Project Guidelines

Everything is due by the end of the day 4/11 (11:59pm EST). You can put your report (PDF version please!) on Github. Your project report and code should be uploaded to your Github repository. See the next page for the structure of your repository.

Your project report should document what you have done including capabilities and limitations of your project. Your report should contain a short (max of a half page) section on related work. Related work involves academic/industry papers that are doing something similar to your project. Citations should be included. Your report should also contain a short plan (max of a half page) of what you will do for Part 2 of your project. You can include one small screen shot of the project that should **not** take up more than a half page! All code listings and additional screen shots should be put in an appendix and do not count towards the minimum amount of pages for your report. Make sure any screen shot and code listings are high quality and can be clearly read! As I mentioned at the beginning of the course, the minimum length of your document corresponds to your team size (>=2 pages per team member). Projects will also be graded based on code you will submit on Github. You should include a README that contains general comments about what each programming file does. The projects will be evaluated on at least the following three criteria, each roughly treated equal for your grade.

1) Cleanliness: You code should contain (right above each function) comments of what each function. Your code should significantly minimize hard coding. For example, directories, important parameters to the ML algorithms, etc. should be passed.

2) Completeness: Your project should be complete and self-contained. There should not be any large glaring gaps in the project. Based on your goals, it should completely automate those goals end-to-end from input of the dataset to output of the results.

3) Complexity: The project should do something non-trivial. Higher grades will be given to more complex projects.

The code you submit should be modular. Each module should have its own script, and each script should be executable as follows `python script.py args`

This course is about advanced analytics and automation highly important, especially when dealing with large data sets. So any "manual" aspects (e.g., having to go into Excel) to perform parts of your project are highly discouraged and should be removed. Your project report should be edited for grammatical and spelling errors. Reading the report out loud helps in spotting grammatical errors.

HINTS:
- Use full names of all authors
- All figures should be high quality. Your figures should not be fuzzy!
- Include several full sentences under each figure, table, and graph that explains the figure, table, or graph completely. The reader should not have to read the paper to find out what your figures, tables, and graphs are trying to depict!
- Use decent sized fonts for your x and y axis labels and your legend labels!
- Make sure your text is fully justified on both margins.
- Your figures, tables, and graphs should fit within the margins! Don't let them go over the margins please.
- Please do *not* include an over abundance of equations! If you want to include equations, add them to an appendix.

Your project and code should be uploaded to your Github repository. Your repository should be in the following format.

- README.md : description of the project
- documentation/… : your report (report-part-1.pdf) and any extra documentation
- scripts/… : scripts shell and python, examples:
  1. python: train a scikit-learn model or transform dataset from CSV format to your own format, ….
  2. shell: run tests, launch server, …
- modules/… : python modules when multiple python files are involved to declare classes and functions, they form a module. A module is python code that is loaded by scripts that use them.
- gui/… : HTML/JS/CSS
- data/… : additional data. *do not* add the original datasets or trivially transformed versions of these datasets. What goes here is data that are specific to your project. If you have transformed the original datasets we provided, please include in the directory "scripts/" the code that does this. Note: Your project should work with any dataset provided the uses the same format as the original dataset (within limits that you can set: size limits, fixed or variable labels, fixed or variable features)