Many of you asked on Saturday for more clarification on what is required for the text-analytics & recommender systems project.

The overall project goal is to test your practical skills in text analytics and recommender systems. Hence we asked for one recommendation system problem to be solved in two ways:

(i) using a content-based approach, this would use text mining/text analytics to generate the product features

(ii) using a collaborative filtering approach using user ratings (implicit or explicit)

We also asked you to combine the results from the above using some simple method (such as averaging) which we will cover on day3. This shouldn’t take too much effort.

Ideally you can find one dataset that allows all of the above. But if you cannot then some other options are:

1. Make use of text-based product review data. There are many ways to use this type of data but we would expect some form of text mining to be used. Below are some examples of how product review data might be used, note that only one of the below would be required:
   1. generate a set of product features for use in a content-based approach (it’s OK if the products that features can be extracted for are only a small subset of the total products)
   2. derive implicit ratings for the product if no explicit rating is present
   3. filter the recommendations that have been made by collaborative filtering (or other) before they are presented to the user. Additional value-add could be (for example) to extract the key positive and negative product features from the reviews and present these alongside any recommendation to give the user more data to help make their decision
   4. use the review data in some other way to support the recommendation process

1. Make use of other text-based data to support the recommendation process

1. Find two separate data sets  (solve two different problems): one dataset for the content-based approach and the other for collaborative filtering. Clearly in the case you will not be able to generate a hybrid (combined) solution afterwards.

Also note that you do not need to generate a user-interface to your system. It will be sufficient if you upload the code that you have developed (and also your data), plus your report.

When writing the report please make sure that you describe adequately: the data set used, data preparation and pre-processing, methods and algorithms used, your test framework and your test results.  Ensure that you include enough details with regard to testing.