identify outliers in current data, as well as create a more accurate average. A greater of trials would also likely help in identifying the average difference in percent error across trials, an average unable to be fully realized with current results. The low number of datasets also limits the conclusions able to be made. In using a greater range of datasets (perhaps datasets made up of eight and twelve duplicated sets of data), the study would be able to better identify points in which accuracy begins to increase, leading to a more comprehensive picture of what impacts the program's results.

6. CONCLUSION

This study, while undoubtably providing clear answers to a number of questions, manages to open more doors than it closes. One can clearly see that the program is capable of providing accurate predictions given a varying level of data as well that the amount of data given to the program greatly impacts the total accuracy. These conclusions could possibly help expand the understanding of AI using low amounts of data, as well as help to show that artificial is capable of functioning even when given only a fraction of what would be considered a "typical" amount of information. Yet these conclusions are, for the most part, already explored and understood. This study does however, create questions to be answered. One question that sticks out is the increase in the ratio of accuracy to data. Future studies may wish to explore this phenomenon, using a more complete range of datasets in order to find when and why this increase exists. A second notable question is the impact the duplication of data has on the program's results. Future studies may explore the difference between a dataset that is wholly unique and one that is made up of a duplicated set of information. Finally, impact of binary datapoints is also a point that may deserve further research. What is the program's accuracy when given a dataset made up of entirely numeric datapoints? How does this compare to a dataset made up of a mix of datapoints? A dataset made out entirely binary datapoints?

A more complete study may also be performed with a much greater sized dataset. While not entirely limiting in scope, the dataset used in this study constricts the conclusions able to made and may be subject to error. A dataset of greater length would likely prove to be a more accurate representation of the program's abilities. Finally, one must note the situation of the study itself. Due to time constraints, a more complete study was unable to be performed. A more complete study would likely include a greater number of trials as well as a greater array of datasets, resulting in a more conclusive study.

7. ACKNOWLEDGMENTS

Those involved with the study would like to thank Dr. Ruthie Halma for her guidance and excitement towards the study. Those involved in the study would also like to thank Mr. Connor McLaughlin for his help in utilizing the ACM format. Finally, those involved would like to thank Mr. Steven Loutzenhiser for his interest in programming and help throughout the years.

8. REFERENCES

[1] Matthew Lynch 390. 2019. Using Machine Learning to Predict Student Outcomes. (May 2019). Retrieved November 12, 2020 from https://www.thetechedvocate.org/using-machine-learning-to-predict-student-outcomes/

[2] Thomas Miller and Melissa Irvin. 2019. Using Artificial Intelligence with Human Intelligence for Student Success. (December 2019). Retrieved November 12, 2020 from https://er.educause.edu/articles/2019/12/using-artificial-intelligence-with-human-intelligence-for-student-success

[3] Olivia Piontek, Jessica Poiner, Erika Sanzi, and Chad L. Aldis. 2019. Can educational data mining predict student performance and enhance personalized learning? (August 2019). Retrieved November 12, 2020 from https://fordhaminstitute.org/national/commentary/can-educational-data-mining-predict-student-performance-and-enhance

[4] Hadelin de Ponteves. 2019. *AI Crash Course.* 35 Livery Street, Birmingham B3 2PB, UK.

[5] Juan L. Rastrollo-Guerrero, Juan A. Gomez-Pulido, and Arturo undefined Duran-Dominguez. 2019. *Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review* (November 2019).

[6] Huaxiu Yao, Defu Lian, Yi Cao, Yifan Wu, and Tao Zhou. 2019. Predicting Academic Performance for College Students. *ACM Transactions on Intelligent Systems and Technology* 10, 3 (2019), 1–21. DOI:http://dx.doi.org/10.1145/3299087