Use of neural networks to analyze and predict the efficacy of learning accommodations for special education students

Nihal Gulati

September 2021

Abstract

This research gauges the usability of artificial neural networks as a method of data analysis in the efficacy of learning accommodations for special education students, using a simulated student data set. The ANN achieved a high accuracy in data predictions, and was able to determine the most effective accommodations per individual students. Data size played a large role in ANN accuracy, with a predicted 20,000 student records giving an acceptable model accuracy. These results are still inconclusive because of the simulated nature of the data set, further research using real student data is needed to gauge true efficacy.

1 Introduction

1.1 Basic Overview of Special Education

Since the 1970s, federal law has required public elementary and secondary schools to provide special education services to students with disabilities. It began with the Education of all Handicapped Children Act of 1975, as the first legislation that protected the educational rights of students with disabilities. This law was replaced and extended by the Individuals with Disabilities Education (IDEA) Act in 1990 and 2004. IDEA's overall purpose is the same as the first act: it guarantees educational rights to all students with disabilities and makes it illegal for school districts to refuse to educate a student based on a student's disability.[2]

To accomplish that goal of a fair education, IDEA also outlines the path for students to receive special education services. Per California's Legislative Analyst's Office (LAO) office report, parents or teachers typically are the first ones to identify if a student might benefit from special education services. In most cases, children are then referred to school district specialists, who evaluate whether the student has a disability that interferes with his or her ability to learn.

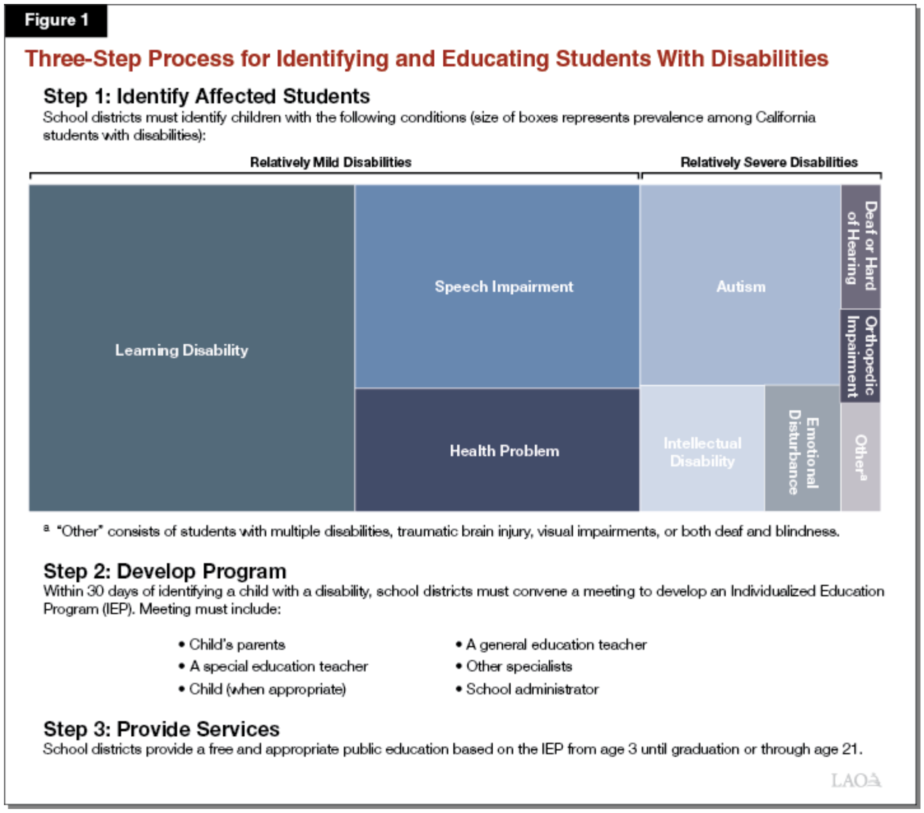

Figure 1: LAO Overview of Special Education Process [1]

If determined to have one or more such disabilities, the student receives an individualized education program (IEP), based on the student's needs, that sets forth the additional services the school will provide. The IEP is developed by a team consisting of each student's parents, teachers, and district administrators, often consulting with special education professionals or specialists.

The IEP may include various types of special education services, such as specialized academic instruction, speech therapy, physical therapy, counseling, or behavioral intervention. [1] An IEP would also include learning accommodations, if any, that would help the student achieve their personal goals listed in the IEP. These accommodations can include anything from extra time on assessments to full separation in exclusive specialized classrooms or schools.

This process has long required those specially-trained professionals, along with parents, to make decisions about how best to help the needs of students with disabilities. Such specialists have to make decisions about unique individual needs based on their own knowledge and prior experiences.

While this system has proven effective in helping students achieve their individual goals, recent advancements of data-driven analysis and modeling, if

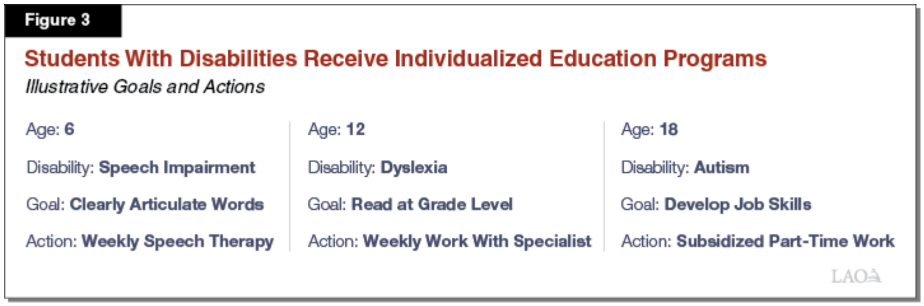

Figure 2: LAO Overview of Special Education [1]

applied to this area, could greatly improve the confidence and ability of learning professionals to help students by analyzing school data and providing them clear statistical success metrics and indicators of effective aid techniques.

Right now, comprehensive data on learning accommodations and their effectiveness, even clear experimental data, is not available at the large level to professionals, limiting them from the benefits analysis of student data could provide.

The compiling of large data sets on students of various demographics, needs, and accommodations and their relative performances would beget a large change in this field. Again, no such large data sets are publicly available. Student learning accommodations as a whole are not commonly tracked by school districts and states. If schools have the information, it is most likely in the form of the IEP, which is limited in that any accommodations listed are part of a general plan which does not include when accommodations were actually given and other specifics. An IEP also exists purely as unstructured data, making it hard to normalize and compile. Private data that school districts or testing vendors possess, IEPs or structured data or otherwise, may also be subject to student privacy concerns that provide ethical barriers to public availability and detailed outside analysis. Future analysis in this field will likely be done only in deep conjunction with education agencies because of these limitations.

This research provides a basis on future data analysis in this field and proves methods that may be effective. Specifically, the focus is on the use of artificial neural networks to train and predict data sets on student attributes, accommodations given, and performance metrics to help identify successful practices.

1.2 Brief Overview of Neural Networks

To give an overview and explanation of what artificial neural networks are for the unfamiliar, an artificial neural network (ANN) is a method of data modelling or prediction based on the human brain, imitating the way that humans think as a 'net' of neurons firing. ANNs are the central part of the field of machine learning and the basis for all other deep learning algorithms. [3]

Functionally, an ANN is composed of layers of 'neurons', called nodes, with an input layer for the data, multiple hidden layers for processing, and an output layer. Each node in a layer connects to every node in the preceding and subsequent layers with an associated weight (multiplier) and threshold, or bias. A single node takes each of the inputs received by the previous layer, multiplies them by their weights, and sums them together. If the sum of inputs to a given node are below its associated threshold, it does not fire and the next layer receives an input of zero from that node. If it does meet the threshold, the next layer receives a one, assuming the simplest neuron activation function for convenience. This propagates through the network until the output is received. [3]

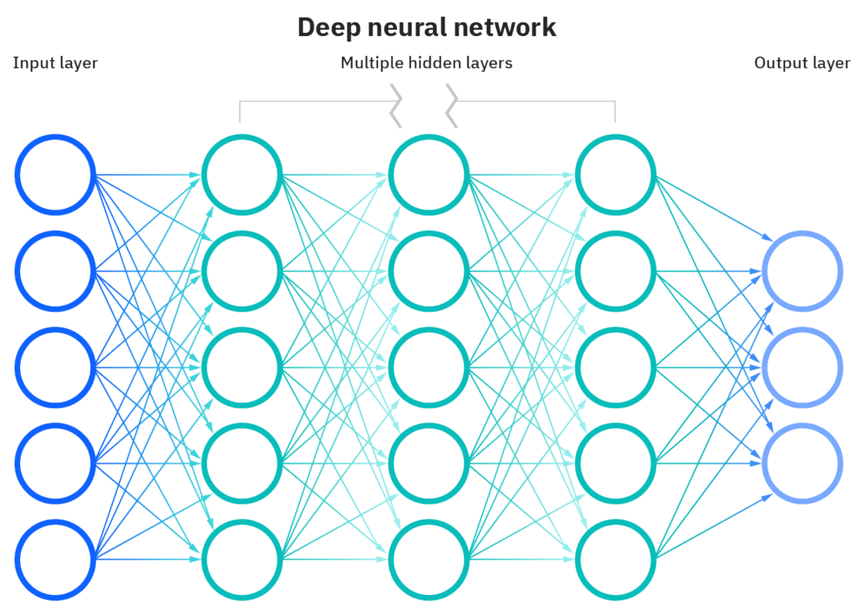

Figure 3: Diagram of Multi-Layer Artificial Neural Network [3]

In this way, neural nets are well built to comprehend complex patterns. Weights can essentially represent the importance of a given input, and then individual nodes in the net can focus on identifying particular factors which then sum later in the net to provide a cohesive output.

As an illustrating example, we can use the question of “Is it raining right now?” which a simple one-node neural network, known as a perceptron, can answer with the inputs of “Is the floor wet?”, “Are people carrying umbrellas?”, and “Is it cloudy?”

Each of the inputs receives a weight. Perhaps the ‘cloudy’ input will have the biggest weight, since in the case that it is not cloudy, the chances of rain are almost zero, and the “wet floor” could have the smallest weight since it is the most ambiguous in determining whether it is raining. These inputs, multiplied by the weights, would then be summed and considered against the threshold. Since any single factor is not enough to say that it is raining, the threshold is likely big enough to require any two of the inputs to be true before the node fires, declaring that it is raining. In this manner, even a simple one-node network can be a useful prediction algorithm.

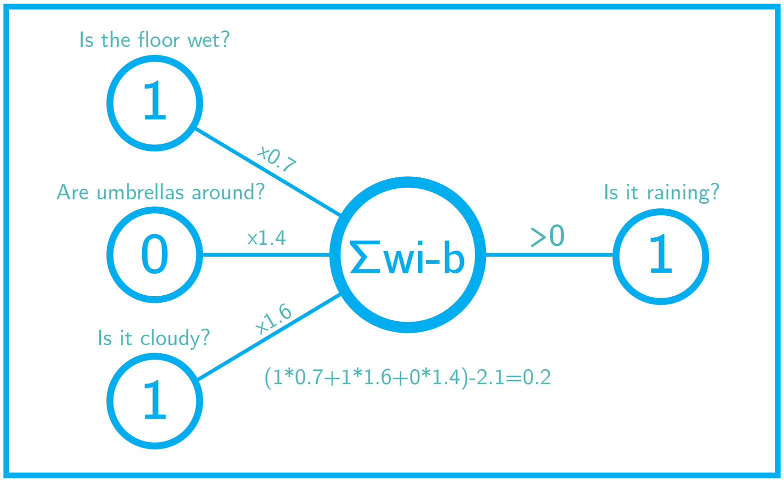

Figure 4: Perceptron Model For Basic Determinations

With large multi-node networks, larger inputs and more complex patterns can be considered. Using the previous example, a set of nodes might be dedicated to considering whether the floor is wet because of rain or because of other input factors, like a spill. A given neuron in this set might only 'focus' on the input of 'wet floor' and 'are we indoors?' If both were true, the neuron might not fire, reducing the odds that the neural net finds rain.

A central question has not yet been answered. How does a neural network change its given weights and thresholds to accurately predict the data? ANNs rely on data, training data to learn and improve accuracy over time. An ANN is initialized with randomized and near-zero weights and biases, and then learns and adjusts its weights and biases with training data. Generally, the larger the data set it is trained on, the more accurate the neural network at its task. The question of how exactly a neural network adjusts its weights to fit training data requires complex math to answer and is beyond the scope of this introduction. Information on this, on back-propagation and stochastic gradient descent, is referenced below. [11]

Once a neural network is trained, they can classify, predict, and cluster data at speeds far beyond traditional statistical methods. Neural networks are also the best method for difficult computer tasks like image and speech recognition, and can even, in the form of reinforcement learning, play video games or execute complex motor actions like driving or walking.

1.3 Application of ANNs in Special Education

ANNs are well suited to data analysis in the field of special education compared to traditional statistical methods because of their overall high complexity and adaptiveness. Given adequate data and training time, ANN's ability to excel in analyzing large, multidimensional data sets where connections may not be clear such as student data, far surpasses regular analysis, with the possible downside of non-deterministic behavior.

ANNs would be able to finely analyze basic data like gradebook information and learning accommodations, seeing which accommodations have been most and least effective for helping students depending on student attributes. Such analysis would provide direct helpful feedback to education professionals and inform their future actions to assist students.

It may also be combined with a recommendation system. A recommendation system together with the ANN could generate predictions or suggestions on effective learning accommodations for individual students based on its previous data. Such a tool could be an immense additional aid to education professionals in their decisions by giving them data-driven, algorithmic suggestions to aid students.

This work, in lieu of real student data, aims to gauge the feasibility of data analysis using ANNs in this field by the use of simulated student data sets to provide a foundation for future research in education.

2 Methods

2.1 ANN Network Creation

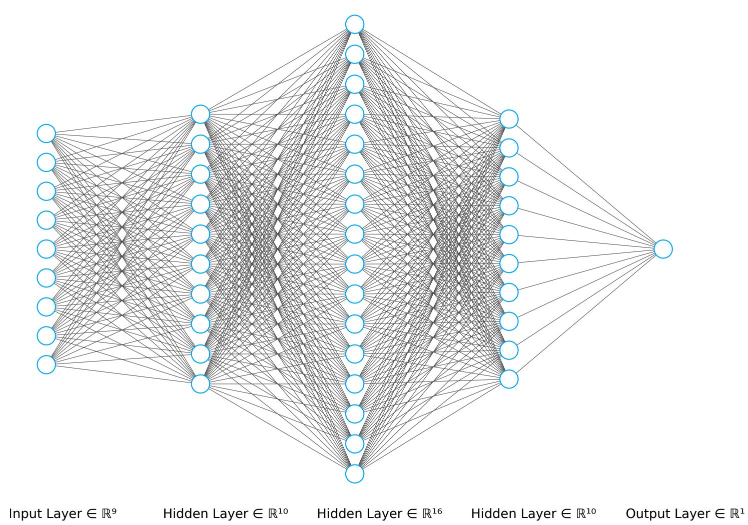

Figure 5: ANN Node Diagram, One-Third-Size Layers For Viewability [4]

The artificial neural network model was made in Python script with the use of TensorFlow's Keras libraries. TensorFlow is a common library in the field of machine learning. Along with the Numpy library for matrix and tensor manipulation, programs can make use of TensorFlow Keras's built-in neural network models.

The model used for prediction in this case made use of Keras's Sequential system base with Keras Dense neuron layers. The architecture chosen had five total network layers. First, the input layer, sized in this case to accept a 1D tensor of size 29 in accordance with the data input, then, three hidden layers with rectilinear, ReLu, activation of 50, 30, and 50 nodes each, and finally, a one-node output layer using the linear activation function. The model was compiled using the "adam" optimizer and using the "mean squared error" as the data loss function. For an illustration of the architecture of the model, with layers one-third the size for reasonable display, refer to Figure 5 above.

2.2 Simulated Data Set Creation

In place of real student data, artificially generated student data was used as a proof of concept. A particular student data row contained 3 quantitative variables and 4 qualitative, plus a data identifier in the form of a randomized student ID for a total of 8 variables in the data set, as shown in the example data set in Figure 6 below.

Out[5]:

	studentIDs	gender	age	teacher_cred	class_size	disability	accomadation	gpadifference
0	17483	Male	16	PhD	32	Autism	Materials in Braille	1.672031
1	69971	Male	15	Associate's	39	Mathematics Disability	Tutoring Sessions	0.809299
2	82568	Female	13	PhD	23	Speaking Disability	Bigger Print Materials	0.056877
3	106957	Female	16	Associate's	38	Autism	Use of Toy in Class	0.558986
4	44373	Female	8	PhD	28	Speaking Disability	Book Buddy	1.637078
...	...	...	...	...	...	...	...	...
49995	97778	Male	9	Master's	31	Dyslexia	AAC Devices	-0.004897
49996	52780	Male	10	PhD	35	Reading Disability	Materials in Braille	-0.250446
49997	35250	Female	16	PhD	23	Speaking Disability	Special Education Classroom	2.395113
49998	44684	Male	10	Master's	37	Visual Disability	Use of Calculator on Tests	-1.484255
49999	80703	Female	10	Bachelor's	37	Low Emotional Intelligence	Text to Speech Devices	0.555461

50000 rows × 8 columns

Figure 6: Example Student Data Set

The specific student variables as shown in Fig. 6, chosen randomly in an approximation of possible real data, are as follows:

- Student gender was chosen randomly from either male or female.
- Student age was a randomly chosen integer from 6 to 18, representing the common age range in K-12 education.

- Teacher credentials were randomly chosen from an Associate's degree, Bachelor's degree, Master's degree, and PhD.
- Class sizes were picked from a simulated random distribution with mean 30 and a standard deviation of 5.
- Student disabilities were randomly picked from 10 possibilities, "Auditory Disability", "Visual Disability", "Autism", "Low Emotional Intelligence", "Dyslexia", "ADHD", "Down Syndrome", "Reading Disability", "Mathematics Disability", "Speaking Disability", and "Developmentally Delayed".
- Learning accommodations given were also randomly picked from 10 possibilities, "Materials in Braille", "Text to Speech Devices", "Breakout Corner", "Use of Toy in Class", "Bigger Print Materials", "Isolated Workstation", "Tutoring Sessions", "Book Buddy", "Use of Calculator on Tests", "AAC Devices", and "Special Education Classroom".

Meaning cannot be assigned to the particular choices of accommodations or disabilities listed in the data, since labels were chosen in an indiscriminate fashion.

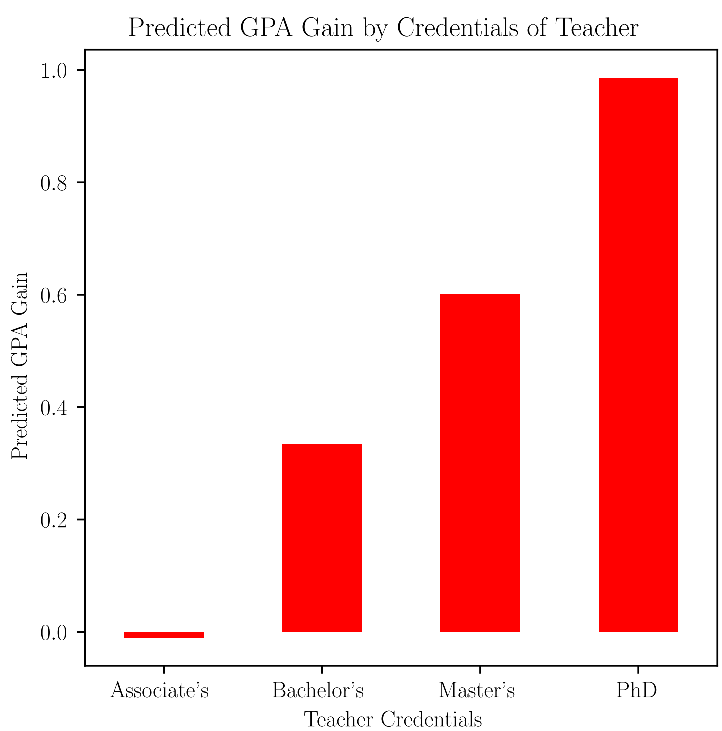

Figure 7: Average Effect of Teacher Credentials on GPA Variable

Finally, the predictor of student success used, the response variable, was the simulated GPA difference between a grade quarter and its preceding quarter. The GPA difference variable was generated in relation to the other data variables, as to place a simulated relationships for the model to find. GPA difference was generated with a linear negative correlation to class size, a positive discrete correlation to teacher credentials (illustrated in Figure 7), a combined addend factor of age and gender, and a more complex addend factor linking disabilities with particular accommodations in a linear format. Values were then given a purely random addend to add simulated noise to the data and standardized, plus an offset, to the 4.0 GPA scale. For more data on the exact calculations used, view the published code repository. [27]

To use a particular example, a dyslexic female student of 10 years old, with a class size or 21, a teacher with a PhD, and given the "tutoring sessions" accommodation would have had the GPA difference generated with a factor of -0.05 for class size, +1.5 for teacher credentials, +0.6 for her age and gender, +0.5 for her disability in relation to accommodation given, and a random factor for an overall GPA difference value of 2.12. This would then be z-standardized with the rest of the scores to a distribution with a standard deviation of 1 and mean of 0.4. The overall distribution can be viewed in Figure 8.

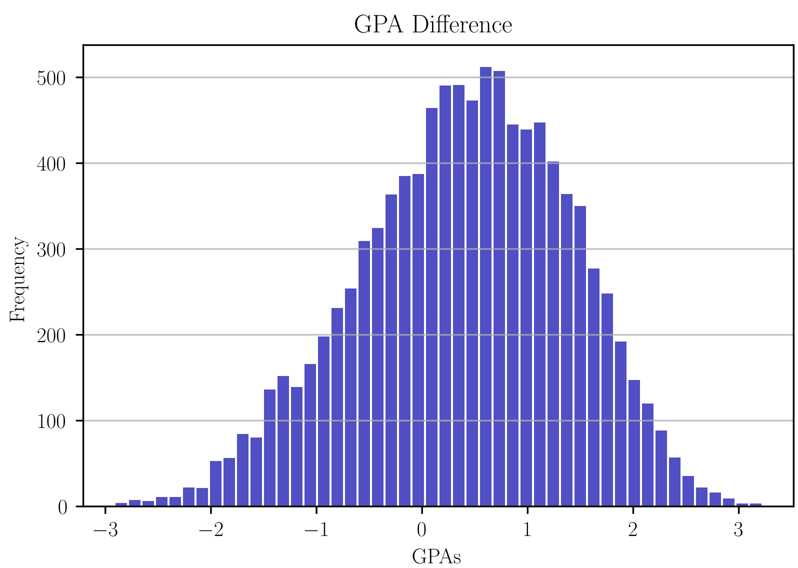

Figure 8: Distribution of Generated GPA Differences

To repeat, meaning cannot be assigned to the particular choices of multipliers for the GPA difference, the impact of different student factors was chosen simply to provide data patterns.

2.3 Data Pre-Processing and Experiment Setup

The student data was generated in five different sizes for different trials, with 2500 student data rows, 5000 rows, 10000 rows, 20000 rows, and 50000 rows. Once generated, libraries from Sci-Kit learn were used to process and package the data for neural network consumption. Gender was encoded using 0s and 1s for male and female, the teacher credentials, disability, and accommodation variables were all one-hot encoded into multiple binary 1-0 rows, one for each possible variable value, and the student IDs were dropped from the data.

All 29 data values were then normalized by sklearn's MinMaxScalar function, which scales all values to between 0 and 1 for easier consumption by the neural network. GPA differences were split from the main data set and prepared separately for the model as the response variable to be predicted by the model.

```
Out[32]: age                                          6.000000
         class_size                                  37.000000
         gpadifference                               -1.604922
         gender_Female                                0.000000
         gender_Male                                  1.000000
         teacher_cred_Associate's                     1.000000
         teacher_cred_Bachelor's                      0.000000
         teacher_cred_Master's                        0.000000
         teacher_cred_PhD                             0.000000
         disability_ADHD                              0.000000
         disability_Auditory Disability               0.000000
         disability_Autism                            0.000000
         disability_Developmentally Delayed           1.000000
         disability_Down Syndrome                     0.000000
         disability_Dyslexia                          0.000000
         disability_Low Emotional Intelligence        0.000000
         disability_Mathematics Disability            0.000000
         disability_Reading Disability                0.000000
         disability_Speaking Disability               0.000000
         disability_Visual Disability                 0.000000
         accomadation_AAC Devices                     0.000000
         accomadation_Bigger Print Materials          0.000000
         accomadation_Book Buddy                      0.000000
         accomadation_Breakout Corner                 1.000000
         accomadation_Isolated Workstation            0.000000
         accomadation_Materials in Braille            0.000000
         accomadation_Special Education Classroom     0.000000
         accomadation_Text to Speech Devices          0.000000
         accomadation_Tutoring Sessions               0.000000
         accomadation_Use of Calculator on Tests      0.000000
         accomadation_Use of Toy in Class             0.000000
         Name: 0, dtype: float64
```

Figure 9: Encoded Data Sample, Before MinMaxScaling

Data was separated under the standard 20/80 test-train split; with 80% of records going to model training and 20% partitioned for testing and evaluation of the model. The ANN model was fitted on the training data with a data batch size of 32 and for a total training length of 25, 50, 100, and 200 epochs in different trials.

Each configuration of a given training length and data size, ex. 50 epochs and 20000 student records, was regenerated, trained, and tested a total of 50 times for repeatable and statistically significant results.

Results were obtained using the standard neural model predict method. The

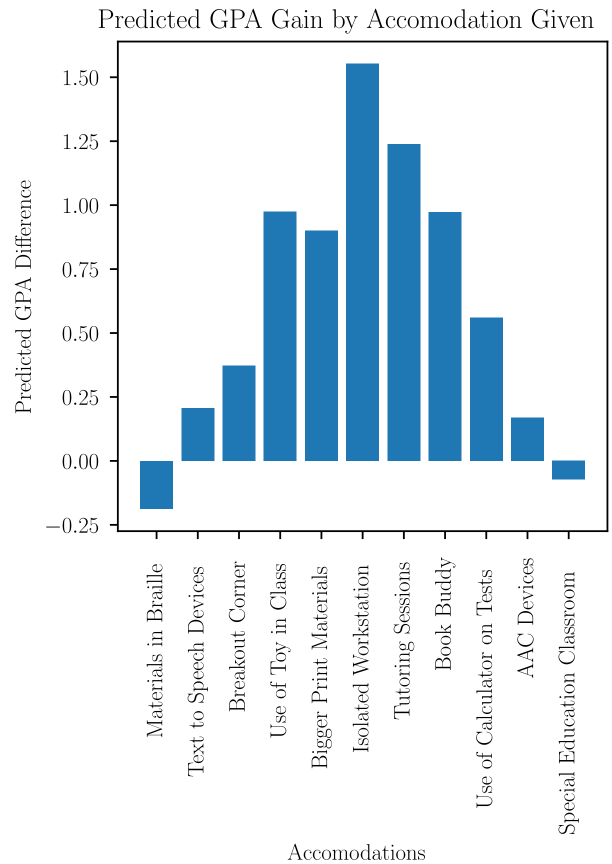

Figure 10: Model Calculation of Best Accommodation for 15-year-old Female with ADHD, Class Size of 28 Students, and Teacher with Bachelor's Degree

model was evaluated on the test set and the mean squared prediction error and explained variance were calculated and recorded.

To establish the feasibility of individual calculation, a single data row was also fed into the ANN model and an estimated GPA difference was calculated. With the use of a simple iteration, the accommodation with the highest predicted GPA difference for a particular student data point was found and graphed by the model. An example on this individual calculation for a student (inputted data: female, 15 years old, class size 28, teacher credentials of Bachelor's, ADHD) can be found in Figure 10. Note, the various performances of each accommodation was programmed in randomly into the generated data set.

3 Results

Reported below is the calculated mean squared error and explained variance of the ANN's predictions given a configuration of model parameters and number of data rows provided, to four significant digits. Each configuration was run 50 times and the numbers below are the average of those 50 iterations for each configuration. The Python code used for the experiment is published online. [citation]

The mean squared error reported represents the 'average' error in a predicted GPA Difference calculated by the model to the actual GPA Difference, so a lower error represents a better data fit and higher accuracy.

The explained variance reported represents the amount of variations or fluctuations in the GPA Difference variable that the model is able to predict, or explain. In a perfect linear correlation, for example, a properly-trained model would have 100% explained variance. With this data, because of the random offset, there will always be an unexplained variance, as in a real data set. An estimated 90% is the upper limit for model explained variance in this case.

Table 1: Averaged Mean Squared Error for All Configurations

Mean Squared Error, Less is Better (GPA difference points)					
Number of Epochs	Number of Student Data Rows				
	2500	5000	10000	20000	50000
25	0.2034	0.1813	0.1612	0.1512	0.1381
50	0.2124	0.1883	0.1650	0.1510	0.1366
100	0.2435	0.2059	0.1754	0.1536	0.1373
200	0.2846	0.2376	0.1905	0.1607	0.1384

Table 2: Averaged Explained Variances for All Configurations

Model Explained Variances, Higher is Better (percent)					
Number of Epochs	Number of Student Data Rows				
	2500	5000	10000	20000	50000
25	80.02%	82.62%	84.09%	85.07%	86.37%
50	78.88%	81.48%	83.61%	85.16%	86.53%
100	75.62%	79.66%	82.61%	84.76%	86.43%
200	71.77%	76.45%	81.10%	84.14%	86.27%

The ANN model was able to easily predict the response variable, the GPA difference, to a high accuracy on all of the configurations, with explained variances between 71.77% and 86.53% and mean squared errors between 0.2846 to 0.1366 GPA points between configurations.

The mean squared error tends to decrease and the explained variance to increase as the amount of data increases, and in this data, the ANN's predictions are always more accurate given more student records. This trend is present on all numbers of epoch trained.

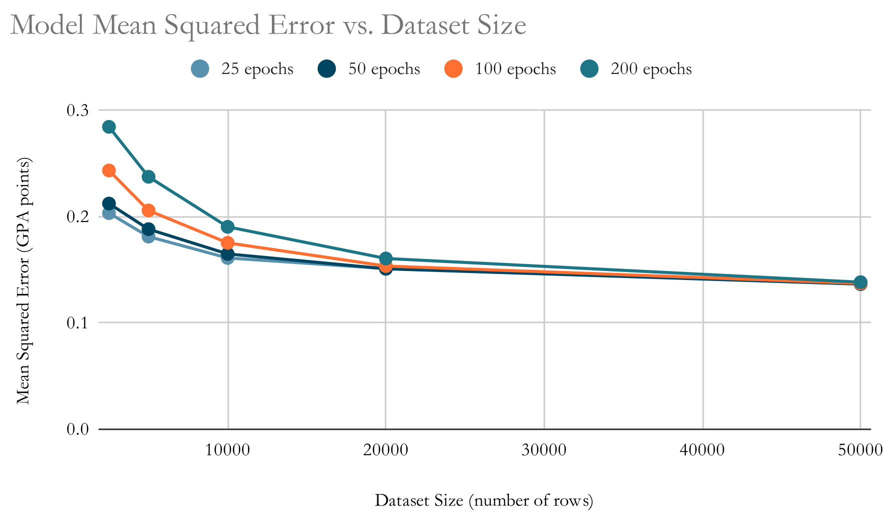

Figure 11: Model Mean Squared Error Vs. Dataset Size

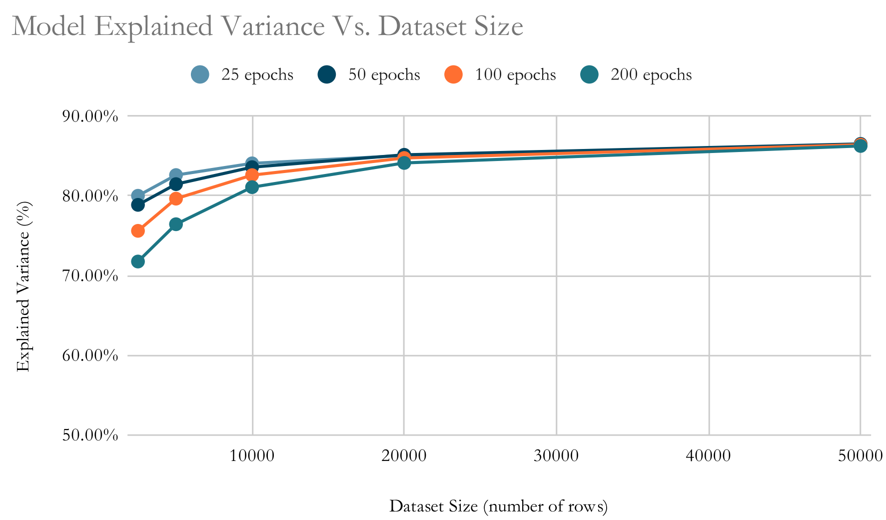

Figure 12: Model Explained Variance Vs. Dataset Size

Going from the smallest to largest data set size, 2500 to 50000 records, gave anywhere from a 6.35% boost in the explained variance from 80.02% to 86.37% to a 14.5% boost from 71.775% to 86.27% and a 0.0653 improvement in mean squared error from 0.2034 to 0.1381 GPA points to a 0.1462 improvement from 0.2846 to 0.1384 GPA points, depending on number of epochs trained.

That relationship with model success and the number of epochs trained was more complicated, with an increase in epochs of training initially decreasing explained variance and increasing error, so decreasing overall prediction accuracy, but leveling out as the data size increased.

At the maximum data set size, 50000 records, there was no noticeable difference in explained variance or mean squared error with respect to the number of epochs trained, presenting explained variances around 86.4% and errors around 0.1375.

At 2500 student records, the least amount tested, there was a 8% gap in explained variance between epoch numbers, with 25 epochs having an explained variance of 80.02% and 200 epochs having an explained variance of 71.77%. For mean squared error, 25 epochs had an error of 0.2034 and 200 epochs had an error of 0.2846. Overall, less epochs trended toward higher model success at 2500 records.

4 Discussion

The ANN was able to easily predict the data set to a reasonable accuracy on all of the configurations. The only limiting factor in its predictions was the randomized offset, without it, the ANN would have likely been able to achieve close to 100

Unfortunately, this overall result is of limited value. Likely, the simulated nature of the data set and its hard-coded patterns were altogether too simple to present much of a challenge to the ANN model. This would not hold true for a real-life student data set with thousands of confounding variables and realistically complicated data patterns. Testing this system on a real education data set is the logical next step and of utmost importance to determine the feasibility of this kind of data analysis in future education research.

Some more specific conclusions can be drawn. As expected, the amount of student data trained on was a large factor in the variation between the model's results. What was not entirely expected, however, was how much data would be needed. Even with the simplistic patterns in the simulated data sets that were relatively easy for the model to predict, improvements in the model's results were still seen upwards of 20000 student records, to 50000 and beyond.

This suggests that this number of student data points needed for accurate real-life analysis will also need to be close to this number to have a reasonable amount of the true explained variance caught by the model's explained variance and have an acceptably successful model, taking into account the much larger complexity of the data.

Below 20000 student records, the model will still return results, but the explained variance in the model will likely be lower than the true explained variance of the data set and accuracy will be impacted. Ideally, upwards of 50000 student records would allow the model to learn the full picture and achieve the highest theoretical accuracy, but this is a much harder data set to obtain given the limitations in the size of student populations in school districts.

Unexpectedly, this same positive correlation did not hold true for the number of epochs trained. Increasing the training time tended to decrease model success overall. This was most evident at the smallest data set size, 2500 records, and not evident at all on the largest set, 50000 records.

The likely culprit of this reversed correlation is overfitting. With a smaller data set of 2500 records and a simplistic enough pattern, the model can overtrain on the training data set, learning to discriminate for that set's particular generation. This model would receive a high success score on the training set, but when evaluated on the test set it would not be able to generalize as well as a non-overfitted model and would receive a lower score. More epochs trained would exacerbate the overfitting problem.

Likely, this may not be a problem in real life data sets, especially if the student record sizes are larger and more complex. Additionally, there are many well-documented ways to reduce model overfit, for example, by adding dropout layers to the neural network or tuning the number of epochs trained to maximize model success. Compared to the problem of obtaining a large enough student data set, dealing with increasing the training time of the model or preventing overfit is trivial.

Going back to the guiding question of this work, whether using ANN systems like this model to analyse student learning accommodations is feasible, the results are not definitive.

The success of the model on this simulated data set suggests that it can, in fact, be used to a reasonable accuracy on real data, but this is an extrapolated conclusion and may not be entirely valid.

Either way, this research makes clear that using ANNs and machine learning on student data is a topic that should be urgently explored with real data to further determine the usability of such a solution, and also determines a likely number of student data records, around 20000, for such analysis to be effective.

If successful, defining success as an acceptable model accuracy, these machine learning models could determine the effectiveness of particular learning accommodations in helping students with disabilities, and could even provide data-based student-individual predictions for effective learning accommodations. As referenced in the introduction, this would be an effective aid for education specialists that generally do not have access to data that could assist their decisions, and push us even closer to IDEA's goal of an equal education for all, regardless of disability.

5 Conclusion

"I always thought something was fundamentally wrong with the universe" [1]

References

[1] D. Adams. *The Hitchhiker's Guide to the Galaxy.* San Val, 1995.