GEB 6895: Business Intelligence
Department of Economics
College of Business Administration
University of Central Florida
Fall 2019

Project Guidelines

Due Friday, December 8, 2019 at 11:59 PM

The objective of this project is to carry out an end-to-end modeling exercise on a topic of your choosing. In teams of two, collect data, analyze it and answer a question that you find interesting. Depending on your interests, the project can take on one of three forms.

Option 1: Regression Analysis

Collect your own dataset and answer a question using regression analysis. The question could use either linear or logistic regression, as suited to the particular application. Be sure to collect a large enough dataset that allows you to split the data into a training and testing sample to validate the results.

Option 2: A Modeling Algorithm

Choose a data-mining or machine learning algorithm and create a report that describes the algorithm and illustrates its use with an example. The goal is to demonstrate the ability to research a new algorithm and communicate its value to stakeholders. Some examples of algorithms include, but are not limited to, Gradient Boosted Machines, Random Forests, Neural Networks, k-nearest neighbors, Support Vector Machines, and Clustering.

Option 3: Testing a Component of a Modeling Algorithm

As another option, the analysis could focus around an improvement to the model-fitting process. For example, the algorithms for regression trees or other classification models, could be augmented by a procedure for bagging (short for bootstrap aggregating) or to implement a particular learning rate. This option is better suited to a student with some previous experience with a particular algorithm and the desire to learn more about the details behind it.

Guidelines:

There are no upper or lower bounds on the page count; use whatever space you need to explain the analysis. However, shorter is better to keep the reader's attention. Browse through some blog posts online to get an idea of the appropriate level of detail and the format. A formally written document in pdf would be acceptable, while a well-documented and illustrated `README.md` file in your repo would be entirely acceptable as well. The goal is to communicate your expertise in an way that allows others to learn from your experience.

As a checkpoint, please provide a proposal for your project by Thursday, October 31, 2019 in the form of a few brief paragraphs outlining your proposed method, the data you will collect and the technical tools you intend to use. Submission can be made by providing a written submission on Webcourses or, ideally, in a README.md file on your project GitHub repository.