Brandon Roten

INFO 3401

Problem Set 4

**1. Peter Naur is famously quoted as saying data science *“deals with the data, while the actual relation of data to what they represent should occur in other fields.”*What might be problematic in this statement? Why do you think he’d choose to frame data science this way?**

He might have chose to frame it this way because it is more of a specific separation between these fields. This could be problematic because it creates a confusing connection for the use of data and the connection to data science. I believe he chose to frame it this way because it is hard to describe the immense impact that data science has in the various fields which it is used.

**2. There was a substantial shift in the ways we define data science between the 1970s and the early 2000s. Describe this shift and why it may have emerged.**

The emergence of the internet in the early 200’s lead to a huge increase in data. It is also acknowledged that there is a presence of data science in almost all fields of science. There can also be an inclusion of legal issues as well as publication. We are also seeing the effects and consequences of data.

**3. The idea of "big data" dominates much of modern data science. However, data is still growing at an exponential rate.**

**A. What factors do you think may have led to this growth? Mention at least three and describe why they have contributed to recent explosions in data volume.**

One factor that may have contributed to this growth is the explosion of technological device creation and use. This has contributed because there are more data that is made available by a larger user base. There is also a strong presence of user research for advertising. This is done by collecting vast amounts of data on users. The presence of social media has also contributed to this huge increase in data growth. Social media collects data on all of its members, and that is millions of people. People are also starting to rely on data much more in today’s world.

**B. Where is this new data coming from?**

The short answer is everywhere. This data comes from any information or exchange online. It can be collected by anyone anywhere. It is scary to think about this, but everyone produces data, and it is being used by somebody, somewhere.