Project Proposal 

#### *Francesco Malagrino*



# Data Labeling Approach

|  |  |
| --- | --- |
| **Project Overview and Goal**What is the industry problem you are trying to solve? Why use ML in solving this task? | Health care and the product should help the doctor to detect based on an image if the child or adult have pneumonia or not |
| **Choice of Data Labels**What labels did you decide to add to your data? And why did you decide on these labels vs any other option? | I choice to add the following labels :  Pnumonia because this job need to dectect if a human have this desease or not  Healthy this means that the human is healhy  Other this means that the person may have some hearth issue like vasculite or something related to the hearth  Unknow means that the human may have anything else related the hearth or the image is poor clear |

# Test Questions & Quality Assurance

|  |  |
| --- | --- |
| **Number of Test Questions**Considering the size of this dataset, how many test questions did you develop to prepare for launching a data annotation job? | 4 |
| **Improving a Test Question**Given the following test question which almost 100% of annotators missed, statistics, what steps might you take to improve or redesign this question? | The lungs is clear ?  Have the hearth some Opacity?  Have the lungs Opacity?  There is a huge Opacity? |
| **Contributor Satisfaction** Say you’ve run a test launch and gotten back results from your annotators; the instructions and test questions are rated below 3.5, what areas of your Instruction document would you try to improve (Examples, Test Questions, etc.) | Test Questions and Examples |

# Limitations & Improvements

|  |  |
| --- | --- |
| **Data Source**Consider the size and source of your data; what biases are built into the data and how might the data be improved? | Adding the value on the label actually there is few 0 and 1 then become empty. With the label based on the number it could be possible already detect if there is some opacity or not. |
| **Designing for Longevity**How might you improve your data labeling job, test questions, or product in the long-term? | We can add more info in the data set adding more and more columns based on the future requirement |