Project Final Report

Duke Computer Science 216 – Everything Data, Spring 2020
Due by 2:00 pm US Eastern Time on Monday, April 27, 2020.

General Directions

The final report should demonstrate a systematic investigation of your data that addresses the research questions set out for your project. The written report should be accompanied by any data, code, or other supplementary resources that demonstrate your effort. The written report should contain at least four parts, which we define below. In terms of length, it should be 5-8 pages using standard margins (1 in.), font (11-12 pt), and line spacing (1-1.5).

You should convert your written report to a pdf and upload it to gradescope under the assignment "Project Final Report" by the due date. Be sure to include your names and netids in your final document and use the group submission feature on gradescope. You do not need to upload your accompanying data, code, or other supplemental resources to gradescope; instead, your report should contain instructions in the methods section on how to access these resources on Duke box, gitlab, or whichever other service you prefer to use for sharing your work.

Part 1: Introduction and Research Questions

Your prototype report should begin by reintroducing your topic and restating research questions from your proposal and prototype. You can start with the text from your proposal or prototype, but you should update your introduction and research questions to reflect any changes in the final project vision, and you should specifically point out what has changed since the prototype. Do not assume that your reader has recently read your proposal or prototype; your introduction should be sufficient to provide context for the rest of your report.

Part 2: Results

The results section of your report should summarize the results obtained in the project. This may include, but is not limited to, results from data scraping and linking, summary statistics, modeling and prediction, clustering, or visualizations. Where possible, results should be summarized using clearly labeled tables or figures and supplemented with written explanation of the significance of the results with respect to the research questions outlined in the previous section. Your results should demonstrate effort and progress beyond your prototype with respect to answering your research questions; specifically, they should reflect an attempt to address the challenges identified in the prototype. Of course, new results may give rise to new challenges or may be inconclusive in some way(s); you should discuss these challenges or limitations in the discussion section below.

Part 3: Methods

Your methods should explain in specific and detailed terms exactly what data were used and how the results from the previous section were generated. For example, if you used a Bayesian inference model to determine certain probabilities, you should precisely specify the model. If you scraped multiple web databases and merged them into a working database, then you should explain how you did so. Your

report itself should include an explanation of your methods, but it should also contain instructions on how to access your full implementation (that is, your code, data, and any other supplemental resources like additional charts or tables). The simplest way to do so is to include a link to a Duke box folder (box.duke.edu) containing all of your materials, but you can also use other services like gitlab (gitlab/cs.duke.edu/) if you prefer.

Part 4: Discussion

Your discussion should explicitly connect your results to your research questions. What insight do your results provide, and what new questions do they raise? Also, what new challenges have you encountered since the prototype, and what are the limitations of your results? To conclude, discuss ways in which your results could be improved or extended, either to address other questions or more convincingly answer your own research questions.

Rubric

The final report will be evaluated on the following rubric. The final report is worth 50% of the overall project grade.

- 100% - Exceptional. Addresses all of the above components with remarkable clarity and insight. Has substantial, feasible, and relevant research questions; the results demonstrate above and beyond effort and provide persuasive and insightful answers to the research questions. Methods are exceptionally sophisticated, appropriate, and well explained and documented. Discussion demonstrates deep insight about the successes and limitations of the project and provides compelling directions for improvement or extension.
- 95% - Good. Addresses all of the above components. Has substantial, feasible, and relevant research questions; the results demonstrate substantial effort and provide appropriate answers to the research questions. Methods are appropriate and well explained / documented. Discussion demonstrates awareness of the successes and limitations of the project and provides reasonable directions for improvement or extension.
- 90% - Developing. Addresses all of the above components but has some minor areas for improvement. For example, the results may not be very well labeled and explained, or the description of the methods may be somewhat imprecise. May only demonstrate minor progress and effort toward addressing the project goals and research questions since the prototype.
- 80% - Emerging. Addresses all of the above components but has areas that need major improvement. For example, the results may demonstrate only superficial engagement with data sources, or the results and methods may not address the research questions directly. May demonstrate little to no progress toward the project goals and research questions since the prototype.
- 60% - Incomplete. Some of the above components are not addressed at all or only in superficial fashion.
- 0% - No submission, or submission only superficially addresses all components.