

## Here's the plan for your final project.

(0) Form a working group with 1-4 people. It is strongly suggested that you work with at least one other person so you can divide up the work, but if you have a compelling reason to work alone, that's ok too.

(1) Choose an active project on [Kaggle](#):

[Forest cover type prediction](#)

[Bike sharing demand](#)

[Random acts of pizza](#)

[Detect the location of keypoints on the face](#) (difficult)

(2) Download the data and figure out how to import it into python/numpy objects so you can process it with SK-Learn. Split the data you get into training and development for running your own experiments. Establish a baseline and submit to Kaggle for verification. For your submission, you should probably train your model on all the data you have. Send Todd and Dan a link to the leaderboard that shows your baseline score.

**Due Date: 3/30**

(3) Schedule a meeting with Todd or Dan to discuss what you've done and what you're planning to do.

**Due Date: 4/6**

(4) Run experiments! Try different models and parameter settings. Engineer new features. Use feature selection techniques. Examine errors and iterate. See how much progress on the leaderboard you can make.

(5) Synthesize your work in an ipython notebook. Target your explanation and analysis to the version of yourself before you started this course. We'd like to make your notebooks public and as useful as possible to people getting started with machine learning. I'm more interested in experiments and analysis that make concepts you've learned clear than your ranking on the leaderboard (though good performance will make people more interested in reading what you have to say).

Try to limit the number of experiments you include in the notebook to just those that improved results or were interesting in some useful way. If you'd like to include extra work, add an appendix at the end. You should include some kind of summary table that shows the relative contribution of each important idea you implemented.

**Due Date: 4/20**

(6) Your group will be responsible for a ~15 minute presentation in the final live session.

**Due Date: 4/27**

**Grading:**

10% Initial setup work

90% Project notebook and presentation

10% Problem description

25% Sensible methods

20% Feature engineering

15% Error analysis

10% Notebook presentation

10% Overall results

10% Presentation