CPSC 325
Dominic MacIsaac
Feb 1, 2023

Research Proposal

My eventual project will mine images and metadata from a social media platform, do some sort of analysis on that data (image recognition, color/hue/composition/histogram analysis, etc), report that analysis using data visualization tools (or prediction/identification tool) on a dashboard which will be web hosted.

The technologies I will need to be using are: the Flickr API, CV2, GCP, and Tableau or Dash. I would like to use the Flickr API to data mine for images. I considered instagram initially but their API is very restrictive on mining public posts. These images are posted with data such as the camera and lens used, camera settings used, geolocation data, and time that could all be useful in analyzing images. Next I will be using the python library, CV2, to analyze images as well as do machine learning on the data if needed. CV2 is a huge open source computer vision library that provides many functions that will be important for extracting data out of images. Next, I will be using GCP for both cloud computing as well as for web hosting my site. Finally, I will be using either Tableau or Dash to create a data visualization dashboard to present my findings.

I have worked extensively in Python but have not used any of these technologies before so there will be a lot of learning involved. Due to this I will need to learn all the technologies listed above (Flickr API, CV2, GCP, and Tableau/Dash).

Each of these technologies has a range of options for learning. There are many tutorials available on youtube to learn the Flickr API as well as documentation. For CV2, I found many extensive tutorials on youtube as well as an IBM course on coursera (1). I found a course offered by OpenCV but it was very expensive so I most likely won't do that one. I would like to take the GCP course offered by google. Lastly, I have also found a coursera course from IBM on Dash that I would take to learn more about data visualizations (2).

Links:
   (1) https://www.coursera.org/learn/introduction-computer-vision-watson-opencv
   (2) https://www.coursera.org/learn/python-for-data-visualization

Hours:
These two coursera courses are very long, roughly 20 hours each so I probably won't complete either during the 4 weeks of research. I will likely target parts of the courses that are relevant to this project and finish the courses on my own time. The Flickr API shouldn't take me a tremendous amount of time to learn and tutorial courses on youtube are rather short (about an hour total in content). Lastly, the GCP course is roughly 10 hours, I believe (I couldn't find that information online without signing up with a credit card).