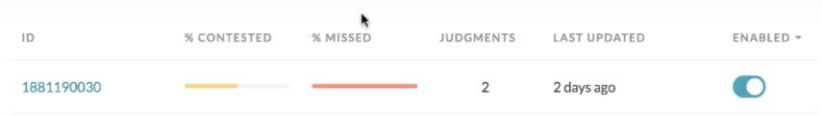
Project Proposal

May Abudujayn

Data Labeling Approach

Project Overview and Goal What is the industry problem you are trying to solve? Why use ML in solving this task?	**Build a product that helps doctors quickly identify cases of pneumonia in children.** ML could use labeled dataset that distinguishes between healthy and pneumonia x-ray images to improve pneumonia diagnosis and reducing of misdiagnosis.
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 had chosen (Yes, No, unknown) labels, to distinguish pneumonia x-ray images from any other unclear cases.

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've added 10 test questions to cover all possibilities of chest x-ray images.
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?	 It depends on question type or annotation type so I may add a text area that users can type the reason for there answers, or changing the labels, or rephrase the question, or add more examples to cover most scenarios.
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.)	 Rewrite the instruction section and include more specified and classified tips and examples of most common and tricky scenarios.

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?	There are no biases until now. Also, this data not quite large data that can made a powerful ML model. So, first I need to increase data size by diversification of sources to cover all possible scenarios.
Designing for Longevity How might you improve your data labeling job, test questions, or product in the long-term?	By continuously adding new data to keep learning with updating the instruction, tips, question and labels if needed.