Prof. Sergio A. Alvarez
St. Mary's Hall, room S255
Computer Science Department
Boston College
Chestnut Hill, MA 02467 USA

http://www.cs.bc.edu/∼alvarez/
alvarez@cs.bc.edu
voice: (617) 552-4333
fax: (617) 552-6790

CSCI 3346, Data Mining
Fall 2018
Term project report guidelines

So, you're finishing up the work for your term project. Now you need to tell others, including me. The oral presentation will be an important part of that public relations campaign, but, as you know, a written report is also required. This document is intended to provide some guidance for the written report.

Overall organization

In brief, the report should be structured like a research paper. See the many papers linked from the Data Mining course site for examples. The elements described below should be included. The page counts in parentheses are suggestions. Feel free to deviate from these suggested lengths. The use of figures, subsections, and selective highlighting is encouraged, since they facilitate rapid visual parsing of the text and can help communicate key points more effectively.

Specific elements to be included

1. Title and authors (group members). If specific responsibilities were assumed by different members, note the details in a footnote on the title page.

2. Abstract (200 words or less). The abstract should provide a synopsis of the work. It should be a more formal version of an "elevator pitch", that is, a brief description of the project's main aims, challenges, and accomplishments (conclusions), without entering into too much detail. The abstract should convince the reader that the report is worth reading.

3. Introduction (2 pages or so). This section sets the stage by explaining the context of the problem to be addressed, why it is important, and what challenges must be confronted. The introduction should cite references that an interested reader can consult in order to learn more about the history of the problem, different manifestations of it, and published work that proposes particular solution approaches or methods.

The introduction also brings the reader up to speed on the basic terminology, and describes any nuggets of prevailing wisdom in the field (e.g., "recommendations based on item content and user relationships combined is usually assumed to lead to better performance than either of the two sources alone"). Again, cite pertinent references.

4. Methods ($3-4$ pages). Describe the methods selected to address the main problem. The rationale behind any design choices should be discussed. That is, state not only what methods you selected, but also why you selected those and not others. Are you addressing a classification problem? Regression? Clustering? What features of the target problem suggested a particular type of predictive model? Include pseudocode descriptions of the main algorithms. The methodology used to evaluate performance should also be described. What specific experiments were carried out? Give specific metrics (e.g., accuracy, area under ROC, RMSE, fpr, fnr, custom cost, running time, memory) to be used, and explain the reasons for your choices in terms that relate to the target task. Explain how you insured against overfitting and overly optimistic performance assessment (e.g., cross-validation, independent test set, etc.), and how you tested for statistical signifiance (p-values). State what software platform(s) were used for experimental evaluation.

5. Results ($3-4$ pages). Present the results obtained in the experiments that you describe in the Methods section. Include tables of values (three significant digits is plenty), as well as visualizations of results (e.g., plots of error rate as a function of hidden layer size). Point out salient findings (e.g., running time increased linearly with the number of hidden nodes, or error rate decreased up to $k = 3$, and then stabilized).

6. Conclusions (1 page). Synthesize key findings from the Results section. How do these results compare with your expectations going into the project? Were your expectations confirmed? Any surprises? What are the "take-home" messages? What work would you suggest as a natural next step that starts from where you're leaving off?

7. References. Provide a list of references that are relevant to the proposed project, and that you cited in the main body of the report. These could be books, papers, or web pages, though most should have appeared in refereed professional venues (as opposed to, say, personal blog entries). Provide full publication information for each (authors, title, venue, volume and issue number, date, pages, publisher).

8. Appendices. If you wish, include system block diagrams, code listings, calculations, and any other details that might clutter the main text too much but that a reader might need in order to follow the details of the main discussion.