

## Final Project Guidelines

*Instructor: Thorsten Joachims*

# 1 Overview and Timeline

There are two kinds of projects that can be done: *Self-Defined Projects* and *Kaggle Competition Projects*. **Choose only one of the two.**

## 1.1 Self-Defined Projects

The final project is intended to be a limited investigation in an area of machine learning of your choice. The purpose of the project is to enable you to study an area of your interest in greater detail in a practical way. The project can take on many forms, including but not limited to,

1. Projects that explore the application of machine learning ideas to an interesting “real-world” problem.
2. Projects that involve a theoretical or empirical study of aspects of a learning method or model.
3. Projects that do an experimental, comparative study of various machine learning methods.

Doing such a project gives you more flexibility and allows you work on something of your liking. However at the same time, this may potentially require some additional effort (depending on your problem) such as data collection or coming up with suitable baselines or such.

## 1.2 Kaggle Competition Projects

Alternately, you can try your hand at Kaggle Competitions. On the website you can find and choose from a number of interesting machine learning competitions. Upon joining a competition, you will be provided with a training and testing sets, and your performance will be measured with specified metrics and ranked with other competitors on the web. You can even win money if you do very well!

The system will score your predictions and inform you of the same. Your scores will be uploaded to a leader-board for you to compare your performance with your classmates as well as research groups from other institutes.

**Note that performance on the different metrics is not the critical factor in your grade on the project.** While doing well on the competition will help, you are expected to come up with interesting ideas to solve the task you choose, which is what we will primarily look at.

While the data is easier to obtain for such a project, there is less flexibility and more emphasis on coming up with interesting methods.

### 1.3 Common Guidelines and Timeline

Please propose a topic to us in your project proposal, and we will give you feedback on the feasibility. After the project proposal, you will be assigned a contact TA that you can use as a resource for questions and advice. If you are working with other faculty or students on the project, please mention that clearly in your project proposal.

We recommend meeting with your contact on a regular basis, so that you identify potential problems before it is too late.

You will submit a status report three weeks into the project, indicating progress made. You will present the results of your project both in a poster session, as well as via a final project report. More details are given below.

**Peer Reviewing:** We will be performing peer reviewing for the project phase of the course. There are several benefits to peer-reviewing. Most importantly, it helps you understand and appreciate work from other students and groups, and it provide more feedback to everybody about the projects.

Peer-reviewing means that each one of you will be given a few submissions of your classmates to read and *grade*. This essentially involves providing some brief comments to help each other out. Please be as fair and impartial as possible during this reviewing. We will be having TAs evaluate and provide feedback as well. You will be graded on how well you review other projects and how insightful your comments are. This will be part of the project grade

There will be three phases for the peer reviewing corresponding to the project proposal, the poster presentation and the final report. The different phases use different peer-review models, called “double-blind” and “single-blind” reviewing:

**Double-Blind Reviewing:** This means that the reviewer does not know whose project he or she is reviewing, nor do the authors of the project know who is reviewing them. Even more, a reviewer is not allowed to disclose who he or she is reviewing.

As an author, DO NOT put your name or netid on your submission or into your submission filenames.

**Single-Blind Reviewing:** This means that the reviewer knows the identities of the authors of the project he or she is reviewing, but the authors of the project do not know who is reviewing them. Again, a reviewer is not allowed to disclose who he or she is reviewing.

In all models, the course staff will know the identity of everybody.

We will be using Microsoft's CMT system for the peer reviewing.

The URL is <https://cmt.research.microsoft.com/CS47802014/Default.aspx>. Please use the same account for both reviewing and submitting your project proposals, progress report, posters and reports.

**IMPORTANT: You will receive an email from me inviting you to participate as reviewers for the course. Please accept this invitation ASAP.**

**The last day to accept the invitation is 10/20.**

**If you do not do so by then, you will not receive credit for the project.**

**This is a HARD deadline and there will be no exemptions.**

**The process itself is simple (Accept the invite, Click the forgot password option and set your password) so please make sure you complete it asap.**

Further information will be made available before the first peer reviewing deadline. If you have any further questions we will be happy to address them on Piazza.

The important dates are:

**21<sup>st</sup> October:** Submit project proposal for feedback (via CMT).

**24<sup>th</sup> October:** Peer reviews for project proposal due (via CMT).

**11<sup>th</sup> November:** Submit progress report (via CMT).

**04<sup>th</sup> December:** Poster presentation. Submit poster (via CMT).

**05<sup>th</sup> December:** Peer reviews for posters due (via CMT).

**10<sup>th</sup> December:** Final project report (and code) due (via CMT).

**15<sup>th</sup> December:** Peer reviews for final project reports due (via CMT).

**16<sup>th</sup> December:** Author Feedback on reviews for final project reports due (via CMT).

**NOTE: Please remember to form your project groups on CMT well before the deadline.**

**IMPORTANT: Late submissions will NOT be accepted under ANY circumstance.**

**This applies to all phases of the project (proposal, progress report, poster and final report) as well as the peer reviewing. Late days left over from the home-**

works **CANNOT** be used for any project-related submission.

## 2 Project Proposal (Peer-Reviewed, Double-Blind)

The project proposal should outline what you want to do in your project. We will give you feedback on your proposal to make sure the project you are proposing is feasible and appropriate. The proposal should contain the following sections:

1. Size of the Team (Do **NOT** include names/netids). You are expected to work in groups of 3-4<sup>1</sup>. You can find project partners using the Piazza post setup for this purpose. If you are still unable to find a group, email the course staff and we will assign you to a group.
2. Motivation. Explain why this project is interesting and important.
3. Statement of the Problem/Task. A statement of the problem, issue, or task that you're interested in studying. In particular, try to formulate the key questions (2 to 4 questions is probably a good number) that you will answer in the project.
4. General Approach. A high-level description of the general approach you'll use to address the questions. This should include how you will evaluate and what evidence you are planning to gather (e.g. how you can answer the questions through experiments on data).
5. Resources. A list of resources you have/need to conduct the project. This includes reading, software, datasets, etc. How are you planning to get these resources?
6. Schedule. A schedule of work indicating the dates by which you plan to complete components of the project. Make sure the schedule is plausible.

**Format:** Please use the ICML double column format for which you can find the style files/Word template here (look under the formatting header). The proposal should not be more than **one page** in length (excluding references and schedule).

The proposal will be peer-reviewed and should be double-blind (do not include names/netids). Please keep the above guidelines in mind when performing the peer reviewing.

## 3 Progress Report (Staff-Reviewed)

This should be a refined version of the proposal after incorporating instructor feedback, along with a more precise description of your proposed method. *The key addition to the*

---

<sup>1</sup>Larger groups may be accepted only in special circumstances with instructor's permission

*proposal will be a section (titled Progress) describing the progress you have made along with any preliminary results you may have.*

**Format:** Please use the ICML double column format. The progress report should not be more than **two pages** in length (excluding references and schedule)

The progress report will only be reviewed by the course staff (and hence does not matter if you include your names on the report or not).

## 4 Project Poster Presentation (Peer-Reviewed, Single-Blind)

The project will be presented as posters during the last week of class (tentatively scheduled for 12/4 in the late afternoon or evening). Each group will be assigned a group number which will determine where you should put up your poster.

Your poster should contain the following:

1. Provide motivation for your project, explaining why it is important and interesting,
2. Explain your research questions,
3. Provide (preliminary) evidence,
4. Draw (preliminary) conclusion.

You should try to make your posters as self-contained as possible. You should be able to summarize your project (with the help of the poster) within 3-4 minutes. Logistics regarding the poster session will be announced on a later date via Piazza.

The posters will be peer-reviewed, but will be single-blind *i.e.*, you can include your names on the posters. Please keep these guidelines in mind when performing the peer reviewing. You are expected to visit the posters you are reviewing.

## 5 Final Project Report (Peer-Reviewed, Double-Blind)

The final project should be submitted via CMT (not via an email attachment!!) as a .zip file before 11:59 PM on Wednesday, December 10. This .zip file should include at least your writeup (as either a PS or PDF file) and the source code of any programs you wrote for your project. Include other files if you feel they are appropriate, but obviously explain their relevance in a readme. All submissions must be via CMT.

Do not be late with your submissions. This is not a homework you can turn in late at the cost of 5 points per day; peer reviewing and instructor reviewing start immediately.

For additional guidance in structuring the report, take a look at the template structure here. Not every project fits into this structure, and you might choose a different structure instead. The most important goals to keep in mind are

- To motivate your project,
- To make a convincing argument that supports your conclusions,
- To make sure that the reader understands what your project is about and how you came to your conclusions, and
- To make sure that credit is given to all software, literature, etc. that helped you in your work.

**Format:** Please use the ICML double column format. The progress report should not be more than **six pages** in length (excluding references) *Being concise is a good thing, but do not sacrifice clarity and completeness.*

The final report will be peer-reviewed and should be double-blind (do not include names/netids). Please keep the above guidelines in mind when performing the peer reviewing.

For the final reports, you will also have a chance to provide **author feedback** on the reviews you received. This allows you to address concerns or issue clarifications, which the course staff can utilize while they determine their reviews.

## 6 Grading

The projects will be graded in the same spirit as research papers are assessed (though we do not expect you to do original work at the same level). Here is a list of things that we will be looking for:

1. Originality
2. Relevance to course
3. Quality of arguments (are claims supported, how convincing are the arguments you bring forward)
4. Connection to earlier work
5. Clarity (how clearly are goals and achievements presented)
6. Scope/Size (in proportion to size of group)
7. Significance (are the questions you are asking interesting)

Relative to each other, the proposal will account for 5% of the grade, the progress report for 10%, the poster for 15%, and the report for 60%. Your peer reviewing performance

will account for 10% of the project grade.

Feel free to come and talk to us about the various aspects of your project (in fact we strongly encourage you to) so that we can make sure that you are on the right track. Finally, have fun while doing it; its meant to be something that you are interested in doing!

Good luck!