

Project Proposal



<Alex Key >

Data Labeling Approach

Project Overview and Goal What is the industry problem you are trying to solve? Why use ML in solving this task?	Identifying pneumonia in chest X-ray images is currently done by physicians inspecting an image and making a judgement based on a couple of indicators (further, there are other factors to take into consideration besides the x-ray images). This process requires a human assessment – however, there are cases that clearly indicate no case of pneumonia, in which a human inspection is unnecessary. Why use ML? The task is repetitive and, given a large enough dataset, can be taught to a machine through image recognition.
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 choose three data labels: 0 – no indication, to weed out clear cases right away 1 – likely a case of pneumonia – should be inspected by a physician 2 – very likely case of pneumonia. In such cases, inspection by a physician will be needed anyway.

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 prepared 11 test questions, which would amount to almost 10% of the whole dataset. This would probably be too large a share for larger sets of data, but for such a small one it seemed reasonable.

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?

ID	% CONTESTED	% MISSED	JUDGMENTS	LAST UPDATED	ENABLED ▾
1881190030	<div><div></div></div>	<div><div></div></div>	2	2 days ago	<input checked="" type="checkbox"/>

There are a number of measures we can take:

- First, check if the test question was correctly answered by me (obvious, but this possibility needs to be ruled out)
- Make labels more specific
- Give clearer instructions on what the labels mean
- Check if the rest of the instructions needs an update (rules and tips section especially)
- Check if the examples in the annotator instructions are clear and correct

If time and budget allows, I'd change only one of these factors at a time and re-run the evaluation, to see which one has an impact.

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.)

Contributor Satisfaction ⓘ

Number of participants: 20

3.2 / 5

Overall

3.3 / 5

Instructions Clear

2.9 / 5

Test Questions Fair

2.8 / 5

Ease Of Job

3.7 / 5

Pay

In this job, I assume the most likely reason for annotators to struggle is with the level of uncertainty I built into my labels. (there's 0, 1, 2 as answers – whereas 1 or 2 leave room for interpretation). I might consider changing the whole labeling to a simpler set with only 0, 1, and unknown. If we'd really build such an application, one could expect that cases labeled as 0 show no indications of pneumonia. Cases with 1 would show any indication of pneumonia and need to be reviewed by a physician anyway.

Second, I'd also need to check if the instructions are concise and clear, and there are sufficient examples.

Limitations & Improvements

<p>Data Source</p> <p>Consider the size and source of your data; what biases are built into the data and how might the data be improved?</p>	<ul style="list-style-type: none"> - There might be differences in image quality that make it harder to identify cloudy or obscured areas - Some shapes might be misinterpreted – the heart for example might be mistaken as a cloudy area - If the images are all from the same source (the machine is the same model, and/or the same hospital), we'd need to gather data from other sources as well. Otherwise, we can't know if data from other sources might look significantly different - We should ensure that there is an equal number of healthy and non-healthy images in the dataset
<p>Designing for Longevity</p> <p>How might you improve your data labeling job, test questions, or product in the long-term?</p>	<ul style="list-style-type: none"> - The diaphragm shade is an indicator of a healthy person, but it might as well be missing without indicating a pneumonia. So, its absence seems to carry little significance. We could consider leaving it out of the annotation at all and just focus on the cloudy or obscure areas - We should make sure to gather data from different physicians and/or x-ray machines - We could introduce more specific labeling for certain conditions in the image. For example, checkboxes for all different indicators (large clouds, scattered clouds, diaphragm shadow yes/no). Downstream, we might then weigh these indicators in our product (some will indicate less or more likelihood of pneumonia) - It's critical to include professionals (physicians) in the development of the product. Latest, when the first model is trained, the results need to be verified with a professional – ideally, we could involve them already in the labeling. We could even have them create example questions for the labeling job. - There might be even more indicators hidden in these images, about which we can learn by talking to professionals.