Topic: Summary – Seminar talk by Matthew J. Salganik
Date: 11-30-2016
Class: INFX 573 – Extra Credit
Name: Pierre Augustamar

Introduction

"Bit By Bit: Social Research in the Digital Age" is the title of a new book written by the speaker, Matthew J. Salganik. The talk was primarily based on lessons he learned while writing this book. He started by providing a timeline on how data changed from analog to digital. He emphasized the impact that both a social scientist and data scientist have on data behavior. Also, he explained the best ways that research can be done. Finally, he completed with a plea to everyone who is dealing data to have transparency on why and how they are using the data. I will attempt to provide key notes from this talk.

Social Scientist Vs Data Scientist

The speaker Social scientists see the glass half-full while data scientists see the glass half-empty. In other word, data scientists are more optimistic about finding something from the data while social scientists are very much pessimistic. He explained how social scientists would make use of a

model generated by computer scientists to make assumptions. However, as he mentioned there is an intrusion of computer scientists into the world of social data. The effect of this interest of data by computer scientists has resulted in the creation of data scientist. He emphasized the importance of each role and how their differences are based on how each position uses data. Social scientist tends to use the data for a task at hand while data scientist would engage in exploring ways into finding hiding information within the data. He made an interesting analogy of doctors being able to make a somewhat diagnostic based on a patient's symptom, but the sheer data provided by X-rays have helped doctors make a definitive diagnostic of the patient's possible health concerns. He sees the same when looking at social science and computational social science.

He talked about the boom that Data Science has experienced, and foresee that data science not to be a fad but rather an expansion of computational analysis that will eventually reach a plateau of normality.

During this quarter, we talked at length on the importance of data scientist and how customer's purchase tracking behaviors could be a good or bad thing. Target's example was one that showed the creativity of data science being able to pinpoint a social behavior, but one that could have caused some major social impact to the people impacted by this experiment.

Research Design Methodologies

The speaker listed out four main steps that should be followed for research to be impactful.

1. Observing behavior – this is related to the social science of understanding human interaction
2. Asking questions – being able to come up with the right questions will take time and practice.
3. Using experiments – use of surveys and interviews are crucial to perform meaningful experiments
4. Creating mass collaboration – exchanging and collaborating with others.

I am happy to say that we have done quite a bit of observing behaviors and collaboration during this class. We have had multiple labs and projects where we had to observe, analyze and report on

behaviors. A good example was the analysis that we did for the Titanic data. We got to analyze different ways to figure out who will survive the Titanic. As far as collaboration, we got to work as a group in some of the projects, and we had to opportunity to discuss and exchange ideas with others in the team. We got to question the result we got from the models generated from our analysis. At times, we had to figure out whether more cleanups are needed on the data or whether we have enough to perform our analysis. Also, we learned to be critical of the data visualization and understand whether the data at hand is projecting any type of trends. On the other hand, we did not do much of surveys or interviews as part of this class.

In addition to providing relevant examples of research designs such as monitoring wealth and poverty in Africa through cell phone location, and the creation of valuable companies that help reduced energy consumption by analyzing's neighbor's energy consumption, the speaker spent lots of time explaining how surveys are necessary. He said though most data scientists don't particularly see the importance of surveys, they are critical and useful in coming up with a model. But he did also emphasize that surveys need to be modernized to fit our new standard of overloaded social data. He gave the example of the presidential election results and how the exit poll may have been the results of bad surveys or possibly the inability of pollsters to understand voters' true feelings.

Big data collection was another topic of discussion. He explained the limitations that come with accumulating tons of data. He referred to the data stored by the NSA as an example of valuable data that would have been a blast to perform analysis, but are not made available for people outside of NSA to analyze.

Ethics

The speaker seemed to have a strong attitude toward ethics, and he is somewhat challenging those who are in the field of data mining to be a steward of ethical attitude toward the usage of data. He noted that social data had entered a new arena to which there are no rules or laws to follow. He argued that data scientist should use a principle-based approach, but by still following the established rules. He noted that companies like Amazon and Google are using tools like Echo, and Google home to collect data while not being transparent. These companies failed to disclose to their customers that those tools are designed to capture all sort of data without customers being aware of the reason behind these data collection.

Also, in the age of intellectual property being in jeopardy with so many copyrights infringement, the speaker made a bold move by having his upcoming book open for review as open source. It shows his dedication towards openness and transparency.

This class from the start was based on transparency and from the beginning there was an emphasis on being accountable to the work that we do. I think this is a good practice and one that will carry on after school. I think everyone in the program feels the importance of ethical

Conclusion

I liked the passion that the speaker showed talking about accountability and transparency. As he quoted, "with great power comes with great responsibility." Thus, as future scientists that will be dealing with data, we do have a moral and ethical responsibility always to be transparent. I am happy with the lessons learned from this class as they tied directly to the message of this talk.