

Artificial Intelligence in Education (263-5005-00L)

Instructions and Advice for Final Projects¹

1 Introduction

If you're reading this, you have chosen to embark on the exciting endeavor that is a final research project for Artificial Intelligence in Education (263-5005-00L). Hooray! There are a number of steps you should take to help avoid getting stuck in frustrating situations:

- Clearly define your goals. At the very beginning of your project lay out exactly what you hope to achieve and create a detailed timeline and make sure everyone in your team understands them. Make sure they are reasonable goals, taking into account your current research experience, group size, and the time frame.
- Do a thorough literature review. In doing so, you will get a sense of which approaches are likely to fail and, of course, which approaches have already been attempted if you're trying to do something novel.
- Consult with the course staff. If you are feeling stuck, please email the course staff or post on Moodle. The course staff will be able to point you in the right direction.
- Don't be scared of negative results. As a course project, a negative result is not a failure by any means and can certainly earn full credit. We simply want to see a rigorous scientific plan laid out and executed. The findings are less important for your grade.

Expectations

Your project should aim to demonstrate an understanding of the techniques reviewed in class, specifically artificial intelligence methods applied to an educational setting. You are not expected to make a scientific contribution to the domain or beat a state-of-the-art method within the scope of the course project. Instead, we simply want to see an application of the modeling techniques you have learned to an educational task. It is important that your summary paper provides a rigorous explanation of your experimental findings. This includes both qualitative and quantitative analysis. An example of qualitative analysis would be finding specific examples where your model did not work as expected. Importantly, your work will not be graded on the performance of your methods! Rather, projects will be graded holistically, taking into account criteria such as: originality, complexity of the techniques you used, thoroughness of your evaluation, amount of work put into the project, analysis quality, and write-up quality.

2 Choosing a Project Topic

Project Suitability. You can choose any topic related to Artificial Intelligence in Education. In addition, your project should make substantive use of the methods taught in this class. It would not be okay, for instance, to use a sentiment analysis model for movie reviews². If you have any questions about what an acceptable product would entail, please feel free to contact the course

¹ Credits: The project instructions have been adapted from Prof. Ryan Cotterell's Introduction to NLP class (<https://rycolab.io/classes/intro-nlp/>)

² This is not a judgement about the usefulness of the technique. Rather, we simply require that you demonstrate mastery of mostly techniques that are the subject of this class.

staff.

Project types. Here is a non-exhaustive list of possible project types:

- Reimplementing a model from literature without public implementation;
- Writing a systematic review of
 - state-of-the-art models in EDM for a concrete task (e.g., personalization using dialog) and shedding light on their limitations;
 - one notable AIED system (e.g., an intelligent tutoring system that is not covered in the class³) and the nuts and bolts of how it works;
- Contributing to an EDM shared task;
- Providing an experimental or theoretical analysis of an existing model, e.g. performing a replication study of an existing paper with ablation studies or a few more experiments to assess generalizability on other data sets;
- Proposing a novel method with applications in EDM.

Projects outside of these types are allowed but please talk to the course staff over email or on Moodle discussion forum about the appropriateness of your project before the proposal deadline if your project does not fall into these categories. Our goal is not to restrict you, but to make sure you succeed!

2.1 Literature Review

You will be expected to perform a thorough literature review such that you can both summarize and cite relevant work in your write-up. We recommend doing this sooner rather than later as it is generally much easier to define your project if there is existing published research on a similar task or modeling technique. Identifying existing relevant research—including published code—will ultimately save you time and make the process more enjoyable. Prior works can give a sense of best practices for approaching projects, vet out initial ideas, and lay the groundwork for your own research. Finding relevant research is not always easy. However, there are a number of places to start looking:

- Follow-up references from papers discussed in the class. Please contact TAs in case you would need access to the papers for future class topics to develop your project proposal.
- Recent publications at any of the top venues where AI in Education research is published: e.g., [AIED](#), [LAK](#), [EDM conference](#), [ACM L@S](#), [CSCL](#), etc.
- Use a keyword search in the following search engines/repositories:
 - <http://scholar.google.com>
 - <http://dl.acm.org/>
 - <http://arxiv.org/>
 - <https://search.library.ethz.ch/>
- Look at publications from research groups. For example:

³ E.g., Cognitive tutor (MATHia) for algebra/geometry, ASSISTments, Betty's Brain, Why2-Atlas, etc.

- <http://www.mrinmaya.io/publications/>
- <https://www.gse.upenn.edu/academics/faculty-directory/baker>
- <https://www.si.umich.edu/people/christopher-brooks>
- <https://adl.asu.edu/leadership/danielle-s-mcnamara>
- <https://cas.gsu.edu/profile/scott-crossley/>
- <https://www.cs.washington.edu/research/nlp/publications-by-year>
- <https://scholarphi.org/>

2.2 Finding Data

Here we present options for finding data for your project.

Shared Task datasets. Shared tasks (also called bake-offs) are competitions to which researchers submit systems that address specific, predefined challenges. Often, the data provided by the organizers of these tasks is well documented, clean, and standardized. This often makes these datasets incredibly user-friendly. You can find examples of such tasks here: <https://sites.google.com/view/tipce-2021/shared-task> and here: <https://eedi.com/projects/neurips-education-challenge>.

Using publicly available datasets. There are a number of online resources that host standard educational datasets. We present a few here:

- [Papers with Code](#)
- Social Media and Online Platform Data: many online platforms provide APIs or readily compiled data dump files. Consider Twitter, Reddit, YouTube transcripts, Wikipedia, Reuters.
- [Kaggle](#): a platform for hosting competitions on tasks that also provides many datasets.
- [Datashop](#) : a repository that contains educational datasets (including [KDD workshop](#))
- [Education data on data.world](#): contains some open-source educational datasets
- Emailing the Authors: If you come across an interesting dataset that is not publicly available, you may consider contacting the authors. Often, authors of academic papers are more than happy to email you a copy of their dataset.
- [Assistment](#)- data from an online tutor that simultaneously teaches and assesses students in grade school mathematics
- [EdNet](#) - dataset of all student-system interactions
- [Coursera discussion forum threads](#) - discussion threads from the forums of 60 Coursera Massive Open Online Courses (MOOCs)
- [XuetangX](#) - data from massive open online course (MOOC) platform initiated by Tsinghua University
- [Duolingo](#) - [language translations dataset](#), [spaced repetition dataset](#),

Just because a dataset is online and already in a single, structured file does not mean the dataset has been processed in a manner suitable for your project. Take care to analyze any online resources to first ensure their quality and second, determine what additional steps you may need to take in order to properly format the data.

Collecting new data. It is possible to collect your own data for your project, including rich

[multimodal behavioral data](#) - especially if you choose to conduct a project that does not involve coding. However, keep in mind that data collection is often a time-consuming and messy process and is more difficult than it appears. If you choose to do so, make sure to budget the data collection time into your project and provide a thorough explanation of your processing techniques in the write-up. Note that we do not recommend collecting your own data if you plan on implementing or creating a model.

3 Advice

3.1 Data Processing

Data processing, even for already “clean” datasets, can be cumbersome. It is worth investing time in a thoughtful, well constructed pipeline to assure you do not have to redo experiments. The following should all be considered:

Cleaning. You should consider if your data needs to be transformed before modification. You might find the following tools useful:

- [Scikit-learn](#): standard Python Machine Learning library;
- [NLTK](#): a lightweight NLP toolkit written in Python;
- [Spacy](#) : NLP pipeline for several languages.
- [HuggingFace](#): Various pretrained NLP models and easy to use NLP APIs

Splitting. At the beginning of your project, you should split your data set into training data (most of your data) and test data. A typical train–test split might be 80–20. Explicitly, training data can be used to estimate the parameters of your model. The validation data should be used for model selection, but not parameter estimation. Model selection may consist of choosing the best hyperparameters for an architecture or deciding whether to stop your optimizer early to avoid overfitting. The test data should only play a role at the end of your project. You should only evaluate your model once on the test data. Looking at the test data repeatedly is known as [p-hacking or data dredging](#).

Preliminary analysis. Before modeling the data, performing a preliminary analysis of the data is helpful for successfully completing the project. Actually looking at your data may reveal flaws, e.g., missing entries, inconsistencies, or oddities, e.g., unexpected trends, noise. When researchers fail to look at the data, they often discover such inconsistencies at a later and much more inconvenient time. Some recommended approaches include building visualizations and computing basic statistics. For example, does your data follow an expected probabilistic distribution e.g. Gaussian, Binomial? What is the variance of various attributes? Are some features strongly collinear? While we do not require such analysis in your writeup, we nonetheless recommend it. There are a number of online resources⁴ that provide guidance on how to do such an analysis.

3.2 Baselines

⁴ such as <https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4> or <https://r4ds.had.co.nz/>

One aspect of good experimental practice is selecting a good baseline to benchmark your method and properly comparing against that method. Without a fair comparison to previously proposed methods, we cannot tell whether an innovation is actually an improvement over what the field currently offers. For instance, comparing against a simple baseline may reveal that your task does not actually require a complex, resource-intensive model. In general, simpler methods should be preferred⁵. Regardless of your chosen research topic, we will expect some sort of baseline comparison in your final report. Note that some baselines are theoretical. Baselines can include simpler models or models proposed by other works. You should first make an effort to understand what the baselines are.

3.3 Use of Existing Infrastructure

There are a number of software tools and packages that can aid in your research. You are allowed (and encouraged) to make use of these tools, however, if you do so, please be explicit about it in your writeup. These include:

- PyTorch <https://pytorch.org/>
- TensorFlow <https://www.tensorflow.org/>
- scikit-learn <https://scikit-learn.org/stable/>
- HuggingFace Transformers <https://github.com/huggingface/transformers>
- Various other packages in R, Python, etc.
- Open source GUI tools such as [LightSIDE](#), [Orange](#), [JASP](#), [Gephi](#), etc.

As is the case with collecting data from scratch, we do not recommend building complex neural models from scratch. While coding an LSTM from scratch may be an interesting exercise, it will likely take more time than you want to spend. After all, the goal of this project is to answer an educationally relevant research question. Note that training a predefined model in a deep learning library with readily available preprocessed data (e.g., reproducing one of the tasks README) does not constitute a full project! However, making use of pre-trained models in your modeling process is fine. If you are unsure of your practices, please contact the course staff.

3.4 How to Write Systematic Reviews

- Describe the AIED system / EDM method
- Discuss how widely it is or was used, and by who
- Discuss what contexts it was used in
- Describe how it functions (or functioned)
- Describe what pedagogies it supports
- Describe what (if anything) was interesting or historical about this AIED system / EDM method
- Describe how well it worked with real learners (both learning and other outcomes)
- Identify boundary conditions (or, limitations) for when the AIED system / EDM method

⁵ Occam's razor is a good maxim to be aware of when doing research.

- does not work as intended
- Based on the identified limitations:
 - propose two-four concrete ways to redesign either the
 - systems's intelligence (reasoning about student/tutoring knowledge, analyzing student solutions and determining its quality, adapting teaching responses to students' cognition and/or affect), and/or
 - interface (input and responding modality)
 - Support with references from the literature.
 - propose two-four concrete ways to redesign the data mining method based on improvements to the algorithm, expected results and a discussion of why/how those changes may alter the model performance. Support with references from the literature.

4 Deliverables

All deliverables will contribute to your final project grade. After submission of proposals, each group will be assigned to one of the TAs. If any questions arise during your project, you may schedule a meeting with your assigned TA in between each of the deadlines to discuss them. Please reach out to your assigned TA with ample time as they likely have scheduling constraints.

• Proposal

Your initial plan should be 1–2 pages in length (not strictly) and should, at a high level, provide answers to the following questions:

- What are the specific details of the task you are investigating? What is the input and what is the output? Give concrete examples. If the task can't be framed as input and output, what exactly are you trying to achieve?
- What dataset(s) and tools are you planning to use?
- What does success look like for your project?

Please be sure to include all group members' names and nethz usernames in the proposal.

• Final Write-up

The final report should take the form of a research report. Your write-up should, at a minimum, include an introduction, background section, theoretical explanation of modeling techniques, experimental findings, and analysis. You are required to use [ACL 2021](#)⁶. We recommend looking at the structure of other research papers as a guideline for your write-up's layout.

We additionally expect you to submit any code used during your experiments, along with a README giving instructions on how to reproduce your results. Reproducibility is a fundamental part of research afterall. The simplest way of doing this is to create a [GitHub](#)

⁶ Please uncomment the "final copy" command.

repository.

- **Presentation**

Each group must give a 15 minute presentation on their project⁷. While there are items we want you to cover, the exact format is up to you. You may, for example, give a live demo or create a series of infographics. Regardless of the format you choose, note that we are expecting that the presentations provide appropriate background information for the project and discussion of final results of the project. Due to COVID-19, presentations will take place online and we will ask you to pre-record them. You may upload them to polybox or Google drive and share with your assigned TA. More details will follow soon.

5 Other Project Topics

We encourage you to propose your own topic and pick a project of your liking. Alternatively, we also provide [some suggestions](#) for you. Feel free to contact the relevant mentor for these projects for details. Your project score will eventually depend on the novelty and execution of your ideas.

⁷ While we do not restrict you to 15 minutes, all critical information must be conveyed in the first 15 minutes as the TA team may not have time to watch >15 minutes of each presentation.