

ANNOTATION PROCESSING

By: Kumar Ashutosh Anand



SOME INSIGHTS ABOUT PROJECT DATASET

Number of annotators contributed : 22

Number of Images annotated: 9087

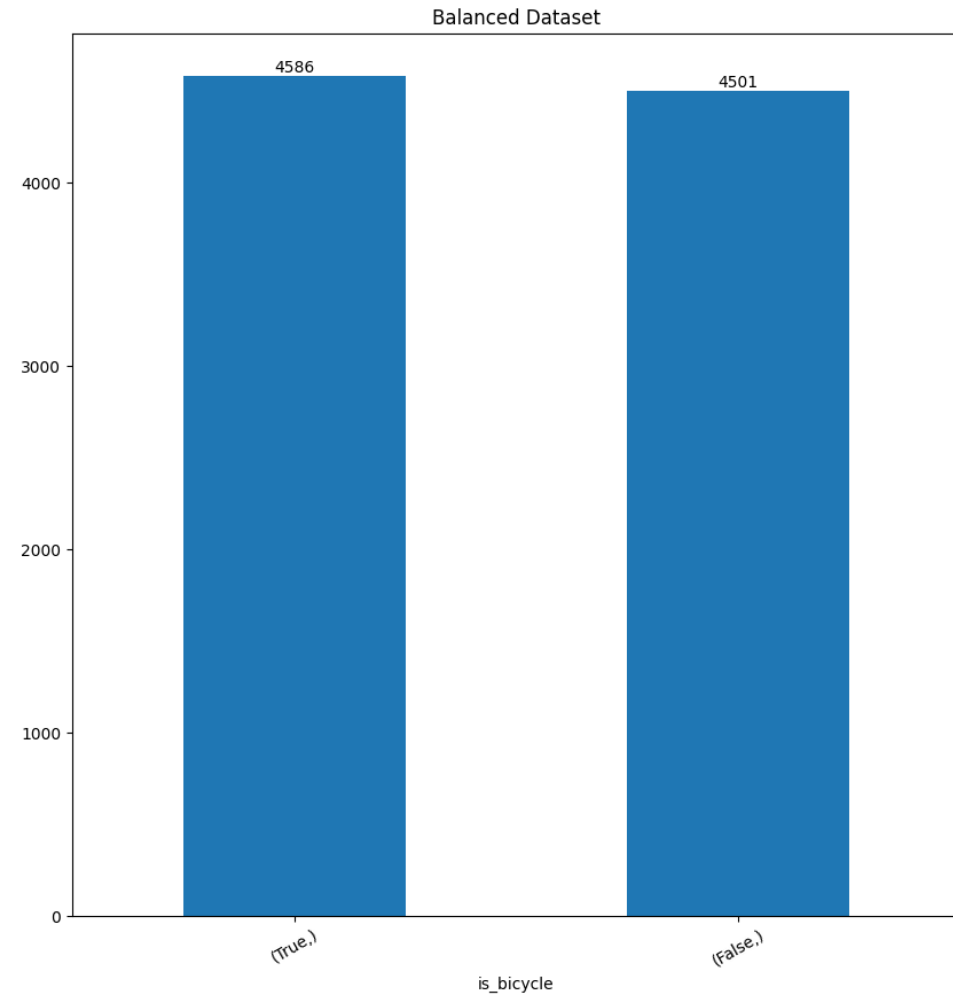
Total number of annotation records in project dataset : 90870

The annotators marked 4 images as “Corrupt data” and 17 images as “Can’t solve”.

Not all images were sent to all annotators.

BALANCED DATASET

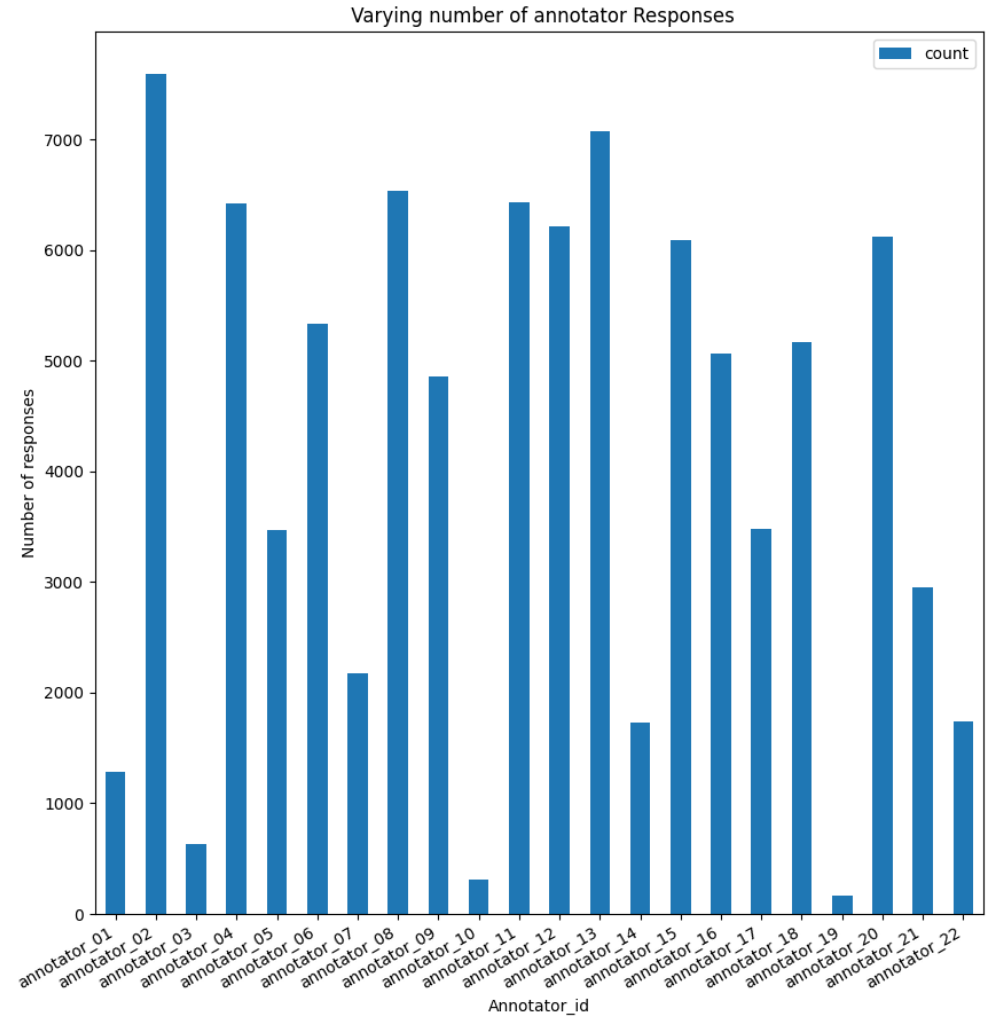
As the number of the 'True' records and 'False' records in the reference dataset(ground truth) does not vary much, it can be considered as a balanced dataset.



ANNOTATOR RESPONSES

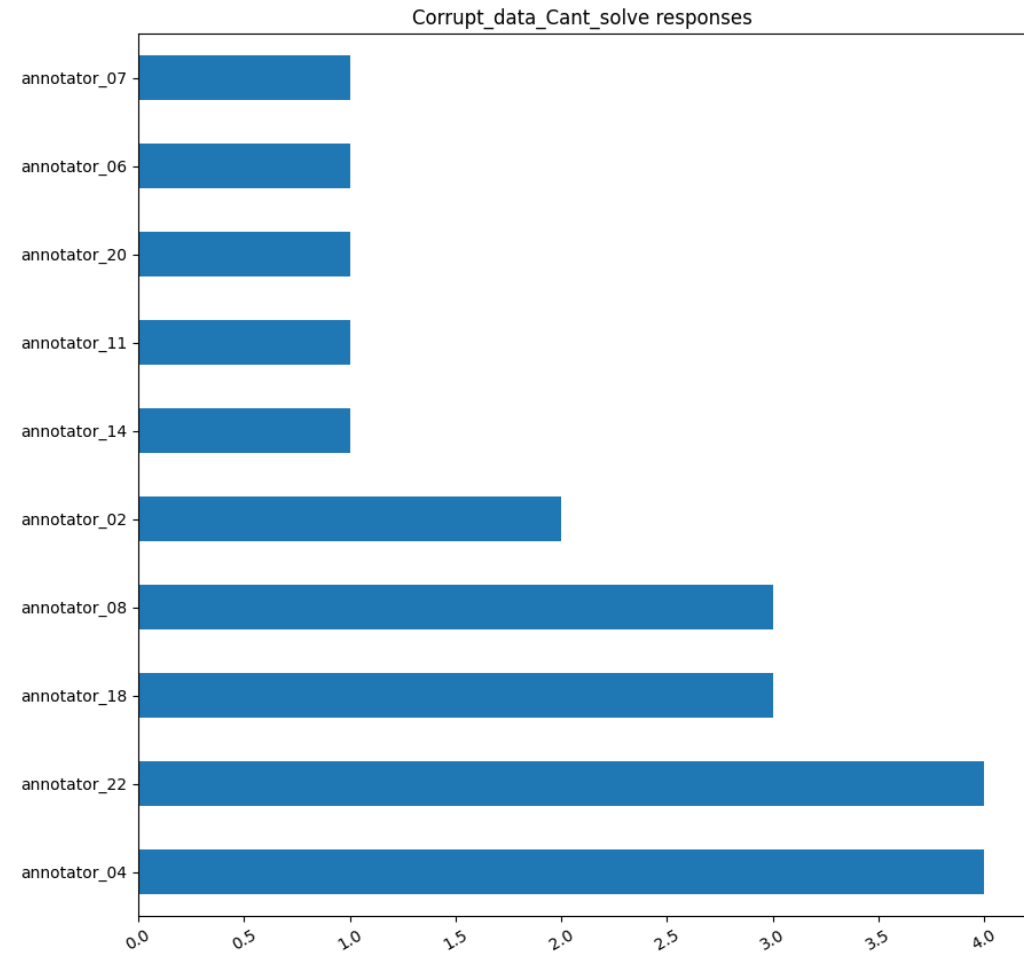
All annotators did not produce the same amount of the results.

Annotators produced varying amount of annotation responses.



UNCERTAIN RESPONSES

Its shows the trend among the annotators that marked the images as 'corrupt data' and 'can't solve'.



INTER- ANNOTATOR AGREEMENT

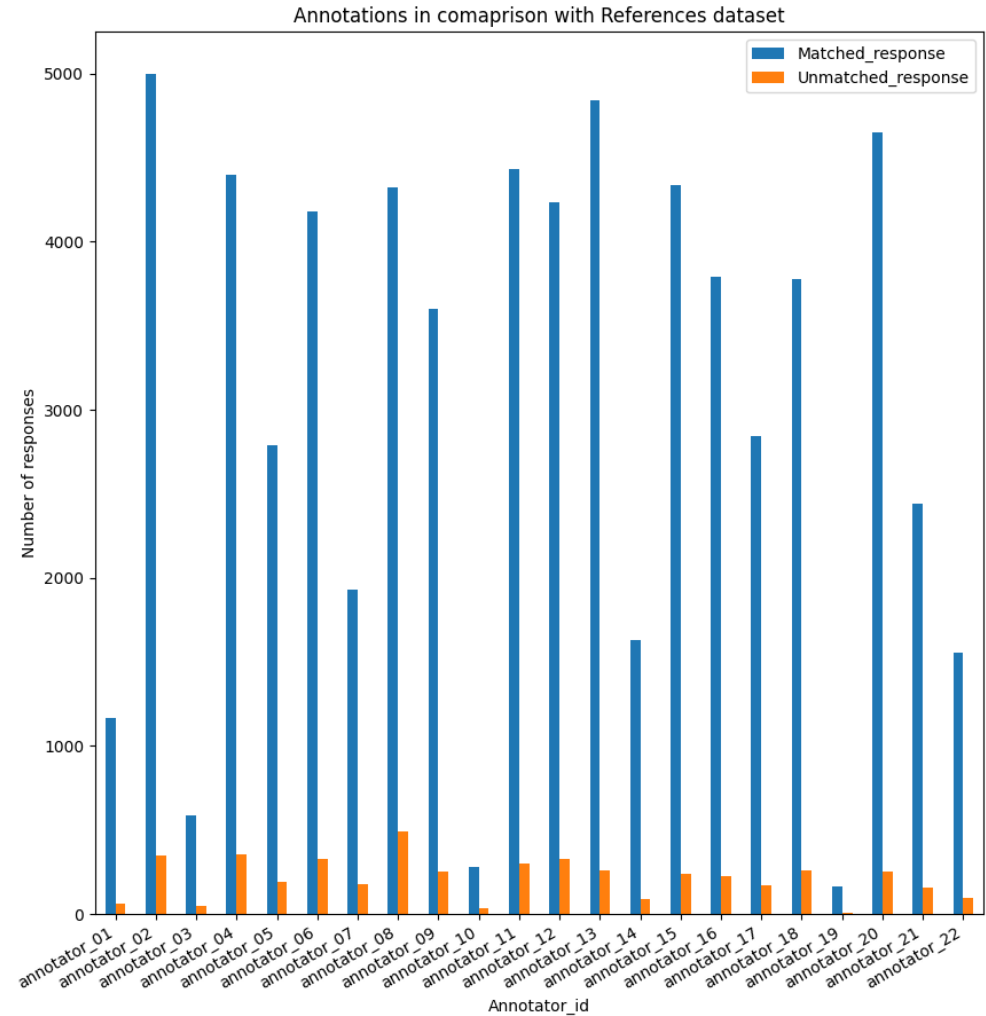
- Inter-annotator agreement for the project dataset: 0.890
- As Inter-annotator agreement i.e., Inter-rater agreement, is within 0.81 – 1.00, so it is highly probable that there are no questions for which annotators highly disagree.
- Metric used for measuring agreement : Fleiss kappa as it must be assessed among more than two annotators.
- kappa Interpretation (source : https://en.wikipedia.org/wiki/Fleiss%27_kappa)

# < 0	Poor agreement
# 0.01 – 0.20	Slight agreement
# 0.21 – 0.40	Fair agreement
# 0.41 – 0.60	Moderate agreement
# 0.61 – 0.80	Substantial agreement
# 0.81 – 1.00	Almost perfect agreement

GOOD AND BAD ANNOTATORS

It shows the number of matched response and unmatched response from the annotators with respect to the ground truth available with reference dataset.

Almost all annotators have different number of annotations responses.



ANNOTATOR PROFICIENCY

It is calculated by using ratio of the number of matched responses to the total number of responses made by each annotator.

$$\text{Proficiency} = \frac{\text{Number of matched response}}{\text{Total Number of responses made}}$$

