CS 443 – Introduction to Data Science
Spring 2015 – Proposals for the Class Project– 100 points
Due 11:00 AM (Eastern) March 12, 2015

Upload (into the Sakai site) a single file, called <Last-Name><First-Name>Proposal.pdf, containing the answers to the following questions. Your file should be **at most** 2 pages long (i.e., ≤ 2 pages).

1. What is the title of your project?
2. Who are the project members?
3. What is the problem definition?
4. Why is it interesting and/or important?
5. Why is it hard? Why have previous approaches failed?
6. What is your intuition about the problem? Give your hypotheses and conjectures.
7. What are the key components of your approach? For example, is your data clean already or will you have to clean and pre-process it? If it is not clean, how are you going to clean it? Are you solving a supervised or unsupervised learning problem?
8. What data sets and metrics will be used to validate your approach? What do you expect to learn? How are you going to show it?

In addition, be prepared to give a 4-minute pitch about your project during class on Thursday March 12th. For the in-class pitch, I will call on you and ask you to go through your answers to the above questions. The reason for the in-class pitch is for other students to see what is being proposed and give their feedback.

Some points to consider for your class project:

- The class project can be done either individually or in groups of two.
- You must have your dataset identified by March 12th. You can find a bunch of datasets at these popular sites. Obviously, this is not an exhaustive list.
  - http://www.kaggle.com/competitions
  - http://datamob.org/datasets
  - http://archive.ics.uci.edu/ml/
  - http://kdd.ics.uci.edu/
  - http://lib.stat.cmu.edu/datasets/
  - http://snap.stanford.edu/data/
- If you are interested in analyzing social media data, I recommending reviewing these
  - http://snap.stanford.edu/proj/socmedia-kdd/
  - http://www.umiacs.umd.edu/conferences/sbp2011/nitin_agarwal_tutorial.pdf (see slides 44 to 72 for data collection APIs)

- Each Web 2.0 site has its own API. Here are some examples:
  - Twitter Archivist: http://archivist.visitmix.com/
  - Delicious' Developer Page: http://delicious.com/developers
  - YouTube's Developer Page: https://developers.google.com/youtube/
  - Flickr's API Gardens: http://www.flickr.com/services/api/
- For details and guidance on the class project, see the other attachment for this assignment. I went through these slides in class.