Project Proposal

Shasu Vathanan

Data Labeling Approach

Project Overview and Goal What is the industry problem you are trying to solve? Why use ML in solving this task?	**Pneumonia Chest X-Rays:** As per the survey, many people are affected by Pneumonia and losing their life. In today's technology, radiologists using the Chest X-rays of the patient to identify pneumonia in them. So, all these chest x-rays need to be verified by the experienced radiologist and they need to show the variation of the complexity of the particular patient. To solve this complexity, we trained the Machine Learning Algorithm to diagnosed pneumonia in Chest X-Rays. So that doctors and radiologists can easily identify the symptoms of pneumonia from the unclassified chest x-rays. In this case, we trained the algorithm with the help of huge datasets produced by the certified hospital or by the x-ray laboring. **Benefits:** Both doctors and patients, in this case doctors can easily identify pneumonia rapidly and suggest the patient to hospitalized and take the needful treatment for them.
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?	**Choices of data label:** 1. Is this is a Chest X-Ray? Ans: (Yes/No) 2. In the lungs area (Lungs, spine, ribcage, heart, diaphragm) are clearly visible? Ans: (Yes/No) 3. Is Pneumonia presents in this X-Ray? Ans: (Yes/No) **Reason:** 1. Instead of chest x-rays, we may get various x-rays. So, the first label to confirm the Chest X-Rays. 2. In the chest x-ray, they need to clearly see the Lungs, spine, ribcage, heart, and diaphragm. 3. The chest x-ray is not clear in the lungs area with cloudy or opaque, we can confirm it as pneumonia is present.

Test Questions & Quality Assurance

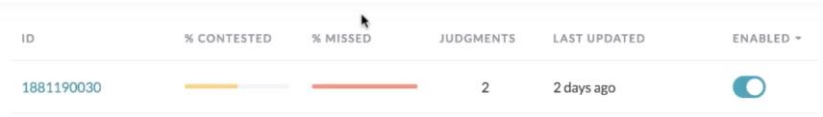
Number of Test Questions

Considering the size of this dataset, how many test questions did you develop to prepare for launching a data annotation job?

I created 20 test questions, so on cover all the possible options presented within the answers.

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?



A circumstance like this can mean the picture introduced inside the test question is intricate and equivocal. I may ensure the appropriate responses are clarified all the more plainly, and furthermore incorporate it inside the Directions itself all together that an increasing nitty-gritty clarification with jumping boxes and hues are frequently given.

In the event that there is a specific viewpoint that is mind-boggling or befuddling, I would also consider improving my general test plan so I gather however much data as could be expected in future runs, and perhaps have a free text input with the goal that annotators can demonstrate in detail a few inputs on why they picked what they decided for the picture.

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.)

The overall scores values, the Directions just as Test addresses need improvement. I may likewise need to consider if this kind of picture comment needs more authority annotators as opposed to lay, annotators, given the simplicity of the activity is evaluated low as well. Any way this can be considered after the directions and test questions are improved.

To improve the guidelines, I would expand more models, give a progressive point by point foundation, and propose a bit by bit for their thought.

To improve the test questions, I would mean to additionally separate the inquiries to catch however much detail as could be expected, conceivably permit free text contribution for questions the annotators reliably get off-base.

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?	The modified version of this dataset provided Kaggle chest x-ray dataset; each image of data is a Chest X-Rays. **The challenge to identify the Pneumonia:** **(Source Udacity)** A **normal**, healthy image will depict clear lungs without any areas of abnormal cloudiness/opacity; there may be structured, web-like vasculature in the lungs but otherwise that area should be clear. In healthy images, you are also more likely to see a diaphragm shadow. A **pneumonia** image may include a few things: areas of cloudiness/opacity in several concentrated areas or one large area. You may also see a general pattern of opacity that obscures the structure of the lungs, heart, and diaphragm. There may likewise be clamor if the pictures utilized for preparing are not well clarified themselves. Or then again, a substantial predisposition towards identifying pneumonia, on the off chance that there are such a large number of pictures with pneumonia Chest X-Rays. The opposite can likewise happen, on the off chance that we have too scarcely any instances of pictures with Pneumonia, it tends to be more diligently to really identify it.
Designing for Longevity How might you improve your data labeling job, test questions, or product in the long-term?	Instructions and questions keep up and change the substance in like manner. A conversation with the clinical network is basic to guaranteeing we remain on target. A further improvement would be ensuring the dataset is constantly revived with more up to date models, that there isn't a predisposition presented by requesting pictures just for certain ethnic gatherings or gender orientations or ages.