Girish Rajani
A20503736
CSP-554 Big Data Technologies
Homework 1

The Parable of Google Flu

1. What was the problem with the Google flu detection algorithm?

The Google flu detection algorithm was tasked with the objective of predicting CDC reports of Influenza-like Illness (ILI). However, instead, it was predicting more than double the number of doctor visits for ILI when compared to the CDC reports. The algorithm overestimated flu spread and missed many instances of Influenza such as the nonseasonal 2009 influenza pandemic, hence, resulting in inaccurate and non-reliable results.

2. What is big data hubris?

Big data hubris has to do with people becoming too confident using big data and thinking that traditional forms of data collection and analysis are obsolete. Instead, both big data and traditional data collection methods should be used together to increase the validity and reliability of data.

3. What approach could have been used to improve the Google flu detection algorithm?

To improve the algorithm, GFT could have combined their algorithm with the delayed reports from CDC data. As shown in the graph from the article, by combining the Google Flu algorithm and CDC data, the prediction of influenza, although not as accurate as CDC predictions, still resulted in a significantly more accurate projection than just using the algorithm itself.

4. What is “algorithm dynamics?”

Algorithm dynamics are considered to be the changes made by engineers and programmers that would improve the service and also by consumers who use that service.

5. What aspect of algorithm dynamics impacted the Google flu detection algorithm?

Many changes that were made to the Google search algorithm as well as the user behaviors impacted the algorithm resulting in an unstable prediction. It was mentioned that within 2 months, Google had made over 86 changes to the algorithm. “Blue team” dynamics have impacted the algorithm whereby the algorithm has been altered to suit the business model.

Set up an Amazon Web Services (AWS) cloud account:

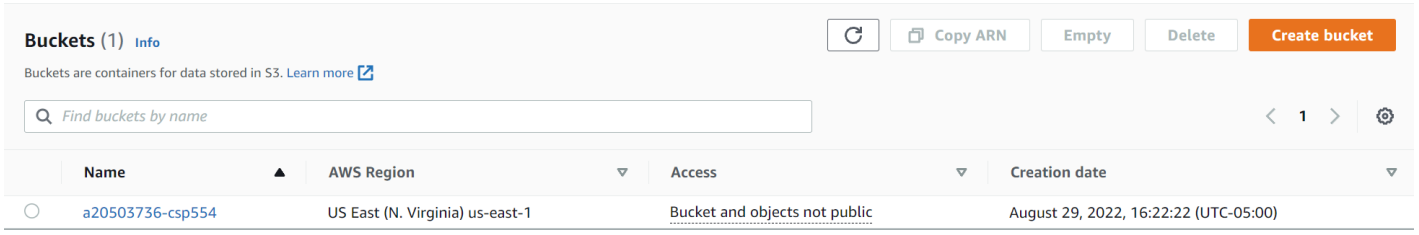

Figure 1 - Showing the S3 Bucket Created called a20503736-csp554

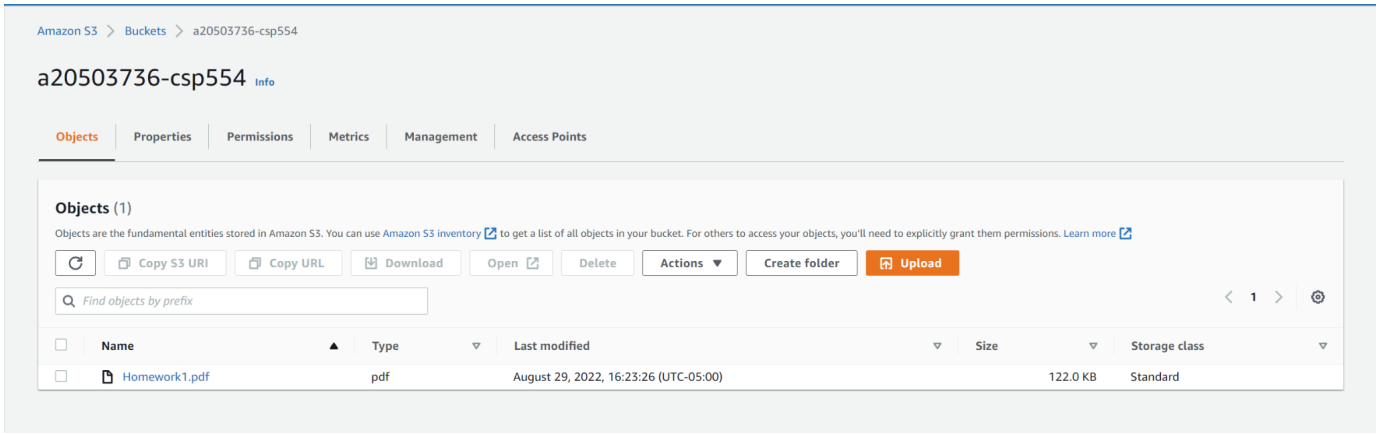

Figure 2 - Showing an Object Called Homework1.pdf in the Bucket Created