

Introduction to Data Science: CptS 483-04 – Fall 2017

Assignment 1: Create Data Science Profile of Yourself and Reflect on an Article on Data Science

Release Date: 08-29-2017.

Due by: 09-05-2017, 4pm.

This assignment consists of two tasks, the first is about a data science profile of yourself and the second asks for your reflection on an article that was a required further reading for a lecture we had the first week of class.

Task 1

As I explained in class, the purpose of this task is to create a visual data science profile of yourself. Specifically, you will create two instances of profiles. The first will show the way you see yourself now. The second will show how you would like to see yourself by the end of the course.

The profile is simple. On the horizontal-axis you will have seven “areas of skills” that could generally be considered important to Data Science:

1) Computer Science. 2) Math. 3) Statistics. 4) Machine Learning. 5) Domain expertise. 6) Communication and presentation skills. 7) Data visualization.

On the vertical-axis you will have a relative scale (think percentage) of your skill level in each of these areas. The area in which you have the strongest skill will be close to 100, and the area in which you think you have very little skill would be close to zero.

As an example, see the slide in the lecture slides of Aug 25 (posted on OSBLE under Lectures/082517) that shows the data science profile of the author of one of the reference books we use in this course (Doing Data Science). As a context, the author, Rachel Schutt, has a PhD in Stat and has held several senior and executive-level Data Science positions in industry.

Your task is to create your own profile – two to be exact, one showing current and the other projected. You are still a student and you may not feel you have a lot of skill in some of these areas. Allow yourself a generous interpretation of skill level and keep in mind that this is on a relative scale. Also, keep in mind that it is perfectly okay to have zero skill level in some of these areas. For example, if you are a computer science major, it is natural that “Domain expertise” would be the area in which you have the lowest skill level among the seven, and it is okay for it to be close to zero.

You can use any tool (Excel, Matlab, R, etc) you wish to make the plots.

Here are a few associated presentation considerations and discussion points you are asked to address as part of this task.

- 1.a. The areas in the horizontal axis could be ordered in a number of different ways. What ordering in your opinion would be most effective and why? Create your profile in the order you chose. In addition, create your profile in which the ordering is as listed earlier (CS, Math, Stat, ML, Domain Expertise, Communication, Data Visualization). A common way to order

the domains could for example make comparison with other profiles easier, which could in turn help in composing an effective data science team.

- 1.b. Is there a skill (bucket) you think should be added to this data science profile? A skill you think should be removed? Specify and justify briefly.

Task 2

As you recall, we discussed the article “Data Science and Prediction” by Vasant Dhar in class in connection with the topic what is data science (a copy of the article is posted under Lectures/082317). Read the article and briefly answer the following questions.

- 2.a. The author identifies a few ways in which data science is different from statistics. What are those ways?
- 2.b. Imagine you were asked to write a “head-line” (as you see in newspapers) for this article, followed by one or two very telling summary sentences. What would your headline and the summary sentences be?