AutoML Modeling Report

Adalberto Gonzalez

Binary Classifier with Clean/Balanced Data

Train/Test Split
How much data was used for training? How much data was used for testing?

Labels	Images	Train	Validation	Test
normal	150	120	15	15
pneumonia	149	119	15	15

We used a total of 150 images and reserved 10% of those images for validation and test.

Confusion Matrix
What do each of the cells in the confusion matrix describe? What values did you observe (include a screenshot)? What is the true positive rate for the “pneumonia” class? What is the false positive rate for the “normal” class?

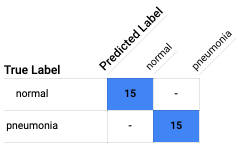
Let’s start with some nomenclature

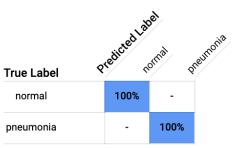
	Predicted Positives	Predicted Negatives
Actual Positives	TP	FN
Actual Negatives	FP	TN

TP - True Positive: correctly classified as positive.
TN - True Negative: correctly classified as negative.
FP - False Positive: Misclassified a positive.
FN - False Negative: Misclassified as negative.

The TP quadrant describes the number of normal images classified correctly, similar situation with the TN quadrant. The confusion matrix shows that the model did a equally good job classifying the images.

What is the true positive rate for the “pneumonia” class?
R=100%
What is the false positive rate for the “normal” class?
R= 0%



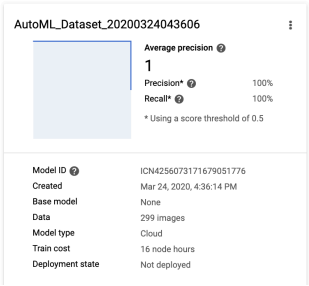


Precision and Recall
What does precision measure? What does recall measure? What precision and recall did the model achieve (report the values for a score threshold of 0.5)?

Precision and recall help us understand how well our model is capturing information, and how much it's leaving out.

Precision tells us, from all the test examples that were assigned a label, how many actually were supposed to be categorized with that label.

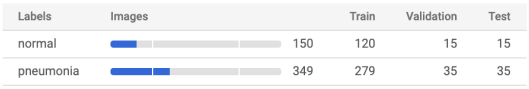
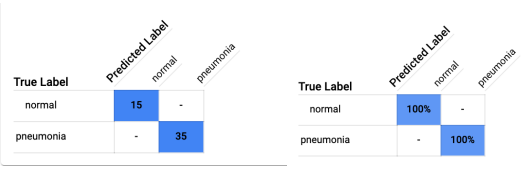
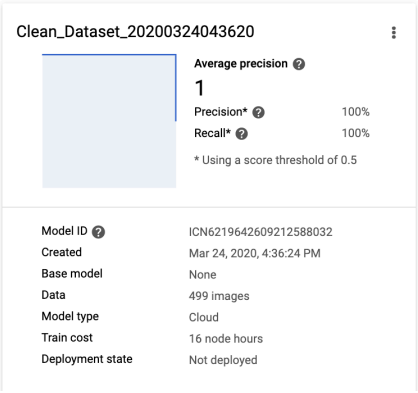
Recall tells us, from all the test examples that should have had the label assigned, how many were actually assigned the label.

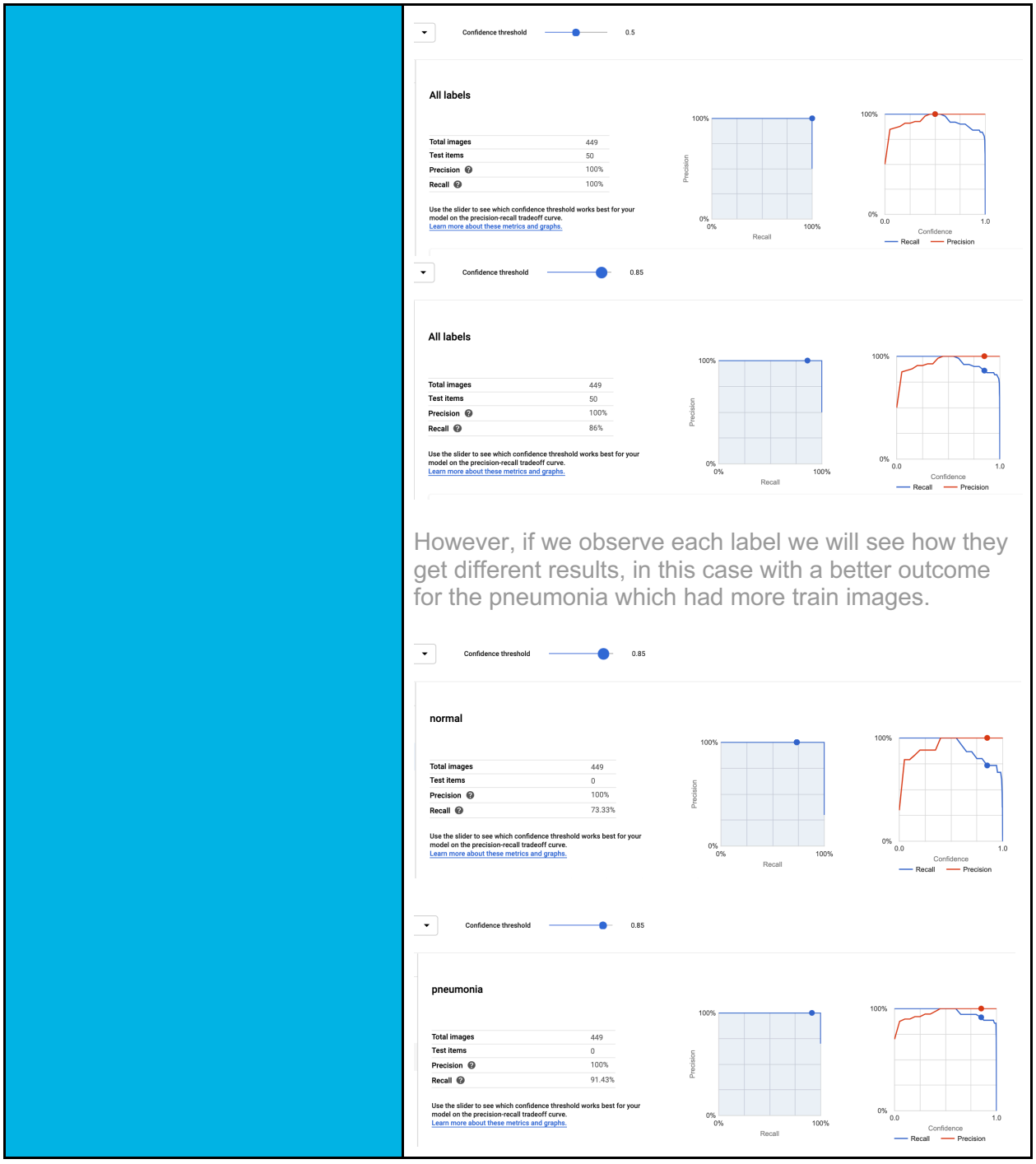


Score Threshold
When you increase the threshold what happens to precision? What happens to recall? Why?

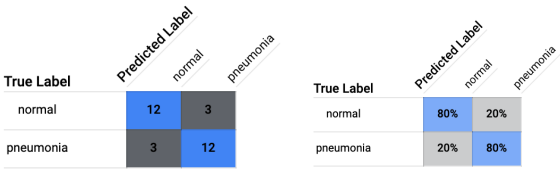
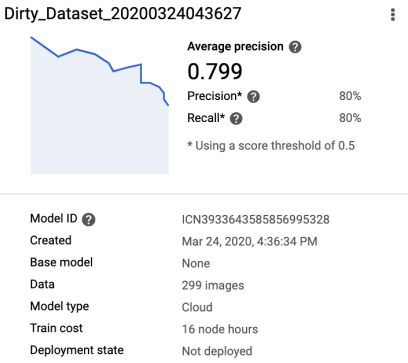
If your score threshold is low, your model will classify more images, but runs the risk of misclassifying a few images in the process. If your score threshold is high, your model will classify fewer images, but it will have a lower risk of misclassifying images.
With that in count, based on our modell, increasing the threshold lowers the recall but the precision stays in 100%.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	 10% of the data was assigned to validation and test respectively.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	 This unbalance did not affect the confusion matrix result.
Precision and Recall How have the model's precision and recall been affected by the unbalanced data (report the values for a score threshold of 0.5)?	 We do not observe precision and recall being affected by unbalanced data.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	I think we have to observe with more detail the results, if we observe the general results, we will see there's not so much effect.

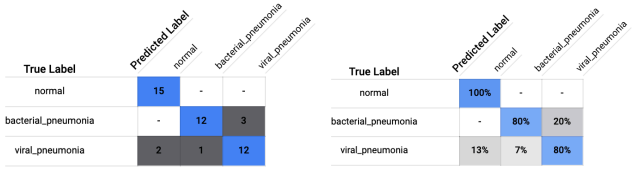


Binary Classifier with Dirty/Balanced Data

Confusion Matrix How has the confusion matrix been affected by the dirty data? Include a screenshot of the new confusion matrix.	 We had a significant increase of FP and FN that is consistent with the % of dirty data we included. With this result (80-20 ratio), we appreciate the importance of using proper data in our datasets.
Precision and Recall How have the model's precision and recall been affected by the dirty data (report the values for a score threshold of 0.5)? Of the binary classifiers, which has the highest precision? Which has the highest recall?	 Precision and recall are lower result in both metrics respectively. Which has the highest Precision? The Clean Balanced and Clean Unbalanced both have the same better Precision and Recall.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	 Dirty data impact the confidence and result of your model, we can observe this directly in the Average Precision. The close to 1.0 score the better our model is performing; A model guessing at random for each label would get an average precision around 0.5. ours is at 0.79% so we are not in the best results.

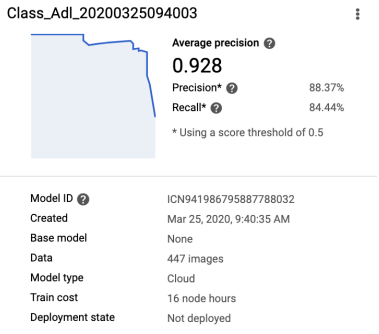
3-Class Model

Confusion Matrix
Summarize the 3-class confusion matrix. Which classes is the model most likely to confuse? Which class(es) is the model most likely to get right? Why might you do to try to remedy the model's "confusion"? Include a screenshot of the new confusion matrix.



The Viral and Bacterial pneumonia are the ones that offer the biggest challenge, since they are subclasses of the label pneumonia.
To remedy this, I would include a larger data set with a proper label for bacteria and viral to create a more stable data set.

Precision and Recall
What are the model's precision and recall? How are these values calculated (report the values for a score threshold of 0.5)?



$$P_{model} = \frac{\sum_{i=1}^{n} P_i}{n}$$

$$R_{model} = \frac{\sum_{i=1}^{n} R_i}{n}$$

Rn=1.00, Rb=0.80, Rv=0.80
Pn=0.88, Pb=0.92, Pv=0.80

Rm= 0.86
Pm= 0.86

F1 Score
What is this model's F1 score?

$$F1 = \frac{2*Precision*Recall}{(Precision+Recall)} = \frac{2*\prod\left(\frac{TP}{(TP+FP)}\right)*\prod\left(\frac{TP}{(TP+FN)}\right)}{\prod\left(\frac{TP}{(TP+FP)}\right)+\prod\left(\frac{TP}{(TP+FN)}\right)}$$

F1 = 0.86