

The Senior Design project is a research project about testing and maintaining machine learning systems in real world systems. From an academic perspective, it involves statistical analysis of data and machine learning systems. One has to be familiar with machine learning concepts such as overfitting, underfitting, dimensionality problems and training schedules. To understand machine learning systems, I need to be well versed in linear algebra. I would also need to use concepts from data structures and algorithms to handle large amounts of data in a quick flexible manner. Furthermore, the research project allows me to understand the process of reading papers, compiling results, writing papers and citing sources. This will help prepare me for my graduate studies.

I am currently taking Machine Learning (CS 5137). The course introduces machine learning concepts such as training, testing and validation and ML algorithms such as linear regressions, SVM Classifiers and Neural Networks. The course relies heavily on concepts taught in Linear Algebra (MATH 2076) and Prob. & Statistics I (STAT 2037). These play a key role in understanding and evaluating machine learning systems. Last year, I took Design & Analysis of Algorithms (CS 4071). It taught me how to compare algorithms in terms of big O, prove their optimality and how to approach designing algorithms. This was built upon knowledge from Data Structures class (CS 2028C). These experiences aid me in my process of writing optimal algorithms to do tasks over large datasets and save time. However, a project requires more than just technical skills. It involves reading and writing papers. I learnt a great deal about Technical Writing in ENGL 4092 (Technical Writing). Furthermore, I have to deliver presentations about my project. The Effective Speaking (COMM 1071) prepared me to deliver presentations in front of a live audience.

My knowledge is not limited to my classroom experiences. I matured a great deal in my co-op experiences. I worked as a ML Engineer for a startup called Omniaz for Fall 2019 and Spring 2020. My work focused on Object Detection & Tracking for mobile systems, Content Based Image Retrieval and OCR. I primarily employed Python with Tensorflow, PyTorch, OpenCV, SciPy and NumPy libraries to accomplish these tasks. In this project, I have to read and understand published papers, scavenge for implementations of papers, be able to manipulate datasets and test machine learning solutions. The projects entailed scavenging numerous datasets and reading published papers for latest solutions. Furthermore, I was in charge of planning projects in terms of timelines and breaking it up into different tasks. As I am the only person on this project, except for my advisor, I have to designate specific timelines to ensure that the project proceeds in a timely fashion. Moreover, I have to constantly be in touch with my advisor to give him timely updates. Lastly, as part of my coop, I delivered presentations to people from various technical and non-technical backgrounds. As part of the senior design project, I believe I will be delivery presentations about the research.

During my coop, I faced the challenge of maintaining ml systems. I didn't know how to empirically establish when to update a ML system or the challenges a ML system faces in the ever-changing real world. This research project will introduce me to literature on maintaining and testing ml systems. The preliminary approach is to investigate effects of image distortion on image classifiers. I am currently finding literature on simulating real world distortions such as rain, snow, fog, various lightning conditions and image degradation due to low quality cameras.

I divided distortions into different categories of distortion: noise due to physical objects in images (rain, fog, snow, objects occluding), noise due to lighting, angles or reflective surfaces, noise due to motion or vibrations and noise due to image degradation either by hardware or software. I will be done if my advisor is satisfied I have found enough credible distortions. I evaluate my work on the number of papers I can find and how many in each category or if I have found any new categories. I will consider myself to have done a good job if I can find 20 distortions.