

CS510 Data Science: Midpoint Report

Scott Rubey

Updates to Project Topic

The intent in my original Research Report was to “explore the technologies and ethical considerations behind automation of data ingestion, storage, cleaning, analysis and modeling.” Ethical considerations I had intended to explore were “implicit bias in algorithms, outsourcing and job elimination.” Due to space considerations, I was unable to discuss ethical issues at any length, thus I opted to eliminate this topic entirely from the aforementioned report. Furthermore, I was unable to expand much on the extreme front end of what is entailed in the data science pipeline.

For my final project, I intend to expand upon these topics individually and in tandem (i.e. as they directly relate to each other). Topics for expansion and/or addition include a description of the steps involved in data ingestion and storage, ethical issues such as outsourcing of jobs, elimination of jobs due to automation, and discrimination, along with active research being performed for the purpose of eliminating bias in algorithms and human involvement.

Planned Research

Several additional research papers have been procured through IEEE. They describe case studies regarding AI as it relates to recidivism amongst Caucasian Americans and African Americans, as well as redlining practices in the financial industry (“redlining” refers to the practice of withholding home loans from people of protected classes). Material has also been found that details research pertaining to pre-processing of raw data for the purpose of addressing algorithmic bias.

Team Structure

As with the Research Report, I intend to work alone on this final project.

Update on Progress

Progress to this point has been slow but steady. I completed my Research Report about one week early, at which point I researched further sources. As of today (Monday, July 20), I have completed my research, compiled my notes, and am in a position to begin expanding upon my paper. My intention is to complete roughly two pages per week through Week 7, at which point I will prepare the slideshow for my final presentation.

My first task at this point is to expand my original paper to discuss ethical issues. This new section will consist of three subsections. First, I intend to discuss the ethics of automation as it pertains to job loss and outsourcing. Secondly, I will discuss the ethics of automation as it pertains to discrimination and algorithm bias. Thirdly, I intend to expand upon my discussion of algorithm bias with an examination of solutions currently being researched.

Following is an update on each of the milestones outlined in my original Project Plan:

- 1) Gather sources – complete, but ready to expand if necessary.
- 2) Complete reading of initial 4 sources, compile notes – completed early Week 3.
- 3) Complete initial 4-page research report – completed early Week 4.
- 4) Read / watch additional sources, compile notes – completed early Week 5.
- 5) Flesh out 4-page research report with 5 additional pages – commenced work during Week 5.
- 6) Prepare presentation materials (slideshow) – incomplete.

Day/Time of scheduled midpoint meeting

Thursday, July 23rd at 2:10pm.