Software Engineering Afterword Matthew Hammonds

Ethics:

When it comes to ethical questions there wasn’t quite many as the project of collecting numerical data for feature extraction isn’t exactly the most disruptive of programs. Though one question we had pondered was how someone could potentially use this in order to detect what kind of cyber-attacks a company is used to being bombarded by. This extraction will be primarily used by cybersecurity groups so that they can detect what different cyber attacks look like, by looking at the header data of TCP and UDP files. When using this a bad actor could use this type of program in order to find what kind of attacks a specific company could be used to and use that knowledge in order to do something different or tweak their attacks in order to have a better chance at completing their task. This could be potentially solved by having tighter security on TCP and UDP data and having the company use the same program to get a better understanding of the type of attack and even further protect themselves from that specific kind of attack. They can also use it to anticipate other bad actors deciding to use other types of attacks and have tighter security on those attacks as well.

Learn:

When it comes to new things learned the concept of gathering tabular data and pcap files as a whole was completely new to me. I had actually never used Wireshark before this as well when figuring out how pcap files work and the type of data that they gather. This led to us taking a couple of weeks to get our heads around exactly what pcap files were and the specific type of data that we were going to collect. The project also started with learning about natural language processing as well and while the project changed just to numerical data, learning about machine learning involving natural language was quite new as well. The only exposure I had with machine learning before this was knowing that the YouTube algorithm uses it.

Contribute:

During the project I contributed by being the one to code how the folder of pcap files will be taken in as input (whether that be as a folder or file) and how it would be outputted as a folder of csv files. I had also created the tests for the different sections of the code. This would include whether it had the right number of columns and that the data itself if correct. I had done this by creating small test files that would test UDP and TCP files.