

AutoML Modeling Report



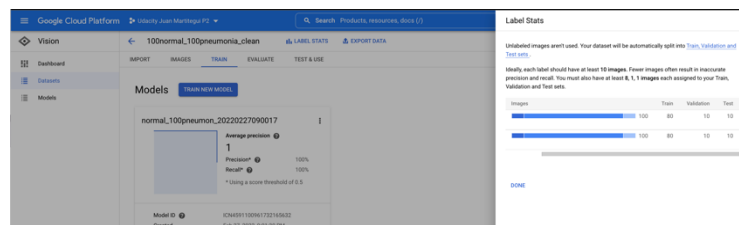
Juan Martitegui

Binary Classifier with Clean/Balanced Data

Train/Test Split

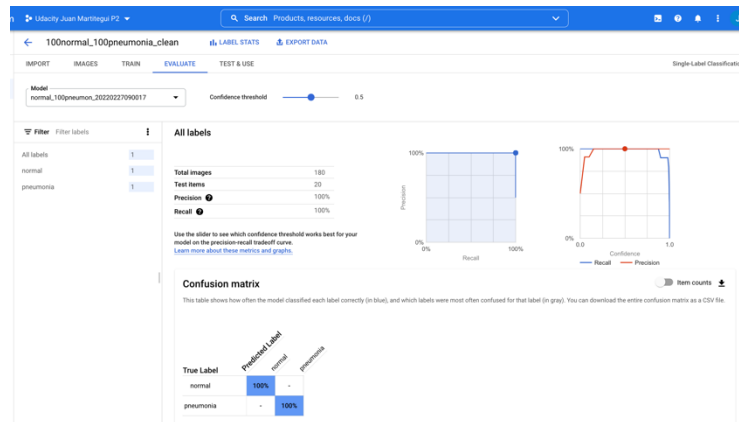
How much data was used for training? How much data was used for testing?

200 Total Images
100 Normal Class (80% Training, 10% Validation, 10% Test)
100 Pneumonia Class (80% Training, 10% Validation, 10% Test)



Confusion Matrix

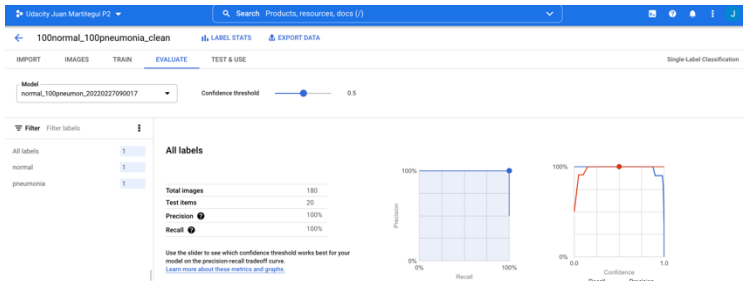
What do each of the cells in the confusion matrix describe? What values did you observe (include a screenshot)? What is the true positive rate for the “pneumonia” class? What is the false positive rate for the “normal” class?



Explanation of each of the cells of confusion matrix

True Positive Rate Normal Class = 100% (which basically means that all the cases are predicted accurately)

False Positive Rate Normal Class = 0% (which means that not a even a single instance/case was confused or

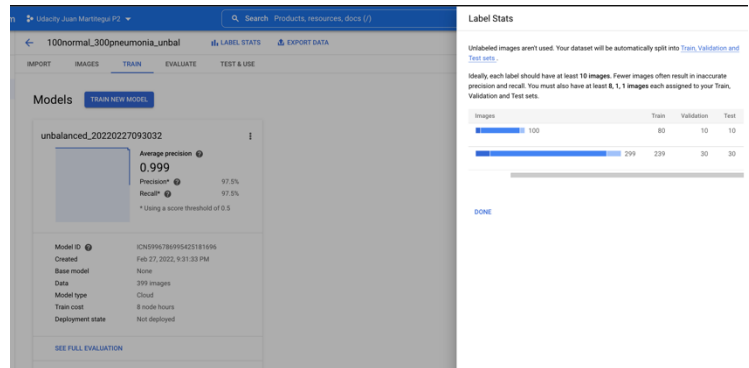
	<p>predicted wrongly or misdiagnosed)</p> <p>True Positive Rate Pneumonia Class = 100% (which basically means that all the cases are predicted accurately)</p> <p>False Positive Rate Pneumonia Class = 0% (which means that not a even a single instance/case was confused or predicted wrongly or misdiagnosed)</p>
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)?	 <p>We call precision to the proportion of relevant results in the list of all returned results, while we call recall the ratio of the relevant results vs the total number of relevant results that could have been returned.</p> <p>This particular model achieved:</p> <ul style="list-style-type: none"> - 100% precision and - 100% recall.
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	<p>When increasing the threshold, the precision seems to go up and the recall seems to decrease. The reason for this is because when you increase the score threshold you want to have higher confidence when you make the prediction.</p> <p>So if you increase the score threshold you'll be classifying a fewer images but with lower risk of mislabeling or misclassifying.</p>

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for	I've uploaded (multiple times) 400 Total Images, but I don't know why I was getting 1 image rejected multiple
---	---

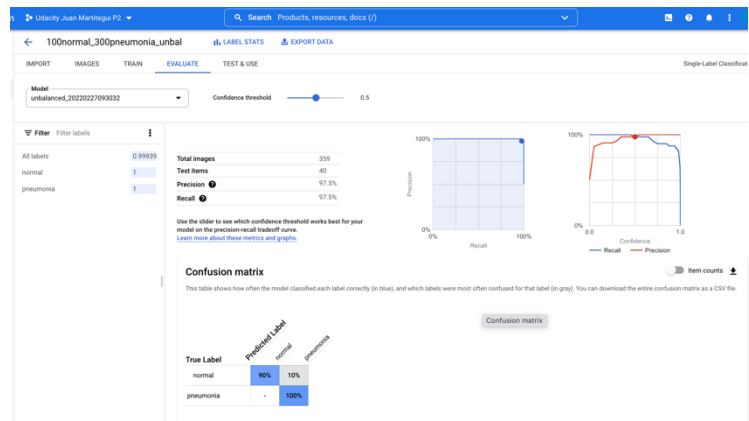
training? How much data was used for testing?

times (and I don't know how to add one single more). As 1 single image wouldn't make much difference I decided to go ahead.
100 Normal Class (80% Training, 10% Validation, 10% Test)
299 Pneumonia Class (80% (239) Training, 10% (30) Validation, 10% (30) Test)



Confusion Matrix

How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.



The confusion matrix was affected a bit by unbalanced data in this case. I don't think the change is due to only one image missing. When you think about it, it only changed because one single case that was "confused"

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)?

They weren't affected by a lot. With the new unbalanced classes.

This particular model achieved:

- 97.5% precision and
- 97.5% recall.

Unbalanced Classes

From what you have observed, how do unbalanced classes affect

It doesn't look like in this particular case having unbalanced data had a HUGE impact on the precision or recall of the model. The % of change was less than 3%

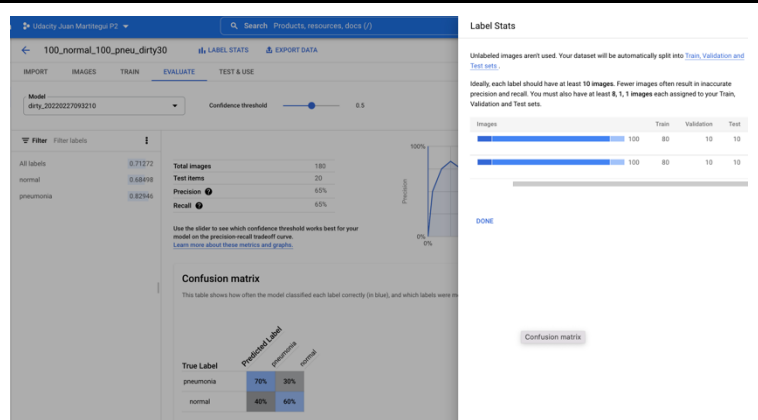
a machine learning model?

and we are talking that we introduced a bias of 3X (3X more images of one class than another)

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.



Wow! I was really surprised to see this change. Here you really understand that this thing is learning and working!

Changes in the Confusion Matrix

In this case you see the following stats on the Confusion Matrix.

True Positive Rate Pneumonia Class = 70%
False Positive Rate Pneumonia Class = 30%

True Positive Rate Normal Class = 60%
False Positive Rate Normal Class = 40%

The presence of "Dirty Data" clearly makes finding patterns and learning a lot more difficult (or even impossible as we are feeding WRONG data on purpose to the ML). Dirty data results in several wrong classifications because the data fed was WRONG, which then results in poor model performance. This goes to show the importance of a proper labelling process for training. If not, it seems that garbage-in, garbage-out type of situation.

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?

This model achieved:

- 65% precision and
- 65% recall

The model with balanced clean data was the best one. It achieved 100% in both scores.

Dirty Data

From what you have observed, how does dirty data affect a machine learning model?

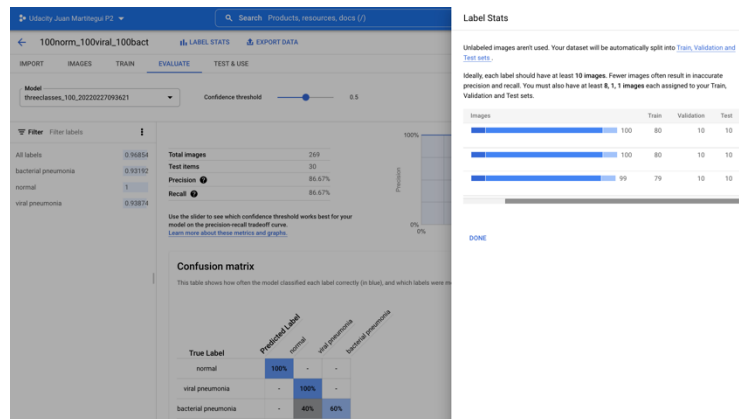
As imagined, AutoML has a very difficult time to find patterns and learn from mixed data (as expected) cause we are introducing confusing inputs/information, especially if we do it in such high % of the total images we are using for learning.

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.

(*) Note: here again despite several attempts of uploading the zip file only 299 images were accepted. There is one that was keep being rejected.



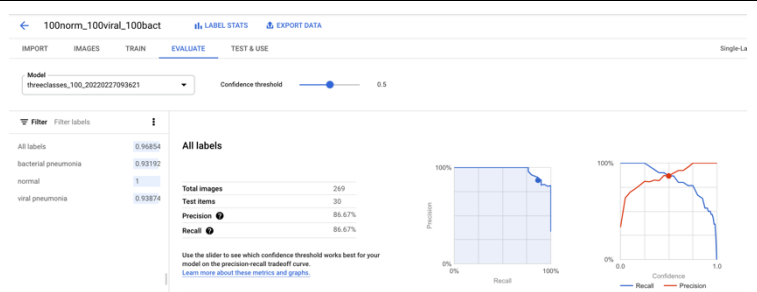
The model is more likely to confuse "Bacterial pneumonia" with "Viral Pneumonia".

The model is more likely to get right "Normal" Classes and "Viral Pneumonia" but struggles a bit with "Bacterial Pneumonia"

Including more images/samples of each class might help remedy this confusion as we are only feeding the model with 100 images each. Having more training cycles might also do the trick.

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)?



This model achieved:

- 86.67% precision and
- 86.67% recall

In a multi-class classification problem these values are calculated by taking the average precision and recall of the individual classes.

The generalization looks something like this:

(*) [source](#)

$$\text{Precision}_i = \frac{M_{ii}}{\sum_j M_{ji}}$$

$$\text{Recall}_i = \frac{M_{ii}}{\sum_j M_{ij}}$$

That is, precision is the fraction of events where we *correctly* declared *i* out of all instances where the algorithm declared *i*. Conversely, recall is the fraction of events where we correctly declared *i* out of all of the cases where the true of state of the world is *i*.

So to calculate the value of precision and recall, the individual values for each class are calculated first and then followed by taking average of those values.

F1 Score

What is this model's F1 score?

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$2 * \frac{0.8667 * 0.8667}{0.8667 + 0.8667} = \frac{0.75116889}{1.7334} = 0.8667$$

