Project Guidelines - INDENG 242

Applications in Data Analysis - Fall 2022

Daniel Pirutinsky, Assistant Teaching Professor dpirutinsky@berkeley.edu
Industrial Engineering and Operations Research Department

Here are some reminders about the details for the project proposal and the project more generally. The requirements have been previously mentioned in lecture and in the syllabus, but here we provide some more details.

Project Related Deadlines

- Group membership due 09/15 (individual)
- Proposal due 10/13 (1 per group)
- Presentations 11/29 (1 per group), we will likely need **two** lecture days to have enough time for everyone.
- Final Project Report due 12/13 (1 per group)
- Peer Contributions Report due 12/13 (individual)

Project Content Details

The final project provides an opportunity for students to apply analytical methods to a problem in a domain of their choosing. You will gather (and clean up) data relating to your chosen problem and use the data analytics techniques discussed in class to solve/answer one or more substantive problems or questions. The project will give students some experience in the kind of work that a data scientist might perform in practice.
The broad requirements for the project consist of the following:

1. Motivation: What problem are you solving? Why?
2. Data: What is the nature of the data used? How did you collect and process the data?
3. Analytics models: What methods did you use? What were the results? How confident are you in your results?
4. Impact: What is the (potential) impact of your work with regard to the problem that you are trying to solve? How might you expand the scope of your analysis to improve its impact even more?

Of course, during the course of your project, some elements of the above may change. For example, during the data collection phase you may realize that it is best to adjust the overall scope of your project. You may find that your models do not perform as well as you had hoped. You may perform

some additional data collection/analysis in between the presentation and the due date for the final report. These types of adjustments are fine. We will evaluate your final report and presentation based on your originality and effort regarding the above four points, as well as the soundness and completeness of your analysis.
Here is are some example projects to give you an idea of what we are looking for:

- Predict code blue events in an ICU, improve response times
- Predict the winning team in the video game DOTA 2, and then optimize hero selection strategy
- Predict which Kickstarter proposals will be successful
- Predict energy demand and then optimize solar panel placement

Note on Data Sources

If you are considering utilizing a simple dataset (e.g., .csv file(s)) retrieved from Kaggle, UCI, or another source, then please note the following. These sites often provide both a completely cleaned up dataset as well as a clearly defined problem statement. As mentioned above, a portion of your grade is devoted to data collection and processing and another portion is devoted to motivation and impact (which involves defining your own problem statement). Therefore, if you simply use the dataset and problem statement provided by the site without additional modifications, then you are in danger of not receiving full credit on these parts of the grade. If you do elect to use one of these datasets, then we recommend that you make sure that your project involves at least two modifications. First, you should modify/extend the problem statement presented by the site so that your project at least partially addresses an extension of the problem presented by the site or perhaps an entirely different problem altogether. Second, you should make sure that your project involves some nontrivial component of data processing. One way to achieve this is to collect some additional data to augment your analysis. This additional data may be in the form of additional features that you augment to your dataset, a new dataset set of related observations to form a comparative study with the original dataset, or some other extension that you think of and discuss with the course staff.

Group Membership

By 09/15 you are required to confirm which group you are in by adding your name to the group membership google doc, Group Sign-up sheet.
Groups may consist of 6 or 7 members. You can use the pinned 'search for teammates' post in EdDiscussion to form a group. If you have trouble forming/selecting a group, please reach out to the teaching staff and we will try to assist you in coordination.

Proposal

The proposal is not graded but is an important opportunity for you to get feedback on your project idea. By 10/13 each group must submit a one-page proposal that outlines a plan to apply analytical

methods to a problem you identify using some of the concepts and tools discussed in the course. The proposal should essentially consist of the exact same components of the project:

1. The problem
2. The data that you have or plan to collect to solve the problem
3. Which techniques you plan to use
4. The impact or overall goal of the project (if you could build a perfect model, what would it be able to do?)

The teaching staff will be available to answer questions and will provide all students with electronic feedback.

Presentation

Project presentations (in PowerPoint or pdf format) will be given live on 11/29. There will be one presentation per group. Depending on the number of groups we may need to reserve the previous lecture day for presentations as well.
Here are some more details about what the presentation should consist of:

- The presentations will be 5 minutes in length.
- Consider the presentation as a five minute "pitch" of your project idea and what you have done so far.
- The analysis does not have to be complete at this stage, you will still have about two weeks left before the report is due. However, you should have done some work in data processing and at least run some initial models by this point.
- The presentation should entice us to read your report to see how the "story" ends.
- The presentation should also adequately address the four components listed in the proposal (adjustments are fine).

Final Project Report

The final project submission will consist of a written report of at most 4 pages (not including appendices) that describes the analysis. By 12/13 each group must submit one final project report. There is an additional individual requirement. All students must submit an individual Peer Contribution Report due by 12/13. This will be a summary of you and your group members contributions to the final project. More details are given later in this document.
A few other important points about the report:

- The report should be approximately 4 pages or so of text/main results.
- This is not a hard constraint, but rather view 4 pages of text as a target.
- You may use an appendix for additional figures, graphs, results, etc.

- Do not sacrifice readability by stuffing everything in the appendix, but please keep things reasonable.
- Tell a complete story that summarizes what you have done (successes hopefully, but also possibly failures)
- Please also submit your code (Python, R, etc...) and instructions for how to reproduce your results.

Peer Contribution Report

Each group member must individually submit a Peer Contribution Report due by 12/13. This should be a 1-page list of all the group members along with the following information about each member (including yourself):

- A brief one or two sentence description of what each member contributed to the project (e.g. Wrote some portion of the code, cleaned the data, helped come up with the problem, solution, etc.)
- A rough approximation of the percentage of 'work' each member contributed to the project.

This information will be used to ensure that all group members are contributing in a fair way to the final project. We may use this information to adjust individual final grades, and/or reach out to you to discuss any potential major discrepancies.

Grading Details

The presentation will constitute 20% of the total project grade, and the final report will constitute the remaining 80% of the total project grade.
With regard to the final report grade (which makes up 80% of the total project grade), as a guide, you may consider motivation and impact as roughly 25% of the report grade, data as roughly 25%, and analytics models as roughly 50%. However, we may not strictly adhere to this guide by providing additional flexibility to your benefit. For example, if data collection/processing was particularly onerous for your project then we will consider that appropriately when judging your overall grade.
Your individual final grade may be adjusted, and/or the course staff may reach out to you based on any discrepancies or issues highlighted in the peer contribution reports.