DATA607 Presentation Narrative

Data Science with Contact Tracing

# Presentation Overview

Data Science is applicable to in virtually all industries. While most corporation’s use this science to make a profit (e.g., retailers, marketing, advertising), there have been some endeavors where data science was used for the greater good, specifically in Healthcare.

# Examples

* Google Flu Trends
* Watson Health AI
* Covid Tracker

# Timeline

* 2008-2013
* 2014
* List objectives here

# Details:

**GFT**

1. Release 2008
2. First published in 2009 on nature.com article “Detecting influenza epidemics using search engine query data”
3. the goal of GFT was to outperform the existing influenza surveillance system used at the time by Center for Disease Control (CDC) and European Influenza Surveillance Scheme (EISS).
4. The traditional approach was using a combination of clinical, virological and physician reporting data of ILI (influenza like illness), to detect outbreaks.
5. Traditional method has roughly 2 weeks.

**IBM Watson AI**

1. Announce in 2014 by IBM
2. The goal was to improve health care by implementing advanced AI in patient care.

**Contact Tracing**

Approach:

**GFT**

1. Google believed during a flu outbreak web search terms would match flu related topic.
2. The method was to use to match Flu terms with web search term to narrow in on regions with outbreak.
3. As reported by datacollaborative.org, the amount of search terms involved was 50 million search, which Google fitted to 1152 data point.
4. Desire goal was to predict seasonal flu and potential pandemic, should there be a mutated strain

**Watson AI**

1. Combined Watson’s Natural Language Processing (NLP), IBM Cloud, massive databank of unstructured text and metadata, for analytics use.
2. The was to determining patient illness, with unbiased, fact driven diagnosing, utilizing Watson’s data bank of historical clinical documentation.
3. Watsons’ ability to understand statements, regardless of wording or in any “Natural Language” was verified on the gameshow Jeopardy. (small nod to Alex Trebek

**Contact tracing**

# Obstacles

GTF:

* The first immediate obstacle is overfitting.
  + The data was entirely broad data
  + It was made to fit the 1152 data points too closely
  + The overfitting resulted in many inaccuracy when put in a live environment.
* [Wired.com’s](https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/) article “What We Can Learn From the Epic Failure of Google Flu Trends” puts the blame on *Big Data Hubris*.
* Data Hubris - “the assumption that big data analytics can be used as a substitute rather than a supplement to traditional means of analytics” [Quora.com](https://www.quora.com/What-is-big-data-hubris)
* As explained in [Globisinsights.com](https://globisinsights.com/tech-innovation/denying-big-data-hubris-3-reasons-machines-still-need-us/)’s article “Denying Big Data Hubris: 3 Reasons Machines Still Need Us”:
  + human intervention is still needed in data acquisition and analysis
  + Machines Cannot Define Goals
  + Machines do not know what data needs to be collected. Example being Garbage-in-Garbage-out (Nokia)
  + Machines cannot interpret data, they can only follow instructions
* The overconfidence in the accuracy of these results, led to Google advertising its predictive model as “nowcasting”. [Wired.com](https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/)
* An added obstacle was the fact that in the same time frame Google introduced “suggestion search” into there web search engine, which skewed the search term results, by directly impacting their users. [Searchengineland.com](https://searchengineland.com/googlecom-finally-gets-google-suggest-feature-14626)

**IBM Watson**

* Accents or various pronunciation of word during data acquisition
* Big Data Hubris and a continued need for a clinician to analyze a patient’s symptoms
* Complexity and nuances of the healthcare profession. Humans, human psychology and anatomy, can vary according to demographic, diet, activities and various environmental factors. (e.g., reported cases of anxiety and mental health issues has increased over the years, with the technological advancement of mobile devices, social media etc.)

# Performance

**GFT –**

* Flu outbreak reported by many news reports including USAToday.com as [“The worst in recent history”](https://www.usnews.com/news/articles/2013/01/10/worst-flu-outbreak-in-recent-history-strikes-united-states)
* GFT missed the mark during the 2013 flu season by an estimated 140% according to [Wired.com](https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/).
* Subsequently Google quietly discontinued the tracker.

**IBM Watson**

* IBM sale of commercial AI in healthcare is relatively minimal over the past 8 years.

[Spectrum.ieee.org](https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care)

* NLP has not shown any relevant improvement on diagnosing. Experts point to the knowledge gap between developers and health professionals, who understand the field from working the front lines.
* Ultimately, the need for a human factor remains constant in the healthcare profession reinforcing the need to avoid big data hubris.

# Ethics

Possible student prompts: What did you learn? What worked well? What was the most challenging aspect of this project? What will you do differently next time?

# Add the link to your project here:

Link to access project