Project Proposal

Problem:

        For the third capstone project, I decided to analyze patient lung x-rays to see if I could create a model that could predict COVID positive patients. The pandemic has stretched many healthcare facilities thin, increasing the stress of all healthcare providers and making it more difficult for patients to get the adequate care they need. The idea behind this model is to try and ease the workload of healthcare providers which in turn reduces their stress and increases quality of care for patients.

Criteria for success:

        This model is intended to help Doctors make quick and accurate decisions on whether a patient has COVID or not. Given that this is a very important diagnosis the criteria for success would have to be at least a 99% accuracy with extensive testing done on a wide variety of patients. However, since that is nearly impossible to achieve given my skill set and would most likely require a team of healthcare experts and high level data scientists, a better criteria for me would be above 90%. In addition, should the model make an incorrect prediction, it would be prudent that false negatives are as low as possible so a high recall would be preferred.

Data Sources:

        The data I will be using comes from [Kaggle](). This data comes with different lung x-ray images with different patients. The data comes with COVID positive patients, negative patients, patients with Viral Pneumonia, as well as a lung map. I will use only the COVID positive patients and the negative patients for this analysis. In total the data contains 10,192 (74%) negative patients and 3616 (26%) positive patients. The data is an aggregation of different images from 43 different publications.

Method:

        I will need to analyze the images to find differences between positive and negative patients. One method is to check the pixel intensities between positive and negative images and see if a difference exists there. In addition, I will be using a Convoluted Neural Network to analyze to create a model that could predict positive and negative patients. Hopefully the model will have an accuracy of over 90% and have a high recall to prevent false negatives from occurring.