

MSAN 630
Advanced Machine Learning
Final Class Project – Final Paper & associated code
Due 3/15

First, reread the final project assignment so you can remind yourself of what I am looking for, and reread the submission you made for the project proposal & status updates as well as any feedback I've given you to date.

A paper describing your project is due Sunday, March 15, and must be formatted in LaTeX. It should be the equivalent of a 6-8 page single-sided paper using regular (12-point) font size and regular (one-inch) margins. The audience of the paper, in addition to Cindi, is someone knowledgeable in the basics of machine learning: you should view this as an opportunity to write a paper that you would be proud to post on a blog or give to a potential employer. Do not assume the reader has read your original project proposal. Here are the components you should aim to cover in your paper (these don't have to each be a separate section of the paper – the organizational layout is up to you):

- Abstract: provide a 1-2 paragraph summary & preview of the paper.
- Introduction, Overview and Motivation: Provide an overview of the project goals and the motivation for it.
- Related Work: Anything that inspired you, such as a paper, a web site, or something we discussed in class.
- Description of your data set and how it was obtained.
- Description of any pre-processing steps you had to take to get the data into shape for EDA
- Exploratory Analysis: What visualizations did you use to look at your data in different ways? What are the different ML and data processing methods you considered? Justify the decisions you made, and describe any major changes to your plans that resulted. Include charts as needed, but an appendix is also appropriate if the graphics are tangential to your main story line.
- How you chose which features to use in your analysis; did this evolve over time?
- Details of the machine learning performed, including how you selected your models and validated them.
- Experimental results, training / validation curves, etc. Again, charts are typical here.
- Your challenges and successes
- Optionally, possible extensions or business applications of your project
- Conclusions and key lessons learned
- Appendix if needed

As this will be your only chance to describe your project in detail make sure that your paper is a standalone document that fully describes your process and results. Do not underestimate the importance of writing the final report. Writing the report is just as important as exploring data, applying models, writing code and running the experiments. Even if you get very, very good results, they will have no value if you are unable to communicate them in writing.

The paper will be graded on the basis of technical soundness, organization, clarity of presentation, grammar and style, and other factors listed below in the rubric. You may consider scanning some papers published in one of the major national or international conferences (e.g., ICML, NIPS, SDM, ICDM) as a model for your paper. You may consider using a conference LaTeX template, such as AAAI's latex macros available [here](#).

Review your BCA notes & tips!

Rubric for paper:

- Presence of the expected content listed above
- Technical quality and justification of applications of the algorithms chosen, feature engineering, and evaluation
- Overall organization clarity & quality
- Quality of grammar / English
- Appropriate use of figures, tables, appendices
- Overall professional appearance and appropriate length of paper for complexity of problem attacked

The advice from Jeff's time series project carries over here and is repeated with some modifications below.

You will lose at least 5% off of your final grade for this project for each violation of any of the following:

- Do not write up a list of bullet points instead of writing an actual paper that flows and is a pleasure to read.
- Title and caption pictures clearly.
- You are forbidden from taking screenshots of images of R or Python modeling outputs and inserting those images into the paper. You should properly write out descriptions of what you did in R or Python (or other). You may of course use ggplot or matplotlib to create figures to illustrate the points you are making in your paper.
- Any accuracy outcomes, confusion matrices, or other raw numbers should be recreated as tables in LaTeX.
- Place adequate text between figures and tables. In other words, I don't want to see stacks and stacks of images without any filler between those images. Limit the use of images to one per page.
- Do not include raw data in the paper. Review your EDA & BCA class notes for ways to summarize data in more meaningful ways.

Implementation, etc.

The remainder of your grade will also be entered in this part of the project "assignment," for Appropriateness & Ambitiousness, EDA, Implementation, and Experimental results (see the original project description). To help me assess this, either provide me with a github repository containing your code, place the code in my Dropbox folder here:

https://www.dropbox.com/sh/nm9gct6ndlb7cny/AAAcyfHBATSIaK6dw2q_GHdma?dl=0, or email me a zip file. **Please do not send me any data, especially for Kaggle competitions**, but if you used additional data sources for your analysis, be prepared to provide them should I so request.