

## Final Project

*Instructor:* Andrew Nobel

*TA:* Alexander Murph

**Project Overview:** The overall goal of the project is to apply a new learning classification technique to answer a research question. The formulation of this question, the learning procedure you will use, and how you interpret the results, are entirely up to you and your team. Since the data you use will be from the real world, there is the expectation that you perform a preliminary exploratory analysis as well as any necessary cleaning and reformatting necessary to apply your procedure. The following submissions will be expected.

1. **Project Proposal:** Due 04/16/2020; no more than 1 page in length. This is expected to be a brief of your project plan, outlining your teams choice of data, research question(s), and learning procedure. Besides explicitly stating each of these points of interest, this proposal should include a summary of your data that discusses its origin, who collected it, why it was collected, and what predictors are available. If you cannot explain what a given predictor measures then you may not use it in your analysis. You should also include a top-level explanation of whatever learning procedure you have chosen.
2. **Presentation:** Due on the final class; all group members must be present. Presentations are expected to be 10 minutes in length and should include well-thought-out slides. They should be rehearsed extensively beforehand to ensure they last for the proper amount of time. Each member of your group should speak for an approximately equal amount. You will be expected to take questions from the instructors following your presentation. Grading will be based on clarity, thoroughness, the degree to which contributions seem equal, and ability to answer questions accurately and concisely.
3. **Final Report:** Due 04/24/2020 by 11:30pm; 8-12 pages in length (including graphics). This should include the following sections:
  - *Introduction:* Should include a detailed discussion of your data and your predictors, why your team chose this data, and what research question(s) you will investigate. You should also include a discussion of any data cleaning that was necessary, and any pre-processing that you did.
  - *Learning Method:* Should outline which learning procedure you plan to use and discuss why it is relevant to your project aims. This section should include an explanation of what your learning procedure does. While this explanation should be in your own words, you should also properly cite any resources you used in your research of the procedure. You should be able to discuss specifically what your procedure does, how it does it, and how it can be used for your classification aims.
  - *Results:* Should include detailed, informative visuals with a discussion of how you got them.
  - *Discussion:* Should interpret the visuals you created in the context of your research question.
  - *Conclusion:* Based on your analysis, what were your conclusions?

<i>Submission Item</i>	<i>Weight</i>	<i>Data Suggestions</i>	<i>New Learning Procedure Suggestions</i>
Presentation	50%	Enron Emails	Random Forrest
Final Report	50%	Twitter Sentiment	Decision Trees
		Wine Quality	Neural Network
		Pima Indians Diabetes	Naive Bayes
		2017 NBA Teams	Boosting

If there are any data or learning procedures that you would like to use for this project that are not listed above, you may do so with prior approval from Murph. Email him at least 24 hours before the Project Proposal is due.