WGU C951

Task 3

MACHINE LEARNING PROJECT PROPOSAL

Amy Saunders

Student ID #000686585

Date: 7/16/2021

Table of Contents

A. Project Overview

We Play Math (WPM) is a math curriculum provider geared toward homeschoolers and independent learners. Math skills are taught conceptually in the form of short video lessons. Earlier math levels mainly utilize animated cartoons while more advanced levels utilize interesting graphics and charts, but all are visual, multi-sensory and interactive. Once a concept has been learned, students are given the opportunity to practice the skills via games. Students advance through math skills sequentially, with each skill building on previously learned skills. WPM also includes supplementary practice exercises and materials for educators.

A.1. Organizational Need

In order to meet consumer expectations and to provide exceptional quality instruction to students, WPM must incorporate machine learning techniques to train AI to fulfill the following needs. First, students will transfer into WPM from different curricula at varying points during their math education. WPM software must include computerized adaptive formative testing (CAFT) to place students at the correct level and also to identify any gaps in previous learning so as to be able to remediate those gaps.

We will also need multiple game-playing AI to direct the skills practice sections of our curriculum and we want an AI to guide students along personalized education paths. The AI guide will continually adjust the difficulty of content in order to keep students engaged but also allow them to progress through the material as quickly and efficiently as possible without having to waste time practicing skills at which they are already proficient. All of these AI needs will be addressed in time, but the focus of this project will be the computerized, adaptive, formative assessments students will take upon entry into the WPM curriculum.

A.2. Context and Background

Before the advent of the internet, knowledge had to be acquired through either reading, listening to lectures or experimenting. Since the advent of the internet and smartphones, we carry access to almost unlimited knowledge around in our pockets. As demand has increased, so have expectations. The clunky online classrooms and learning management systems of ten years ago are laughable today. Not only must modern learning management systems be streamlined and user-friendly, but they must also incorporate smart technologies, like machine learning, to make them as efficient and helpful to humans as possible. Machine learning is especially useful in customizing modern educational software to perfectly meet the needs of students' individual learning styles, challenges and strengths.

Machine Learning has changed the whole test-taking process with its ability to adapt, or personalize the test to each test taker. In schools, students tend to progress through a level of math per year, regardless of their understanding or experience. Students are grouped by age rather than ability. However, homeschoolers and independent learners tend to progress sequentially through math courses at a more natural rate of progress rather than one level per year, often accomplishing learning more efficiently and completing more than a level per year. Another concern is switching curricula and being unable to match up perfectly the scope and sequence between them, so a child that might have learned order of operations in fifth grade in one school moves to a school that learned it in fourth grade, leaving that student short of a necessary skill. Adaptive assessments can pinpoint gaps in education along with misunderstandings and can suggest resources to remediate them.

When homeschooled students switch between paper textbooks, publishers recommend that they take a placement test to find the right starting point. However, with the widely-varying discrepancies between curricula, that could mean a student might have to take three full tests before finding the best fit, and that test would still not pinpoint the gaps in knowledge. A single, computerized adaptive assessment would be able to adjust students automatically between multiple levels of tests, finding the perfect starting point for that student in a much more efficient, enjoyable manner.

A.3. Outside Works Review

While researching this project, I found three especially pertinent resources. In his excellent article, *6 Ways Machine Learning Will Revolutionize the Education Sector*, Matthew Lynch discusses several solutions to problems commonly encountered in education, a few of which directly relate to the organizational needs of WPM outlined above. One of those WPM needs is to individualize learning environments and pathways for students so that they are more engaged with the learning and also so the learning process is efficient. Matthew talks about how machine learning in the form of learning analytics, or computers making connections by rapidly connecting pieces of data, can help learning management systems gain insight and provide direction that would have been impossible for humans. He also talks about how adaptive learning, which analyzes student performance in real time and modifies the learning process based on that data, can be used to remediate struggling students and challenge gifted students (Lynch, 2019).

Another useful article by Olga Pekisheva asks and answers the question, “What role does machine learning play in education in 2020?” It begins with a discussion of three types of machine learning algorithms: supervised, unsupervised, and reinforcement learning. It then proceeds to discuss how the algorithms analyze student data to guide them along their learning journey, whether by making them repeat concepts they have not fully learned or proceed to new material. It also discusses ways that machine learning can help the educators to monitor student progress in deeper, more effective ways than just by grading assignments. Machine learning enables the learning management system to actually predict outcomes for students as well as to pinpoint shortcomings or gaps in knowledge so they can be rectified. The article concludes with some specific examples of actual machine learning in classrooms followed by a discussion of challenges facing machine learning in education and ways they could be overcome in the future (Pekisheva, 2020).

The third article I chose to review is a journal article titled *Components of the Item Selection Algorithm in Computerized Adaptive Testing.* It basically explains the 3 components of a conventional CAT item selection algorithm: test content balancing, the item selection criterion, and item exposure control along with methodologies underlying each component. He discussed multiple algorithms and their applications, including new designs that improve testing, live item calibration, new diagnostic methods and test security measures. He then closed by talking about how exciting CAT research is because the implications are just beginning to be applied to the field of education, especially in regards to formative assessments (Han, K. C. T., 2018).

A.4. Solution Summary

Earlier we identified multiple organizational needs for WPM, each of which could be solved by machine learning. We’ve decided to initially focus on the need for a dynamic, adaptive, formative assessment more in depth and discuss a specific machine learning solution to address this particular need.

Our adaptive assessment will be based on a family of complex mathematical models called item response theory (IRT). An item is a single question in the bank of questions. The item bank must first be calibrated through pre-testing. Each item must be exposed a minimum of 200 times before its data is considered reliable. Item response theory first analyzes test data to develop item response functions (literally a logistic curve) for each response. The x-axis represents the knowledge of the examinee and the y-axis represents the probability that a response is correct. The functions, or logistic curves, representing each test response defines that questions difficulty, discrimination and guessing parameters, and each will be a different shape. Using the exposure data, IRT is then used to calculate IRT parameters for each of the items in the bank. The machine will use patterns extrapolated from the functions and the data to grow increasingly dynamic and responsive, consistently improving the tests so they are better individualized and more useful to each student.

Our machine learning solution to the need for game-playing AI is much simpler. It will mainly utilize the policy of alternating between choosing the branch that maximizes the outcome of winning on the computer's turn and minimize the outcome of the students winning on their turn (called the minimax algorithm). Then each game state can be assigned a value based on the likelihood of that state leading to a win. When the number of game states is too large, we will probably use the Monte Carlo tree search algorithm because it's much faster to search random portions of the tree than exhaustively search every branch. The games can play themselves and provide their own training data from which they can learn and improve.

A.5. Machine Learning Benefits

Individualized tasks and customized recommendations are just a few of the numerous benefits of machine learning applications to education. As students follow a personalized, individualized path through the learning materials they spend more time on concepts they struggle with while glossing over concepts with which they are already proficient. The AI can customize a level that is the perfect level of difficulty for each student so that it feels challenging, increasing engagement and eliminating boredom, but not so challenging that students feel frustrated. High-achieving students can feel challenged while low-achieving students are not just left to guess at questions they've never been exposed to.

Perhaps the very biggest advantage to using machine learning in education is the ability to spot exactly where students are committing errors in an algorithm. In pencil-on-paper math, kids naturally skip steps, making it difficult for teachers to see exactly where their misunderstandings are taking place. Was the incorrect answer due to a misplaced sign or a place-value mistake or a fundamental order-of-operations misunderstanding? Even if students did meticulously write down every step, it would take far too long to figure that out for multiple problems for 25 students per class over six periods, day in and day out. It would not be difficult, however, to teach an AI exactly what to look for using appropriate data sets. Students receive precise, immediate feedback. The AI would then be able to direct students to resources that would correct a misunderstanding and encourage students to practice that skill.

B. Machine Learning Project Design

B.1. Scope

The scope of this phase of the project is solely to create computerized adaptive formative testing for WPM. The remaining portions of the project will happen during later phases. The scope of the test includes:

- Ability to place students at the correct level
- Identify any gaps in learning
- Suggestions for resources to remediate those gaps
- Data set acquisition
- Item bank calibration
- Item bank maintenance
- Test security (and information security)

The scope of development of the adaptive test does not include:

- Proctoring

B.2. Goals, Objectives, and Deliverables

The goal of WPM in using computerized, adaptive formative assessment is to provide the optimal assessment experience for each student. These benchmark assessments will be able to efficiently cover multiple math levels and will be able to accurately pinpoint gaps in student knowledge. Our objective is to have that assessment 95% accurate and for it to take students less than 40 minutes to complete, on average. The comprehensive, computerised, adaptive formative assessment itself is the deliverable. The assessment will be added to the We Play Math site as a useful tool for educators and students who want to use WPM curriculum, but also as a tool for educators (including parents) who just want to address educational gaps.

B.3. Standard Methodology

This project will follow the CRISP-DM methodology for data mining with overarching team-based agile project management.

1. Business understanding: WPM needs an automated, efficient, effective, computerized, adaptive test to benchmark students as they enter the program.
2. Data understanding: We need math questions/answers, but we also need sets of possible incorrect answers and the logic behind each incorrect answer. The data will already be clean. The data set must be large enough to yield statistically meaningful results and representative of the data set as a whole.
3. Data preparation: We need to separate our data into a training set and a test set. We'll use about 80% of our data for our training set and the rest for the test set. Only use the training set to build the models, saving the test set of data to verify the test's ability (never train on test data).
4. Modeling: Our goal is to create a model that generalizes well to new data. The data models are used to represent the data and how it is stored in the database (the relationship between data items). We'll start with a regression algorithm and repeatedly adjust biases to prevent over or under-fitting.
5. Evaluation: From our models, we'll select the one that best meets our business objectives and then determine whether to iterate further or progress to deployment.
6. Deployment: A model is only useful when the customer can benefit from the results. Assessment results will be immediately available in the students interface.

B.4. Projected Timeline

07/01/2021 – The proposal is accepted.

08/01/2021 – A technical proof of concept is presented.

09/27/2021 – Testing begins.

11/01/2021 – CAT Delivered

Sprint Schedule

Sprint	Start	End	Tasks
1	Date 08/02/2021 08/09/2021	Date 08/06/2021 08/13/2021	**Planning and project setup:** Define task and scope requirements Setup project codebase
2	Date 08/09/2021 08/09/2021 08/16/2021	Date 08/13/2021 08/13/2021 08/17/2021	**Data collection and labeling:** labeling documentation data validation revisit step 1
3	Date 08/16/2021 08/23/2021 08/23/2021 09/06/2021 09/13/2021 09/20/2021	Date 08/20/2021 08/27/2021 09/03/2021 09/10/2021 09/24/2021 09/24/2021	**Model exploration and refinement:** establish model performance baselines fit simple model to training data try parallel ideas debug perform optimization revisit steps 1 & 2 to ensure feasibility and data quality
4	Date 09/27/2021 09/27/2021	Date 10/01/2021 10/10/2021	**Testing and evaluation:** Evaluate model on test distribution Write tests
5	Date 10/13/2021	Date 10/17/2021	**Deployment:** Deploy model to a small subset of users

	10/13/2021 10/27/2021	10/24/2021 ongoing	Monitor live data & model distributions Periodically retrain model

B.5. Resources and Costs

Resource	Description	Cost
240 work hours	IT Team (2 members)	36,000
120 work hours	Management Team (2 members)	21,600
60 work hours	back end development	2,700
14 work hours	front end development	630
15% buffer		9,140
	Total	70,070

B.6. Evaluation Criteria

Describe the criteria used to evaluate and measure the success of the completed project.

Objective	Success Criteria
User ratings & feedback	User exit survey scores 70% or better with positive feedback.
Parent/Educator feedback	80% or more parents/educators feel student placement is accurate
User friendly interface	85% or higher students are able to complete test successfully without intervention

C. Machine Learning Solution Design

C.1. Hypothesis

We Play Math would like a computerized, adaptive, formative assessment to offer students who want to use their curriculum and might be switching from a different curriculum. Machine learning is a good solution to this problem because it will enable the adaptive tests to be able to select the next question based on students previous performance and to identify gaps in conceptual mathematical understanding. Machine learning will also enable the assessments to recommend remediation resources so students can quickly rectify misunderstandings without having to move backward entire levels.

C.2. Selected Algorithm

The key component of a computerized adaptive test is the selection algorithm, which should find the best suited items (questions) for each student based on his ability. The item selection method is one of the most critical parts of adaptive tests because different methods may lead to different questions for the same student and, consequently, to different classification results. A few other concerns must also be addressed when selecting the next question in an adaptive test, such as content balancing and exposure control. These are generally integrated into item selection heuristics by adding small probability experiments to limit overexposure of items and by limiting the items available for selection by making the decision about which item to select next dependent on previously selected items (Nurakhmetov D., 2019). We can use a sequential item selection (SIS) algorithm with Reinforcement Learning in order to reduce data consumption and associated costs of item selection during classification. This will not only optimally span the item pool, but also minimizes the test length needed to classify the students and helps determine the policy for test termination. The Reinforcement Learning agent learns which item to select next because it is rewarded for successful classification of the input response pattern. The Maximum Likelihood Estimation (MLE) algorithm is the preferred algorithm of developers to determine student placement.

C.2.a Algorithm Justification

The article I found most compelling suggested literally dozens of different algorithms to take care of all the components of adaptive testing: item selection, content balancing, exposure control and test termination. The author also discussed reasons different developers prefer the different algorithms. At the beginning of the test, there are no data from which to estimate the skill level of the student, so the value is initialized at about the 50th percentile (called the expected value). Test developers usually prefer the MLE method because it lacks the estimation bias of Bayesian methods due to their use of an informative prior. The MLE method is unable to handle special response patterns, so estimation methods are often used in adaptive testing. For example, whenever the likelihood of having all correct or all incorrect responses is high, usually when the response set is very small, the EAP method is used. As the test progresses, providing a larger data set, the MLE method is increasingly accurate (Nurakhmetov D., 2019).

C.2.a.i. Algorithm Advantage

As stated above, an advantage of using the MLE algorithm is that it does not have the estimation biases that Bayesian methods often exhibit because they use an informative prior.

C.2.a.ii. Algorithm Limitation

As stated above, one algorithm limitation of the MLE method is that it is sometimes unusable in an early-stage testing due to its inability to handle special response patterns such as responses that are all correct or all incorrect.

C.3. Tools and Environment

The tools required to build a machine learning model of this type are Python, PyCharm IDE, and the CATSim library for test initialization, item selection, ability estimation and test termination.

C.4. Performance Measurement

Quality and performance of the computerized adaptive test will be measured against the performance of students using conventional tests. Studies comparing CAT and conventional testing programs can differ enormously in quality, depending on factors like data collection, student samples and depth of analysis. Generally, though, CAT are thought to be both more efficient and a better measurement of students' actual abilities.

D. Description of Data Sets

D.1. Data Source

In computerized adaptive testing, the data is collected from the students taking the test. We will provide the items and the correct answers used in the bank, based on public use items from the National Assessment of Educational Progress, but selection will happen based on student data. Training and test data used to calculate the parameters will be obtained from volunteer test-takers.

D.2. Data Collection Method

Data for the initial item bank must be collected and calibrated prior to the testing program becoming operational. Calibration data will be collected using the same administration software as will be used operationally from a large, representative sample of homeschooled students. We will store the data in a database for later retrieval.

D.2.a.i. Data Collection Method Advantage

The advantages of this data collection method are that the data collected will be recent and it will be from a large representative sample of homeschooled students using the software that will be used operationally, so it will be optimally accurate.

D.2.a.ii. Data Collection Method Limitation

The biggest limitation I anticipate is the assumption that questions will perform the same way once operational as they did in the calibration sample. In the event that they don't, the quality of reported scores can be compromised.

D.3. Quality and Completeness of Data

We intend to clean and format the data so that it is as free of errors and as reliable as possible. Data cleaning consists of two stages: error identification and error solving. The first step is to identify any anomalies. Data must be formatted consistently and any missing values should be filled out. Rows with missing data that cannot be filled out should be deleted. Corrupted data must also be removed, as should duplicate rows. Just make sure that rows you delete don't provide data present in other rows of the training data. Next, we will ensure that there are no typos or inconsistencies between upper and lower case. Outliers can be detected by way of a box and whisker plot or a scatter plot. Outliers aren't always bad data, but if there are just a few we can drop them from our data set. If they are substantial, they may merit a second look. Finally, we will reduce our data set so it is more manageable and so it will generate more accurate results.

D.4. Precautions for Sensitive Data

Student data, because students are typically minors, is especially sensitive and governed by FERPA laws. In order to protect individual privacy in the big data sets machine learning requires, different anonymization techniques have been used: k-anonymity, I-diversity and t-closeness. Further, identifying information can be deleted from specific columns. However, these practices may not be sufficient protection for sensitive student data. One of the most effective approaches to privacy within machine learning is differential privacy. It presents a stringent privacy notion guaranteeing that no individual student's data has a significant influence on the information released about the dataset (A. Dorschel, 2020).

References

Machine Learning in Education: The Future is Closer Than You Think. (2020, May 24). Husky Jam. https://huskyjam.com/blog/machine-learning-and-education-industry-future/

6 Ways Machine Learning Will Revolutionize the Education Sector. (2019, March 6). The Tech Edvocate. https://www.thetechedvocate.org/6-ways-machine-learning-will-revolutionize-the-education-sector/

Han, K. C. T. (2018). Components of the item selection algorithm in computerized adaptive testing. *Journal of Educational Evaluation for Health Professions*, *15*, 7. https://doi.org/10.3352/jeehp.2018.15.7

Nurakhmetov D. (2019) Reinforcement Learning Applied to Adaptive Classification Testing. In: Veldkamp B., Sluijter C. (eds) Theoretical and Practical Advances in Computer-based Educational Measurement. Methodology of Educational Measurement and Assessment. Springer, Cham. https://doi.org/10.1007/978-3-030-18480-3_17
https://link.springer.com/chapter/10.1007/978-3-030-18480-3_17

Dorschel, A. (2020, May 8). *Data Privacy in Machine Learning*. Luminovo. https://luminovo.ai/blog/data-privacy-in-machine-learning