Project Proposal 

#### *<Fatimah Aljuhani>*



# Data Labeling Approach

|  |  |
| --- | --- |
| **Project Overview and Goal**What is the industry problem you are trying to solve? Why use ML in solving this task? | **The goal is build a product that helps doctors quickly identify cases of pneumonia.** **Using ML to make it easier for the doctors to identify the disease** |
| **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? | For the first question (Are there pneumoniain in this image?) I decided to choose (Yes, No , don’t know)  because it is clear and direct |

# 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? | 10 |
| **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? | Rewrite the question in another way, try to write clear question |
| **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.) | Make the instructions more clearance  Add more instructions  Add more examples to be clear for annotators how to solve test questions  Make sure that questions and labels are clear |

# 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? | Increase the size of the data (add more x-ray images) will be better to understand more |
| **Designing for Longevity**How might you improve your data labeling job, test questions, or product in the long-term? | I see the test launch if the results from annotators and test questions are rated below 3 , then I need to improve my data  Maybe data labelling are not appropriate try to make it easier and more clearance for the annotators  Also try to rewrite questions in understandable way |