Project Proposal 

#### *<Erkan Hatipoğlu>*



# Data Labeling Approach

|  |  |
| --- | --- |
| **Project Overview and Goal**What is the industry problem you are trying to solve? Why use ML in solving this task? | **My goal is to build a product that helps doctors quickly identify cases of pneumonia in children**.  My first task, as a product manager, is to build a labeled dataset that distinguishes between healthy and pneumonia x-ray images. Later, by using this dataset, ML can help doctors to flag serious cases, quickly identify healthy cases, and generally act as a diagnostic aid for doctors. Therefore, doctors can focus on treatment which is a more serious task. |
| **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? |  |

# 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? |  |
| **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? | <your text here> |
| **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.) | <your text here> |

# 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? |  |
| **Designing for Longevity**How might you improve your data labeling job, test questions, or product in the long-term? |  |