Project Proposal 

*<*Aisha Saad Al-Shammari*>*



# Data Labeling Approach

|  |  |
| --- | --- |
| **Project Overview and Goal**What is the industry problem you are trying to solve? Why use ML in solving this task? | It is building a product that helps doctors quickly recognize cases of pneumonia in children.  1\_ Recognize the disease quickly  2\_ Help doctors classify and diagnose  3\_ Contributes to the advancement of medical science  The reason for using ML is accuracy in data and ease of classification |
| **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 added the name of a small cloudy area and also a large cloudy area, and the reason for choosing is that it is scientifically accurate and clear in form |

# 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? | I started with two questions, the first was whether the lung picture was healthy while you had a problem, and the second was what kinds of symptoms are in the lung in the picture |
| **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? | Provide sufficient data on the question, taking into account its accuracy and the clarification and timeliness of the source |
| **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.) | Reframe questions, instructions, options, 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? | We always start with the process of analyzing the data and collecting the data you need, amending it and fixing it to make sure that the data is of the highest quality |
| **Designing for Longevity**How might you improve your data labeling job, test questions, or product in the long-term? | Continuously developing data and continuing changes in the annotation function, and keeping abreast of changes in reality |