Reflection Essay

Ming Chen

Department of Human Development, Teachers College, Columbia University

HUDK4050: Core Methods in Educational Data Mining

Dr. Zhichun Liu

Dec 18, 2021

As a student majored in Cognitive Science in Education, I expected that I would encounter a lot of situations that need me to solve questions based on large sets of data. So far, I used to use SPSS to process data, but sometimes found it limited. I really expected to take a course to learn something more in the data world, to find new methods to analyze data and better ways to depict them.

The course actually brought me more than I thought. Before taking this course, I just wanted to learn some coding skills with respect to analysis methods. After classes, I have understood that coding is just a small part of this field. Actually, as an educational data miner, what is more important is how to link our questions with specific methods that may help to solve them. Coding itself is an endless topic. Because people from different domains keep creating and updating new codes and methods, it is impossible for us to learn everything about coding. I remembered that one time I failed to get an expected result by trying to build a classification through statsmodels. To figure out what happened, I searched for many guidelines for the codes, thinking of how the codes work and trying my best to understand its syntax. I also received great help from Lukas. The problem finally was solved through adding a threshold to its output, which is related to what I have learned through this process (such as understanding how it works and how to change the result by editing the specific syntax). I also realized the flexibility of such programming process. Even if we feel unsatisfied with the results a single code brings to us, we can combine different codes from different package to finally achieve what we want. The final presentations are really interesting and inspiring. I think sometimes I have a mindset about how to organize research, like searching for relevant literature reviews, looking for new variables, and finding some relationships among these variables. Sometimes the process seems a little bit result-orient but not question-orient. My classmates'

presentations are enlightening, I noticed how they find a problem (e.g. new policies, new trends and needs) and how to use different EDM methods to address these problems, this also gave many indications about what these methods can do.

Of course, in addition to the added bonus, I definitely achieved what I expected in the course. Although I have programmed a little through MATTLAB and E-prime (Visual Basic), they are all very field specific (as tools for specific psychological experiments). The knowledge and ability of using general programming tools and popular programming languages are very important for future works, so I felt great to have chance getting familiar with both R and Python, and using Anaconda Navigator to manage the packages and codes. Also, the basic and mostly used methods and their codes are well arranged to learn, they helped me to build a general frame to analyze the data set.

I planned to do some research works in the future, so all the knowledge and tools I learned will be very useful. I will keep on learning new methods and codes while analyzing different data set from different research.