Final Project Guidelines
Econ 4676: Big Data and Machine Learning for Applied Economics

Due Date: December 10 at 6:00 pm

The repo link to create your submission is
https://classroom.github.com/g/3isi41i7

The purpose of the final project is for you to showcase the skills that you acquired in this course. You can choose three routes for the project (a) a novel research paper, (b) replicate a paper that you find interesting, or (c) deploy a machine learning model in a web application that allows practical business decisions based on data. If you choose either (b) or (c) you should talk to me first.

There are three stages for the final project:

1. First submission is due on **October 22nd at 6 pm**. This submission should be a brief statement of what you plan to do. I expect that you provide sufficient background so that the reader can understand the problem/research question you want to address. It should include an outline of how you plan to accomplish it, including the data you have/or plan to acquire, the methods, and any other relevant information you deem necessary. It should be at most 2 (two) pages in length. I'll provide feedback and will ask for changes if needed. **This is worth 5% of your final grade.**

2. Presentation of your work. **Presentations will be in class, November 30th, December 2nd, and 3rd**. The order will be randomly assigned. It should not be longer than 15 minutes. You have to treat this as a seminar/stakeholder presentation. For presentation tips, you can check this document or last semester's seminar class. **This is worth 10% of your grade.**

3. Final submission is due on **December 10th at 6 pm**. The final document should not be longer than 5 (five) pages (not including the title page with abstract, and references). When writing up the document, it should contain the following:

   - Title
   - Abstract (200 words limit)
   - Introduction. It should contain at least: the problem/research question clearly defined, antecedents of your work, your value-added (i.e. why your project is interesting/novel/different), and a preview of results and takeaways.

- Data. Treat this section as an opportunity to present a compelling narrative to justify or defend your data choices, walk the reader through your reasoning of how you think you got the right data for the task, describe it accordingly with descriptive stats, graphs, etc.

- Model. Present the model you are using. Be sure to argue why this is the best model for your task. When writing this section think on the following questions: did you apply other models? Is this model the most accurate at predicting? Is this the only model that you can use? etc.

- Results. Here you should present your results. I understand that a semester is a short time to have a full paper, so preliminary results are fine.

- Conclusions and recommendations. In this section, you should state the main takeaways of your work.

This is worth 15% of your grade.