Project Proposal

#### *Jude Anlasun*

# Data Labeling Approach

|  |  |
| --- | --- |
| **Project Overview and Goal**What is the industry problem you are trying to solve? Why use ML in solving this task? | According to the United Nations Children’s Fund (UNICEF), a child dies of pneumonia every 39 seconds globally. Pneumonia is the number one killer infectious disease among children under 5 claming the lives of about 2,200 children every day.  The most common diagnosis of pneumonia by healthcare providers is the full examination of patient health history and physical exams and if pneumonia is suspected, it is complemented by Chest X-Ray, which may show an infiltrate, which is a collection of pus, blood, or protein in the lung tissue. During the examination of Chest X-Rays healthcare providers can make mistakes or take alot of time in their diagnosis process.  To reduce the mistakes and the time taken by health care providers in diagnosing pneumonia in children, a Machine Learning product that can help doctors quickly identify cases of pneumonia in children would be built.  Using Machine Learning (ML) would make it easier and quicker for doctors to identify serious cases of pneumonia using images of Chest X-Ray. Provide a quicker way to identify healthy cases and also act as a diagnostic aid. |
| **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? |  |

References:

[1] https://data.unicef.org/topic/child-health/pneumonia/

[2] https://www.verywellhealth.com/diagnosis-of-pneumonia-4160855