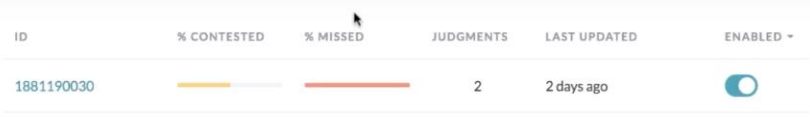
Project Proposal

Etendra Verma

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 is one of the hardest disease to detect especially in it's first stage. So our goal was to make an easy to use application that can predict if someone has Pneumonia or not given its chest x-ray. if there is pneumonia in a x-ray image "Pneumonia" or "Normal" or "Unknown". Machine learning the model is fed parameter data for which the answer is known. The algorithm is then run, and adjustments are made until the algorithm's output (learning) agrees with the known answer. At this point, increasing amount of data are input to help the system learn and process higher computational decisions.
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 have add two data labels in my job. 1:Pneumonia 0:Normal I am use these two data labels because of easy to configure the result of the x-ray image and easy to understandable the data labels in the dataset. If I am used other labels not easy to classify the accurate result also not to understandale easily.

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?	Creating accurate test questions is the best way to ensure high quality results from a job. Test questions should be hard enough to test a contributor's performance, but easy enough for those workers who are trying to follow your instructions honestly. Not all data rows make for great test questions. Don't create test questions that take you significantly more time than an average data row to complete or that don't have a correct response in regards to the instructions. I have created 22 test questions for the image annotation job.
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?	 1.Provide more training / improving the training method for the annotators so they are better equip. 2. Improving the test questions by evaluating the questions – look out for any confusion in the questions. 3. Conduct more check-points with the annotators to ensure the work to ensure any deviation can be detected earlier. 4. Monitoring each step carefully one by one.
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.)	 I will improve accuracy by adding more test questions, accurate images data, more options that making it easy for predicting the labels.

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 first during model training and again when your model consumes the labeled data to inform future decisions. To create, validate, and maintain production for high-performing machine learning models, you have to train and validate them using trusted, reliable data. There are 117 data sets and only 2 possible cases. Case 1: If there was Pneumonia not present in x-ray image then the result show “Normal”. Case 2: If the Pneumonia present in x-ray image then the result show “Pneumonia”. Note: If the x-ray image is difficult to configure then the result show “Unknown”. If we increase the size of datastet then the more accurate result will be obtained.
Designing for Longevity How might you improve your data labeling job, test questions, or product in the long-term?	1.Adding more test question in a job the faster job will be completed. 2.Adding appropriate labels which can easy to configurable and understandable. 3.Increasing the size of the dataset get the much more accuracy of the job in long term. 4.figureout the correct labeling and more labels help us to configure fast to complete job.