IE 529
Stats of Big Data and Clustering
Group Project

I. Overview:

- This project will involve: (1) reading an assigned journal article or book chapter; (2) writing a summary report of the reading; (3) completing an example computational implementation related to the topic of the reading; and (4) making and giving a 40 minute slide presentation to the class.
- There will be 4 students in each group, with one or two groups possibly having 5 students.
- A list of the articles/chapters will be posted on the course Compass site; notification will be sent when this is posted.
- One day after the readings are posted, a Google form will be opened for students to sign up for their preferred reading.
- Sign-ups will be first-come, first-served.
- **The sign-up form will be open for students to select their readings for 36 hours**; after this time groups will be compiled using the survey responses, and an email will be sent containing the group, the reading, and the presentation date. Students that do not complete the survey in the allotted time window will be assigned to groups arbitrarily.
- Student groups will be responsible for arranging their own meeting times.

II. Summary report - Due Monday, Nov. 7: The report should be 5-6 pages (typed, line-spacing 1.2-1.5) and should contain the following elements:

1. A general summary of the main results of the reading.
2. An overview of the algorithmic, probabilistic and/or mathematical methods used or referenced, with some background provided as needed (e.g., if an approach not previously seen in the course is used, a brief introduction should be given and any technical terms defined).
3. A brief discussion of any application areas either discussed in the reading, or that might be expected by the student group. If there are substantial application results these should be discussed in the summary as well.
4. An overview of the example application planned by the student group, including a description of the data to be used. Student groups should either create their own simulated data sets, or obtain data from a depository online (there are many classic data sets available online).
5. Pseudocode for the example application.
6. An outline or draft of the slide presentation.

The report should be submitted as a pdf file.

III. Presentations:

- Presentations will be scheduled, one group per date, for the dates of November 9, 11, 14, 16, 18, 28, 30; December 2 and 5.
- Professor Beck will set the schedule.
- The latter 4 dates will be for those readings that are *starred* on the reading list; these readings are those that are somewhat more technical or mathematical.
- All students are expected to attend all presentations (role will be taken on these days).
- All students NOT presenting on any given date should submit a one-paragraph summary of the reading being presented that day, at the start of the lecture time.