Week 12 Progress Report
DATA 450 Capstone

Andrew Nemkov

Week ending: 4/19/23

Time Log

Day	Time	# Hours	Task(s)	GH Commit(s)
Mon.				
Tue.				
Wed.				
Thu.				
Fri.	[9:00pm-11:00pm]	[2]	[background research on blog posts]	
Sat.	[3:30pm-7:30pm]	[4]	[further research on neural networks and OCR usage]	
Sun.				

Total # of hours worked this week: 6

What you accomplished (or attempted) this week:

The majority of this week I spent time research on further improvements of changes that my code would need for the future, since I know my current code can make significant adjustments. On Friday I spent a bit of time looking into making data science blog posts as I watched a video online that delved into the usage of sites such as medium.com, GitHub pages, or dev.to to make blog posts for projects, as well as what needs to be in a blog post in terms of content and description.

Additionally, I spent time researching the basics of development of neural networks, as in how they function on a basic level using the concepts of linear algebra and calculus to "learn". I also

spent time researching on further usage of OCR method on Japanese language and I was able to find an interesting video where researcher Bing Wang went into analyzing performance of three types of OCR models on Japanese medical documents. This by far was the most useful information I saw, as one of the three models she went over was PaddleOCR, which I am currently using in my code. I also discovered interesting things from her research, specifically that all the three models' accuracy improved with an increase of image quality. Her final note for her presentation was that overall, the Tesseract model seems to be the best performer in terms of accuracy rate and time efficiency compared to the other OCR models publicly available.

This is the OCR presentation I watched below for further context: https://www.youtube.com/watch?v=OitWeF

Any setbacks/roadblocks you experienced:

I was expecting to find it very difficult to find information regarding OCR usage for specifically the Japanese language, but I think I was very fortunate to find the video I talked about above. So I didn't really run into any roadblocks this week for my project.

What you plan to work on next:

I plan to start working on my project's blog post next week to finalize all my research and findings into a single area for simplicity and ease-of-access.

Are you on track? If not, how will you get back on track?

I am very much on track for my project and it seems interesting how this project progressed over time. I am eager to complete it in the future.