

AutoML Modeling Report



Manuel Quijada

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	170	136	17	17
pneumonia	169	135	17	17

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?

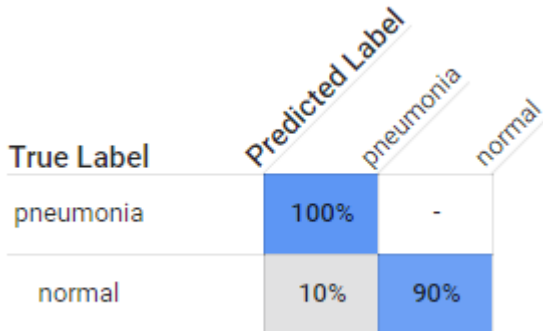
True Label	Predicted Label	
	normal	pneumonia
normal	94%	6%
pneumonia	18%	82%

The Confusion Matrix is showing a strong diagonal from the top left. What this is telling us is that for all of our normal. When we uploaded our images to AutoML, it automatically split our images into training and testing. Therefore, taking most of our images, using those to train the model, and then it reserved a subset of our images to see how the model performed on images that it had never seen before. So what this is telling us is that for all our normal images in our test set, our model was able to identify 94% of them correctly, which is very good and for our pneumonia images our model was able to identify 82% of them correctly, which is pretty good as well.

94% True Positives	6% False Positive
18% False Negatives	82% True Negative

<p>Precision and Recall</p> <p>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>	<p>Precision*</p> <p>88.24%</p> <p>Recall*</p> <p>88.24%</p> <p>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.</p>
<p>Score Threshold</p> <p>When you increase the threshold what happens to precision? What happens to recall? Why?</p>	<p>When the confidence threshold increases Precision increases and Recall decreases. When the confidence threshold decreases Precision decreases and Recall increases. The score threshold tool allows you to explore how your chosen score threshold affects your precision and recall. As you drag the slider on the score threshold bar, you can see where that threshold places you on the precision-recall tradeoff curve, as well as how that threshold affects your precision and recall individually. This can help you find a good balance between false positives and false negatives.</p>

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	<table><tr><th>Labels</th><th>Images</th><th>Train</th><th>Validation</th><th>Test</th></tr><tr><td>normal</td><td>100</td><td>80</td><td>10</td><td>10</td></tr><tr><td>pneumonia</td><td>299</td><td>239</td><td>30</td><td>30</td></tr></table>	Labels	Images	Train	Validation	Test	normal	100	80	10	10	pneumonia	299	239	30	30
Labels	Images	Train	Validation	Test												
normal	100	80	10	10												
pneumonia	299	239	30	30												
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	<div></div> <p>If we look at our images in our test set for all our pneumonia images in our test set, our model identified 100% of them correctly. The best accuracy is 100% indicating that all the predictions are correct. For an imbalanced dataset, accuracy is not a valid measure of model performance. For a dataset where the default rate is 5%, even if all the records are predicted as 0, the model will still have an accuracy of 95%.</p>															
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)?	<table><tr><td>Precision</td><td>97.5%</td></tr><tr><td>Recall</td><td>97.5%</td></tr></table>	Precision	97.5%	Recall	97.5%											
Precision	97.5%															
Recall	97.5%															
Unbalanced Classes From what you have observed, how do unbalanced classes affect a machine learning model?	<p>For an imbalanced dataset, accuracy is not a valid measure of model performance. For a dataset where the default rate is .05 confidence threshold even if all the records are predicted as 0, the model will still have an accuracy of 95%.</p>															

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.

True Label	Predicted Label	
	normal	pneumonia
normal	80%	20%
pneumonia	23%	77%

80% True Positives 20% False Positive
23% False Negatives 77% True Negative

So what this is telling us is that for all our normal images in our test set, our model was able to identify 80% of them correctly, dropping from 94%. For our pneumonia images our model was able to identify 77% of them correctly, dropping from 82% as well.

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

78.26%

Recall

78.26%

The normal binary classification model performed best.

Dirty Data

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

Dirty data can be a major hindrance. The data is inaccurate, incomplete or inconsistent data. It runs contrary to widely shared standards and images of what a person or group should be

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.

True Label	Predicted Label		
	bacteria pneumonia	normal	viral pneumonia
bacteria pneumonia	70%	10%	20%
normal	-	100%	-
viral pneumonia	10%	10%	80%

The model class most like to confuse is the normal image. The model classes most like to confuse are the bacteria and viral pneumonia images. Our model is only as good as the training data that we give it. If we look we can see that a percentage of normal are being mislabeled and if we click on images we can see where our model is getting confused that could have been labeled incorrectly. Therefore what the confusion matrix can help us do is identify where we need to go back and improve our training data.

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)?

Precision

86.21%

Recall

83.33%

Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives. While **recall** expresses the ability to find all relevant instances in a dataset, **precision** expresses the proportion of the data points our **model** says was relevant actually were relevant.

F1 Score

What is this model's F1 score?

Model's F1 score = .81