

Assignment 1: Topic Selection

Due date: Thursday, January 4, beginning of class

Assignment

This is an **individual assignment**.

Step 1. Download a PDF copy of Charu Aggrawal's *Data Mining: A Textbook* book. You can visit Charu's web page for the book at

<http://www.charuaggarwal.net/Data-Mining.htm>

You can then download the PDF by following the appropriate link from the page to the publisher's site. When downloading, please make sure you are on Cal Poly campus, or are connected to the campus VPN - the book is available to you through Cal Poly's Springer subscription. You can also download the table of contents as a separate document. For the purpose of this assignment you primarily will be working with the table of contents.

Step 2. Study the table of contents for the book. Note, the table of contents is fairly verbose, with both meaningful long-form chapter/section titles, and names of different algorithms. You can look up specific topics in the PDF you downloaded, but generally speaking just do a thorough pass through the table of contents. Please note, topics in this book have non-linear dependency structure, so chapter number/page number is not necessarily a good predictor for how simple/complex/accessible/interesting/important etc... a specific topic is (and thus you should read the table of contents through the end).

Step 3. Answer for yourselves the following questions:

- **Question 1:** Which topics mentioned in the table of contents do I know sufficiently well?
- **Question 2:** Which topics mentioned in the table of contents I do not know, but would really like to learn about?
- **Question 3:** Which topics mentioned in the table of contents I do not know, but am not particularly interested in?
- **Question 4:** Which Data Mining/Machine Learning topics that I am interested in are **not covered** in the book?

Step 4. Fill out the survey found here: (this link will also be posted on the slack channel)

<https://forms.gle/MPmuUBReJNgwoFsQ9>

Step 5. Make sure you have a copy of your responses to the survey (or written notes). Bring it to class on Thursday, January 6 - you will be comparing your notes with others in the class.

During the Thursday, January 6 class, you will form ad-hoc groups of 4-5 people and will be asked to discuss your responses and create a combined list of topics that you feel have been covered well elsewhere, topics that you would really love to see covered in the course, and topics that, while not covered elsewhere, are not of significant interest to you.