ISyE 7406 **Data Mining & Statistical Learning**

The Final Report of Course Project

The **final written report** (14 points) of your team project is due at Canvas (One submission per team). Either word or pdf file is fine.

1. In your writeups, we expect clear explanations of models chosen, hypotheses tested, and findings analogous to what you would produce for a consulting project. Code is mandatory for the report; a 20% of the grade (2.8 point) penalty will be applied if code is missing (Please attach the code as Appendix at the end of the report; no page limit on the Appendix)
2. **Mandatory subsection in the final written report: the lessons you learned** (you can also use other related names for this subsection). For the purpose of this class, at the end of conclusion section of your final report, please **add a subsection for lessons you learned** from this project or this course. You can also write any suggestions to the instructor. The instructor/TAs will read this subsection, so that we can improve our teaching in the future.
3. **Peer evaluation form of your teammates** if there are two or more students on a team, each teammate should also independently submit the completed **peer evaluation** form at Canvas (at the same due date of the final written report). The peer evaluation of your teammates will be private and submitted individually as in the homework or quiz. If you conduct the project individually by yourself, you do not need to submit the peer evaluation form.

   This peer evaluation is to discourage free ride, and allows the Instructor to adjust an individual student's score based on the teammates' peer evaluations if needed. In general, the ideal is for all team members to receive the same grade on the final project. However, individual deductions from the team's final project grade will be assessed for failing to contribute a fair and significant share to the team's project, as determined by the teammates' peer evaluation and the instructor.

Peer Review Comments: each student will be required to review three other projects and provide detailed peer review comments.

Report Grading: The TAs will assign the grades to all projects, based on the following grading guidelines (please note that the TAs keep the right to assign any other reasonable grades). At the high level, the grade will depend on the team's selecting and adhering to a logical and readable format for the report; on the appropriate use of whatever data mining technique the team uses; on the appropriateness in the conclusions of the report; and on the readability and understandability of the report when technical material is needed.

- 14 points (=100%,A+) if you think this is an outstanding project that deserves possible publication
- 12.6 points (=90%, A) if you think the written report is excellent and the project is solid from the technical viewpoint
- 11.2 points (=80%, B) if you have some minor concerns on the project or report
- 10.8 points (=70%, C) if you think the project contains some critical technique errors or the report has poor presentation
- 9.4 points (=60%, D) if you think the project or report is not understandable
- 0 points if no submission

Note that the Instructor/TA will read the mandatory subsection on the lessons learned, so as to improve our future teaching. It is okay that you can write any negative or positive comments on the projects, courses, TA or instructors. We keep the right to deduct 2 points if the team misses this mandatory subsection on the lessons learned, but we will always give you credits as long as your team writes *any kinds of comments, negative or positive*.

As always, if you or your team has a concern about peer grading, please feel free to let the instructor/TA know asap at piazza: we will read your report on our own to make a final decision, although please do understand that ultimately the grade on the final written project will be subjective.

The main body of the final written report is generally 5 ∼ 12 **pages**, and the report shall be **not be longer than** 25 **pages** including the appendix. If you have more materials, you can put them in a separate supplementary file (e.g., more plots/tables, and computing codes or data, etc.).

In general, you want casual readers to quickly understand the main body of your report. For that purpose, only very relevant plots and tables shall be included in the body of the report (please do not copy and paste Python/R codes in the main body of the report!). Some useful materials can be included in the Appendix, and the rest can be put in the supplementary file. To give you a bad example, in the past years, one student submitted a report of 100 pages with lines of python/R codes, and do you think whether TAs or instructor will be able to read through the report to find the main message or not?

When writing up your summary report, it is useful to ask yourself the following questions: What is the work? Why is it important? What background is needed? How will the work be presented?

Here is a suggested format for your summary report.

1. **Title Page**: Project Title, author(s) (your name, the last three digits of your student ID, and email address), the submission date, course name/number;
2. **Abstract:** informative summary of the whole report (100-300 words).
3. **Introduction** includes problem description and motivation, data mining challenge(s), problem solving strategies, accomplished learning from the applications and outline of the report.
4. **Problem Statement or Data Sources**: cite the data sources, and provide a simple presentation of data to help readers understand the problem or challenge(s).
5. **Proposed Methodology**: explain (and justify) your proposed data mining strategies.
6. **Analysis and Results**: present *key findings* when executing the proposed data mining methods. For the benefit of readability, detailed results should be placed in the Appendix. Reference of computer softwares to implement your proposed data mining methods (even it is a web page) should be given.
7. **Conclusions:** Draw conclusions from your data mining practice. Unfinished or possible future work could be included (with proper explanation or justification).
   *__A Mandatory Subsection of "Lessons we have learned":__ *at the end of conclusion section, please add a subsection for lessons you or your team learned from this project or this course. Please feel free to write any comments/suggestions/remarks, or share your experiences of data mining.*
8. **Appendix:** This section only includes needed documents to support the presentation in the report. Feel free to divide it into several subsections if necessary. Do NOT dump all computer outputs unor- ganized here.
9. Bibliography and Credits.

Parts 3-6 constitute the main body of the paper for your primary audience. Usually, as with fictional boss in this example, your audience is intelligent but unschooled in Data Mining or Statistics. So these parts should have as little technical material as you can possibly get away with.

It is appropriate, and even recommended, to refer the reader to the appendix in part 8 if you need to provide a more technical explanation for something. Part 8 is your secondary audience - me - and should follow closely enough the "story" of parts 4 – 6 that it is easy for me to see what technical material backs up with results and discussion.

It is not necessary to number these parts 1-9 or name them as-above-mentioned. Please feel free to merge some parts or provide more informative section names if it seems natural to do so.

A good on-line resource for writing reports is http://www.ccp.rpi.edu/. This site has links to writing centers at universities around the country, many of which in turn have pages that describe how to put together different types of reports.