

Guidelines for Project 2

This document contains guidelines, requirements, and suggestions for Project 2.

Team Effort

Before anything, remember that projects are a group effort: Working closely with your teammates is a requirement. This helps teach real-world collaborative workflows and enables you to tackle more difficult problems than you'd be able to working alone.

In other words, working in groups allows you to **work smart** and **dream big**. Take advantage of it!

Project Proposal

Before you start writing any code, your group should outline the scope and purpose of your project. This helps provide direction and prevent [scope creep](#).

Write this as a brief summary of your interests and intent, including:

- The kind of data you'd like to work with and the field you're interested in (e.g., trading, quantitative analysis).
- Possible source for such data.
- Candidate machine learning or statistical models that you may want to explore.

This constitutes your Project Proposal and Outline, and it should look something like this:

Our project is to predict credit card fraud. We'll fit a random forest algorithm to lending club data to create a predictive model. We will also fit a deep neural network model to the data and compare the performance of both models to determine which has sufficient predictive power.

Finding Data

Once your group has written an outline, it's time to start hunting for data. You are free to use data from any source, but we recommend the following curated sources of high-quality data:

- [data.world](#)
- [Kaggle](#)
- [Data.gov](#)
- [Public APIs](#)
- [Awesome-APIs List](#)
- [Medium APIs List](#)

Chances are you'll have to update your Project Outline as you explore the available data. This is fine—adjustments like this are part of the process! Just make sure everyone in the group is up to speed on the goals of the project as you make changes.

Make sure that your data is not too large for local analysis. Big-data datasets are difficult to manage locally, so consider a subset of that data or a different dataset altogether.

Data Preparation & Model Training

Machine learning typically requires extensive data preparation before the model can be trained. Use Jupyter, Google Colab, or AWS SageMaker to prepare a training and testing dataset and to train the machine-learning model.

Use the training data to fit the model.

Model Evaluation

Use the testing data to evaluate the model. Create any necessary tables, charts, plots, or statistical analysis to report the model performance.

Predictions and Conclusions

Summarize your conclusions and predictions. This should include a numerical summary (what data your model yielded), as well as visualizations of that summary (plots of the final model evaluation and predictions).

Optionally, deploy the model as an interactive dashboard or interface (Lex or SageMaker).

Finally, be sure that your projects meet the [technical requirements](#).

Presentation

After you've tweaked your model to your satisfaction, you'll put together a presentation to show off your work, explain your process, and discuss your conclusions.

This presentation will be delivered as a slideshow, and it should give your classmates and instructional staff an overview of your work. PowerPoint, Keynote, and Google Slides are all acceptable for building slides.

As long as your slides meet the [presentation requirements](#), you are free to structure the presentation however you wish, but students are often successful with the format laid out in the [presentation guidelines](#).

© 2019 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.