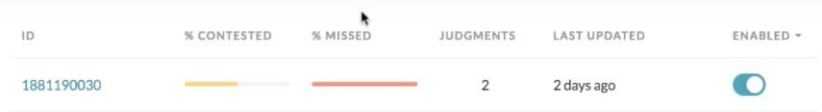
Project Proposal

Rimas Alshehri

Data Labeling Approach

Project Overview and Goal What is the industry problem you are trying to solve? Why use ML in solving this task?	As an AI product manager, I aim to create a tool that can assist doctors in promptly detecting instances of pneumonia in children. To achieve this, my initial objective is to establish a categorized dataset that can differentiate between normal and pneumonia-afflicted x-ray images. With this dataset as a foundation, machine learning can be employed to alert doctors to critical cases, accurately identify healthy cases, and generally serve as a diagnostic aid to medical professionals. This will enable doctors to concentrate on the crucial task of treatment, which demands their utmost attention.
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?	▪ Yes If yes then what do you see: • Lungs • Diaphragm • Heart • Other In the case of other type what is the other in the box ▪ No ▪ Other Yes implies that there is a pneumonia and after that based on this answer I ask the user what they see, so that I know based on what they decided so, since Yes means the annotator is totally confident in his selection. No, indicates that there is no sign of pneumonia at all, so no need to question them anything else. On this my role as an AI PM is to do manual check on the images I uploaded and measure its accuracy. Other, indicates that they see something else that could be not related to the image provided, in this case it could be that the image isn't helping in making the decision in which on my behalf I have to do manual check on the images I uploaded and measure its accuracy. A drawback of the current labeling system is that it involves three distinct labels, whereas what we actually require is a binary classification scheme that can differentiate between healthy and unhealthy instances. To address this issue, we will need to develop a strategy for reducing the number of labels to two once the annotation process has been completed. If the number of 'Not Sure' responses is minimal, we could attempt a manual check. Alternatively, if such responses are numerous, we may need to compute the mean of the scales and make a determination based on that. For instance, if the average score is less than 2.5, we can classify the instance as 'no', whereas if it is greater than or equal to 2.5, we can classify it as 'yes'.

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?	We have 101 unlabeled and 16 labeled data. As a result, the total number of data is 117. As suggested by Appen, I have developed 11 test questions. The percentage of your developed 11 test questions to the labeled and unlabeled data would be: (11 / 117) x 100% = 9.4%
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 appears that the instructions and/or sample questions are inadequate. Initially, I must review the incorrectly answered question in detail. Afterward, there are a few potential solutions to consider: • I could attempt to modify the instructions and/or • I could create additional examples to provide further clarity on the question that was missed. • Evidently, the test question in question is quite challenging. If I believe that there are enough other test questions for Quality Assurance purposes, I may also generate an example based on this particular 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.)	 Based on the outcomes, it appears that the annotators had difficulty comprehending the instructions, and they encountered problems when responding to both the test questions and labeling tasks. As a result, the primary goal should be to revise the examples in order to make the labeling tasks more comprehensible. Providing additional examples may help to improve their understanding. Additionally, since the annotators found the instructions to be unclear, I may consider enhancing the 'Steps' and/or 'Rules' sections. At this point, there is no need to modify the test questions. Once I have made the required changes, I can launch a new dataset and evaluate the feedback again. If necessary, I may make further adjustments.

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 dataset is notably small, comprising 101 unlabeled and 16 labeled instances, resulting in a total of 117 data points. Consequently, there is a high likelihood that our final predictions may exhibit significant sampling bias. Moreover, based on the project overview and prior experience, we are aware that the images differ slightly in size and were captured with slightly varying exposure times. As a result, we may encounter some measurement bias in our final predictions. To enhance the quality of our data, we must take the following actions: • Obtain a substantially larger amount of data. • Ensure that the new data shares the same size and exposure times.
Designing for Longevity How might you improve your data labeling job, test questions, or product in the long-term?	In my opinion, our data is subject to change over time due to advancements in imaging technologies and the emergence of new symptoms or diseases (such as Covid-19). Therefore, it would be beneficial to employ a dynamic model that is continuously trained on new data so that it can continue to learn from fresh input. In order to work with this type of data, we may need to revise our annotation process and update our data to incorporate more pertinent definitions, examples, and/or test questions.