IDSC 4444: Term Project

Spring 2016

Project Overview:

Data mining / knowledge discovery project. Use R to explore/analyze/mine a dataset of your choice, then present and describe your findings. You are expected to describe: (a) the data and how / where it was collected, (b) at least one interesting / important question that could be explored / answered using this data), (c) the tools and techniques that you used, (d) the results of your analysis, and (e) any conclusions that you have come to and what actions one might take in response.

Note: Please speak with either myself or the TA about your proposed topic before you begin, to ensure the question you are planning to tackle is appropriate and the dataset is reasonable.

Suggestions for Data Sources:

- Many links can be found *on* KDNuggets
- UC Irvine Machine Learning Repository
- Data.gov
- FedStats.gov

Note: Feel free to identify your own data source, if you wish. Again, just be sure to run your project idea past myself or the TA before you proceed.

Project Deliverable:

You are expected to prepare a written report, not to exceed 15 pages, single spaced, Times New Roman, 12-point font, with 1" margins. The report should be submitted in PDF format via Moodle. Your report should address all of the points laid out in the above description of the project overview. If you draw upon any external references, please make sure that you provide appropriate citations.

Note: 15 pages is the maximum, but this does not mean you are required to write 15 pages, nor should you, unless you feel you need to, to get your point across.

Project Evaluation: (15 points)

- Background / Overview & Business Question (2 points)
    - Provide a background on the business context (i.e., where the data comes from).
    - Discuss / motivate the question you plan to answer in your study.
- Methodology (6 points)
    - Data collection & description
    - Data exploration (e.g., summary statistics, correlations, distributions and so on).
    - Analytical approach (i.e., identify the technique you are using for the analysis, and justify its appropriateness).
    - Data pre-processing (as needed)
- Results & Interpretation (4 points)
    - Present & interpret results.
    - Draw conclusions, suggest courses of action.
- Presentation (3 points)
    - Formatting & Aesthetics
    - Appropriate Use of Plots / Visual Aids
    - Writing & References