Heinz 95-845: Project Proposal

Firstname Lastname   ANDREWID/EMAIL@ADDRESS.EDU

Heinz College
Carnegie Mellon University
Pittsburgh, PA, United States

Firstname Lastname   ANDREWID/EMAIL@ADDRESS.EDU

Heinz College
Carnegie Mellon University
Pittsburgh, PA, United States

1. Project Details

Your project will involve the use of the machine learning pipeline. This is an opportunity for you to explore some interest you have in an applied domain and the machine learning suitable for the task.

The purpose of the project is to conduct an analysis that is novel in some way. The novelty could be in terms of development of machine learning, the assessment of a wide variety of machine learning algorithms at a focused task, or the application of a single machine learning algorithm that solves a real societal problem.

A list of exemplary papers (in health care) are available in the Possible Data Sets slides on Canvas. The examples may be helpful in identifying how you conduct your study and prepare your write-up. A much larger list of public domain data sets is available at the url: `https://github.com/awesomedata/awesome-public-datasets`. We recommend you do not choose a fully pre-processed data set. We do recommend you choose a data set that will fit in memory (or that can be run on your laptop) so that your machine learning process will be manageable.

The proposal for the project is due on **March 26th**. Please use this TeX template in Section 2 and submit on Canvas **a link to a git repository with instructions on access (particularly if it is a private repository)**. You may find the online editors ShareLatex or Overleaf helpful in drafting your TeX file. However, in order to learn git version control (which will help you checkpoint during your project), we require the submission to be in a git repository.

1.1 Objectives

The objective of this project proposal is to generate a proposal for your course project. It should be concise and describe the following components:

- Construction and description of a research framework that motivates the use of machine learning for your task

- Presentation of machine learning techniques appropriate for the task

- Description of the data

- Description of possible limitations of the study

- Description of the likely analysis outcomes and their impact.

1.2 Parameters

The project will be conducted in pairs. **If you intend to work individually, because of data privacy, computational frameworks, or other reasons, you must use the link on Canvas to do so prior to March 3rd.**.

The project you propose should be different from an existing analysis either in the literature or from other class projects of yours. It is permissible to perform an analysis in data that warrants a second analysis. My guideline here is that the analysis must be greater than 50% new. To get approval for these studies, please describe the existing project and highlight the difference and contribution of this class's project. Provide any relevant documents (proposals, manuscripts, and/or citations). If the project has overlap with work from another course, you must also provide documented approval from the other faculty member/research collaborator(s).

Your team is free to use programming language(s) of your choosing, however, we may only be able to support your endeavors in R.

2. Proposal Details (10 points)

Please provide information for the following fields. Your proposal write-up should be less than 2 pages.

2.1 What is your proposed analysis? What are the likely outcomes?

2.2 Why is your proposed analysis important?

2.3 How will your analysis contribute to existing work? Provide references.

2.4 Describe the data. Please also define Y outcome(s), U treatment, V covariates, W population as applicable.

2.5 What evaluation measures are appropriate for the analysis? Which measures will you use?

2.6 What study design, pre-processing, and machine learning methods do you intend to use? Justify that the analysis is of appropriate size for a course project.

2.7 What are possible limitations of the study?

References