Recap of Part II: Applying Data Science

Reaching the end of Part II, we hope our detailed discussions of the initial six examples and the somewhat more cursory discussions of 26 more have been enlightening.

- While we have not surveyed all possible applications, we hope we demonstrated data science's valuable potential in domains ranging from research to entertainment to medicine to commerce to finance to government, and more. In many cases, we also explained a bit of their "how-to" and demystified how many applications operate.
- We showed that data science is often applied in different ways to multiple subproblems of an application. For example, in a video streaming application, we illustrated copyright identification, video recommendation, search, and advertising. Video streaming could also benefit from additional applications of data science, such as for closed captioning, language translation, summarization, and more.
- By our repeated referral to the Analysis Rubric, we illustrated seven important considerations in applying data science to a problem. We believe that examining an application in detail against the Analysis Rubric elements teases out its design's hard parts.
- Whether contemplating a new data science application or evaluating an existing one, careful consideration of data science's unique and complex aspects is essential. While the Analysis Rubric elements are of necessity listed in a particular order, we fully recognize that the application of the Analysis Rubric may be done in a more bottom-up or top-down way, depending on the application.
- Our analysis shows a few applications to be straightforward, many to be challenging, and some to be very hard or perhaps presently impossible. When difficulties occur, they are usually because:
  - Data capture of sufficient quality and scale is impossible for some reason.
  - Existing technical approaches are insufficient.
  - The cost of achieving dependability is too high.

- Opaque approaches in either sense of the term are insufficient; furthermore, there may be a particular need to prove causality.
- Objectives are in dispute.
- Failures are essentially intolerable.
- Insurmountable difficulties arise from the ELSI criteria.

Finally, we resumed our ethics thread. We explored how the principles set forth in Chapter 3 can guide us when applying data science to gnarly problems.

All of this sets the stage for Part III, a considerably deeper discussion of data science challenges.