MAIS 202 - PROJECT DELIVERABLE 2

1. Problem statement

I will be implementing a model that classify X-Ray scans from patients with pneumonia. I will use the dataset from Kaggle. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

2. Data Preprocessing

Firstly, I will read the images from 3 folders (Train, Val, Test). Each of them contains chest images of people who have pneumonia and people who don't. Since the input dimensions of the images are too big, I will then use PCA to do the dimension reduction. Finally, I will plot the images corresponding to their labels of NORMAL and PNEUMONIA.

3. Machine learning model

I will train the dataset by creating a CNN model in KERAS. What I noticed from the dataset is that the images of normal and pneumonia chests are imbalanced, so I will use imgaug library to balance the overall distribution.

If the accuracy of the validation set is similar to the accuracy of the training set, my model is good. However, if it is much lower than the training set, it means my model is overfitting. I will plot the graph of the model performance to improve the accuracy.