Guideline for the Final Project

API 222: Machine Learning and Big Data Analytics

October 8, 2019

1 Due Dates

- November 6 (for Section A) and 7 (for Section B): Final Project Groups
- November 20 (for Section A) and 21 (for Section B): Final Project Title and Abstract
- November 25 (for Section A) and 26 (for Section B): Final Project Slides
- December 2 and 6 (for Section A) and November 26 and December 3 (for Section B): Final Project Presentations in class
- December 4 (for Section A) and December 5 (for Section B): Final Project Write Ups
- Note: All dues are by the beginning of the class time.

2 Final Project

Below are the set of due dates for each component of the final assignment as well as descriptions of what is due on the corresponding date.

2.1 Final Project Groups

If you are not a PhD student, we recommend you do the final project in groups of 5. PhD students and other students who prefer to work alone may work individually, though we expect the quality of work done by an individual to be of the same caliber as the quality of work done by a group of 5. We will create the final project groups without members on Canvas. You will then join the same group number as the rest of the members of your final project group. Then, as a group, please submit the Final Project Groups assignment. This assignment requires that:

1. All students in your final project group have joined the same Canvas group before submission.
2. You submit in the text field a list of all students in your group so that the teaching staff can cross-reference to be sure all students are accounted for. If you are a PhD student or are planning to work alone, you need to join one of the Canvas groups and be the only member of that group. If you are working alone, when you submit your name in the text field, please also indicate whether or not you are a PhD student.

2.2 Final Project Title and Abstract

The class will vote on which projects it would like to see presented in class. In order to inform the voting, please submit the name of your project and a short abstract of at most 150 words. An assignment has been created on Canvas for your project title and abstract. Please submit in the text entry of the assignment using the following format, where you replace `XXXXXXXX` with your own content:

```
Project Title:  XXXXXXXX
Abstract:  XXXXXXXX
```

2.3 Final Project Presentations in class

Due to time constraints, only 8 groups or individuals (per Section) will have the opportunity to present their project for 15 minutes to the class. These groups or individuals will be selected by a vote from the class; however, all groups will be required to submit presentation slides by the due date specified above. The presentations should contain four slides, and one slide should be dedicated to each of the following topics:

1. The problem motivation, including currently used methods to approach the policy question and how machine learning can add to the existing approaches
2. The model you chose and why
3. Main results
4. Main open questions or concerns that might hinder adoption in the real world

2.4 Final Project Write Ups

The final project write up is an opportunity for you to demonstrate an innovative and thoughtful application of machine learning to a real world policy problem. You should think of your audience as people who care deeply about the problem you are tackling and who are very smart but unfamiliar with machine learning.

The write up should be 10 pages double spaced in size 12 Times New Roman font.

Your goal in the write up should be to demonstrate the value of your methods and model to the people who care most about the policy question. This will include clearly explaining how your model works, how it differs from current approaches, in which ways it is better, in which ways it may not be better, and what further analysis should be done to determine whether or not to implement your model at scale.

In addition to the aforementioned points, please be sure to address the following points:

1. The motivation for the problem you are tackling
2. The data you used
3. The process you used to clean the data, including any new features you added
4. The methods you applied, including:
    - Methods you tried but did not use and a discussion of why you thought to use those methods and why you did not choose them as your final model
    - A clear explanation of the model you used that could be understood by a smart person with no background in machine learning
    - A discussion of your results, including numeric results and a numeric and thoughtful analysis of bias and fairness in your model
    - A discussion of what you would expect for out of sample performance of your model
5. The conclusion you drew from your model
6. A discussion of how your model differs from existing approaches to tackle the problem at hand, which may be both qualitative and quantitative
7. Recommendations for implementing your model in the setting you address

The grading rubric is attached.

For PhD students, the final project serves as an opportunity to get initial results for a research project that will hopefully evolve to a published paper and chapter in your dissertation. Please talk to Amy or Prof. Saghafian for guidance in your project. You will have the same rubric, except "Description of final model "that could be understood by a smart person"" will instead be Description of final model that could be understood by a person in your field with no background in machine learning.

Criteria	Points
Problem Motivation	1
Description of Data	1
Description of data cleaning or feature engineering	1
Explanation of reasonable alternative machine learning methods to chose from	1
Justification of final model	1
Description of final model "that could be understood by a smart person with no background in machine learning"	1
Presentation from Results	1
Discussion of results Including "a discussion of what you would expect for out of sample performance of your model", "numeric and thoughtful analysis of bias and fairness in your model"	1
Contribution "a discussion of how your model differs from existing approaches to tackle the problem at hand"	1
Recommendations for implementation	1
Total	11