Team 7:
Project Title: Reinforcement Learning Model to Improve Student Study Habits
Team Members: Alex Ogle, William Sessoms, Daniel Moon, Benjamin Belandres, John Cordwell III

Project Proposal Q&A

1. **Who is the end user? Who will be using your machine learning model?**

The end user of our model is any student who wants to improve the strategy behind their study habits.

2. **How does your product make the end user's life significantly better? What is the problem to solve?**

The product improves the end user's life because it will allow the user to make more informed decisions about what they choose to study. Through that, the user's grades will hopefully be improved.

3. **Is your problem a classification, regression, segmentation, recommendation, etc.? Why is Machine learning suitable?**

The problem is the reinforcement learning model and will be useful since there is no set equation for determining how well a student will do on an exam. There are guesses we can make based on previous decisions and set amount of study time but none of these will always give the same exact output. The use of machine learning can help give us a probabilistic equation guiding students on what they will most likely fail in based on their habits.

4. **What machine learning technique(s) will you use to solve your problem? Why?**

Techniques required for machine learning are a reward and value system for assisting the model in determining what is "right" and "wrong". We will also be using positive reinforcement rather than negative reinforcement for the model. There is the possibility that we might be able to combine reinforcement learning with logistic regression and log loss for the model. Our team believes that logistic regression will help more with determining what is a more probabilistic outcome, and the paper recommended the use of reinforcement learning for training a model with the associated data. A combination of the two might be able to provide more useful results of the data.

5. **What is the dataset for your project? Do you have enough data for the selected machine learning technique? How much effort will it take to get the data ready for ML training?**

Our dataset is EdNet, which tracks student interaction on the Korean tutoring platform *Santa*. The dataset contains information on 131,441,538 interactions from 784,309 students, so we

believe this dataset is sufficient in size. EdNet is already very well organized, but we may need to do some cleaning and apply techniques like one-hot encoding to make the data easier to work with.

6. **What are the baseline model(s) and paper(s) you will use for inspiration? Are you replicating results, improving performance metrics, reducing inference time, or enhancing interpretability?**

Our baseline models and references begin with the EdNet paper, *Ednet: A Large-Scale Hierarchical Dataset in Education*, which introduces the dataset and highlights key applications such as knowledge tracing and dropout prediction. This paper provides the foundation for our work and examples of benchmark results. For our modeling approach, we will use both classification and reinforcement learning techniques. We will implement logistic regression, which allows us to model the probability that a student answers the next question correctly based on features such as prior experience and skill tags. This will help us compare more advanced models.

We hope to establish strong classification baselines and push toward adaptive recommendation strategies that improve student outcomes over time.

7. **How will you measure model application performance?**
   Examples: Classification: Accuracy, F1, ROC-AUC, Precision/Recall; Regression: MSE, RMSE, R^2; Segmentation: IoU, Dice score

There is an existing leaderboard with models trained on the same dataset, and they use Accuracy and AUC, so we'll probably do the same. It can be found here.

8. **Do you have enough computational resources to handle the machine learning technique and the data? Can you train the model with the resources you have?**

It appears that the strongest computer we have for machine learning is a laptop with an NVIDIA Geforce 4060; however, we also have quite a few other powerful PCs on the backburner (Radeon 6700XT, Geforce 4050).

9. **What are the key milestones in your roadmap?**

   Add or delete rows as needed. Grid is helpful way to visualize the milestones you envision

#	Tasks Description	Priority	Deadline
1	**Rough planning of overall structure** **Division of Testing, Validation, and Training datasets** **Preprocessing Pipeline 1 Planned** ● Missing data handling ● Initial feature extraction and selection	**High**	**9/23/25**

2	**Preprocessing Pipeline 1 Finished** **PreProcessing Pipeline 2 Planned** ● Feature scaling ● Dimensionality reduction ○ Feature selection ○ Feature extraction	**High**	**9/30/25**
3	**Model 1 created** ● Analyze performance after first test run ● Plan future turning steps	**High**	**10/7/25**
4	**Model 1 revision** ● Tune reinforcement parameters ● Adjust dataset if needed	**Medium**	**10/14/25**
5	**Midterm Project Report**	**High**	**10/21/25**
6	**Begin Draft for Technical publication** ● Develop skeletal structure of our method, results, and uses of model	**Medium**	**10/27/25**
7	**Completed draft of Technical publication**	**High**	**11/3/25**
8	**Technical Publication Revision #1**	**Medium**	**11/6/25**
9	**Technical Publication Revision #2**	**Low**	**11/13/25**
10	**Delivery of project presentation**	**High**	**11/20/25**
11	**Delivery of final project report**	**High**	**12/02/25**

10. What is the expertise of each team member?

Name	Expertise
Alex Ogle	Novice in machine learning with some experience in python
Daniel Moon	Python programming, very little machine learning knowledge
Will Sessoms	Python programming, minimal ML experience
Benjamin Belandres	Python programming with limited machine learning experience

John Cordwell III	Python Experience, general software experience, and a singular ML project with YOLOv8

11. How will the project tasks be distributed within your team?

Task/Role	Name
Code Lead	Will
Primary Coders	Will, Daniel, John
Group Leader	Alex
Head of Research • Reading related works • Looking into models that use the same dataset • Finding answers to bottlenecks in coding	Ben(jamin)
Machine Learning subject matter expert	Alex
Quality Assurance	Will, John, Daniel, Ben(jamin)

12. Do you want to produce a technical publication out of this project?

Yes.

References: [Add references or URL to papers, websites, blogs, datasets, etc. used in this proposal.]

Dataset: https://github.com/riiid/ednet?tab=readme-ov-file

Leaderboard: http://ednet-leaderboard.s3-website-ap-northeast-1.amazonaws.com/