Rakshith Churchagundi Amarnath

[A20424771]

CSP-554 Assignment 1

Question 3.

Prior Experience:

I have taken courses related to data science like - Data Mining [CS422], Artificial Intelligence [CS480], and Biometrics [CS559]. I have done a summer internship [2019] in BMW Technology Corporation where I worked on Microsoft Azure.

Big Data interests and personal learning goals for the course:

I became fascinated with data when I was working on one of my research paper where I explored the possibilities of detecting pulmonary diseases using Computer vision models by inputting x-rays. My current goal is to learn more about the tools and techniques required to process big data and get a better understanding of it.

Additional topics in the scope of the course of special interest to you:

I am very excited to learn and work on AWS, Kafka, Hadoop and read new articles on the upcoming data processing and visualization techniques.

Other Necessary Accommodations:

None.

Question 5.

Problem with the Google flu detection algorithm – It predicted more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from different labs across the US. This happened even though GFT was built to predict CDC reports. The article also explores two issues of GFT – Big Data Hubris and Algorithm dynamics.

Big Data Hubris can be defined as the assumption that **big data** analytics can be used as a substitute rather than a supplement to traditional means of analytics [scientific analysis]. The overestimation of influenza cases by GFT is a prime example for Big Data Hubris.

What approach could have been used to improve the Google flu detection algorithm?
The approach of combining GFT and lagged CDC data, as well as dynamically recalibrating GFT, we can substantially improve on the performance of GFT or the CDC alone.

Algorithm Dynamics – Algorithm dynamics are the changes made by engineers to improve the commercial service and by consumers in using that service.
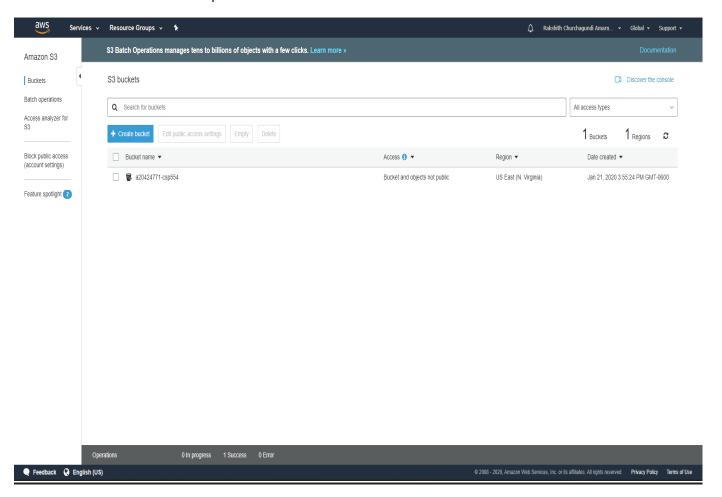
What aspect of algorithm dynamics impacted the Google flu detection algorithm?
The GFT was an unstable reflection of the prevalence of the flu because of algorithm dynamics affecting Google's search algorithm. One of GFT's error was a media-stoked panic in the last flu season
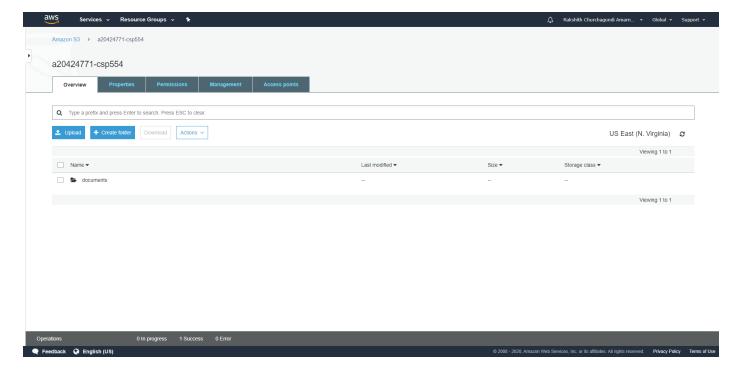
Question 6

AWS Screenshot:

Bucket Name – a20424771-csp554

Folder inside the Bucket - documents

Object name – CSP 554 Assignment 1.docx