**CS-522**

**Project Progress Report I - Data**

**Project Progress Report - I Data/Implementation**
Below is a sample outline for the first project progress report. The goal of this report is twofold:
1) Finalize the goal and the scope of your project, including the data that you will use.
2) Make sure that you are all set for the actual experiments after this point and won't have any issues with getting the data, software etc.

Maximum number of pages: **2**.
The report should be informative, include the required details but concise.

**Abstract**
Provide a 5-10 sentences summary of your project. What problem will you try to solve (or, what questions about algorithms you will try to answer). What data are you going to use. What approach you are going to use. This may change later as you get more into the project. But do the best you can right now to have a good start for your work.

1. Data

1.1 Data Source

Describe the data you are going to use – the source, the size of the data, the type of the data (some will use text, some may use financial transaction or crime data). **You should download the data for this report.** Describe the size of you data on disk, discuss if it can be processed in memory. If not, how are you going to work with your data.

1.2 Data Format

Describe the format of the data that you have. Use the tutorials etc from the source of the data if you like, however, do not cut and paste a lot, describe the data format briefly in your own words.

1.3 Programming Language, packages

What programming language are you going to use. If you need any particular packages, for example, machine learning package, or name entity extraction package **make sure you download them for this report**, list the packages that you will use.

2. Preliminary Experiments

2.1 Preliminary Results

Preprocess a small sample from your data to make sure your environment is ready for your experiments. This make include reading a sample into memory. If it is text, running some preprocessing steps we discussed. For other data, performing a couple of steps that you will do for larger data during your experiments. Describe briefly what you do.

2.2 If you have time to do more, describe your experiments