Tutorial Presentation Assignment – 20% of course grade (20 points/100 total course points)

SI 699 Big Data Analytics
Jeff Sheng, PhD, MS - Assistant Professor
School of Information, University of Michigan

Due Date: Presentation in class Wednesday, February 8.
Deliverables (Materials uploaded in Canvas due: February 7, 11:59pm)

- Readme file (a document with basic instructions on how to set everything up).
- Dataset (link to download should be in the readme file).
- Code script (i.e., python notebook).
- Slide deck for presentation.

Additional deadline: Because of the number of students in the course at 25 where combinations of 4 don't quite work, I am expanding the possible group sizes to 3-5 students per group. There may be some adjusting that we need to do, but groups will be finalized by the end of class on 1/25.

Learning Objectives: In this assignment, students will work in teams to create a coding tutorial and presentation to teach their classmates a valuable data science method that others might want to use for their own final projects. The end goal is to have a shared folder in Canvas with everyone's tutorial as a resource to have for the class and for future projects.

In doing this, students will learn:

- How to take methods from data science research papers and the value of replicating them for self-learning.
- To become extremely comfortable with a specific method and how to use it.
- How to put together a sample dataset that can be run at scale in some way with "big data."
- To refresh their own data science skills in preparation for the final project.
- A set of new data science skills and code that can be used for inspiration or even a second replication for their final project.

Assignment:
In this assignment, you will work in teams of 3-5 to teach a "big data" computational method that is inspired by one of the research papers that have been presented in class (either through the pair presentations or one of the readings assigned for the class). Tutorial Presentations will be in class on February 8. Each team has up to 30 minutes to complete the presentation though it is likely that you should only need about 20-25 minutes. You do not need to copy the method directly as it is done in the paper but may do something that is a variation or inspired by that paper. The main component of this assignment is a set of materials (code, readme, and data) that other students can run on their laptops and an accompanying presentation that explains and helps teach this method (see slide specifics for more information). For the presentation, every team member should actively contribute, but this can be divided between those who present the slide deck while others help classmates setting up their code.

Tutorial Presentation Assignment – 20% of course grade (20 points/100 total course points)

Slide deck format:
The following slides should be included in this suggested order. Some of these, like the part where you go through your code, should be multiple slides. Others, like the title slide, should be just one slide.

1. Title slide: Use a title that is related to your method. Your own names should be here as well.

2. Paper: What research paper is your method based on?

3. Origin story: Why did your group pick this method? (For example, do most of you have experience with it? Did you think you might want to do a final project using this method?)

4. Methods literatures: What previous methods or statistics-based literatures are mentioned by the authors? What years were these published? Look up at least one of these papers and tell us more about it.

5. Data: What data did the paper use? Is your tutorial based on the same data or did you have to obtain a different dataset?

6. Method: What method are you planning to teach us? Describe it at a high level. How does the paper use this method? Is this the same or slightly different than what you plan to show us?

7. Set up instructions/Readme file overview: Have a slide that summarizes the basics of what students should do so they can set it up on their laptops. While some students might not be able to get things working on their own laptops, ideally at least some can follow your set up instructions in class.

*** Break in presentation for everyone to download your materials to their laptops ***

8. Code walk-through (multiple slides): Screen shot your code in run order and walk through the code with us. Each slide should have an explanation of some sort of the code through in line comments (you can also add text separately on the slide itself to help explain).

9. Suggested variations of the code: What extensions or variations could you imagine with this dataset or method?

10. Your own thoughts and impressions about the method: What do you like about it? What did you not like? Where can it be useful? What are its limitations?

Tutorial Presentation Assignment – 20% of course grade (20 points/100 total course points)

Grading Rubric:

- **5 points for following all the instructions (**baseline grade).

- **0-5 points: Success at teaching**. How successful are you in teaching your peers? Do a lot of your classmates find it easy to download and set up the code based on what you provide them? Are people on your team able to help people get set up? Is the readme file you submit clear and simple to understand such that a user can read it and get the code set up on their laptop?

(0 points: Nothing works, and the code is completely broken and/or wrong -> 3 points: Code works but is confusing for students -> 5 points: Flawless teaching, this could be a Coursera course!)

- **0-5 points: Method Selection**. How useful, difficult, or multifaceted are the methods being taught. Usefulness is judged by how beneficial this might be to a wide variety of students. Difficulty is judged by the method being taught – is this a challenging method either through the code being difficult or that it requires a complicated in-depth understanding? Multifaceted is judged by the number of variations you offer of the method being taught.

(Your points here are graded by either depth or breadth. This means, you can either score very high on one of these three factors (depth) or tackle all three and do all three decently well (breadth). For example, if a team picks a very complicated method and can execute the teaching of it well, then that is a 5. If a team picks an easier method, but it is useful to students and there are a few interesting variations of the method that they can use, then this is also a 5).

- **0-5 points: Overall Presentation Score**. How prepared and polished is your team in the overall presentation of both the slides and coding materials.

(0 points: No presentation -> 3 points: Decent presentation but there are some mistakes and seems like the team only rehearsed a few times -> 5 points: This feels like a well-executed presentation by the CEO team at a large tech company, great job!)