

Auto ML Modeling Report



Tamer Abdelaty

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		80	10	10
PNEUMONIA		80	10	10

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	90%	10%
PNEUMONIA	10%	90%

The values in the blue squares refer to true predictions positive or negative while the other non blue light gray squares refer to false prediction positive or negative. So in the second row PNEUMONIA was predicted (true positive)90% of 10 tested images (9 times) and was only once predicted as normal (negative) or 10%. and in first row NORMAL was predicted 90% positive and once (10%) as PNEUMONIA(negative).

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 measures true prediction over total prediction, While Recall measures true prediction over ground truth.



Score Threshold

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



Precision score is proportional to confidence and vice versa With recall.

Whenever the threshold increases, the model becomes more confident that the label is correct and so precision increases and recall decreases.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split

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

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.

True Label	Predicted Label	
	PNEUMONIA	NORMAL
PNEUMONIA	100%	-
NORMAL	-	100%

Yes, It is greatly improved.

We should expect a small portion of normal at least to be misclassified as pneumonia due to the presence of more pneumonia in the training dataset

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)?



It is increased to 100%

Unbalanced Classes

From what you have observed, how do unbalanced classes affect a machine learning model?

It gets higher scores due to higher exposure to training.

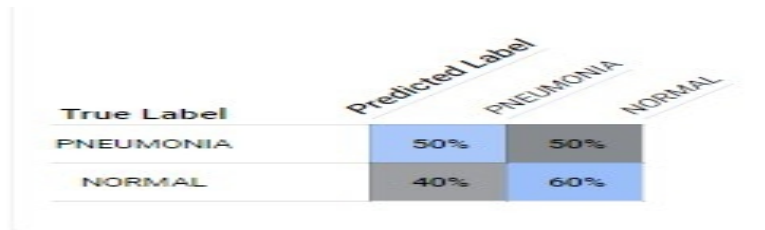
Unbalanced data could create a bias or tendency in the model to one of the predictions.

The model is most likely to be biased towards predicting Pneumonia.

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.



The confusion matrix gets worse due to the dirty data 70/30 Normal and 70/30 pneumonia potentially caused these results.

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?



All labels

Total images	180
Test items	20
Precision ?	55%
Recall ?	55%

It become very sensitive to confidence level with overall decrease in performance by 35% to 55% so that clean/unbalanced data classifier has the highest precision and recall.

Dirty Data

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

The dirty data causes multiple misclassifications which impact the result negatively.

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		
	NORMAL	VIRAL_PNEUMONIA	BACTERIAL_PNEUMONIA
NORMAL	90%	-	10%
VIRAL_PNEUMONIA	-	70%	30%
BACTERIAL_PNEUMONIA	-	30%	70%

Viral and bacterial Pneumonia is equally confused with least true prediction score both(70%). while normal is most likely to get right with the highest prediction score (90%).

To remedy that ,I tried another model with a two classes data-set(bacterial and viral) and got the following result.

	Confidence threshold	0.5
All labels		
Total images	179	
Test items	20	
Precision ?	65%	
Recall ?	65%	

	<table><tr><th rowspan="2">True Label</th><th colspan="2">Predicted Label</th></tr><tr><th>BACTERIAL_PNEUMONIA</th><th>VIRAL_PNEUMONIA</th></tr><tr><th>BACTERIAL_PNEUMONIA</th><td>70%</td><td>30%</td></tr><tr><th>VIRAL_PNEUMONIA</th><td>40%</td><td>60%</td></tr></table>	True Label	Predicted Label		BACTERIAL_PNEUMONIA	VIRAL_PNEUMONIA	BACTERIAL_PNEUMONIA	70%	30%	VIRAL_PNEUMONIA	40%	60%
True Label	Predicted Label											
	BACTERIAL_PNEUMONIA	VIRAL_PNEUMONIA										
BACTERIAL_PNEUMONIA	70%	30%										
VIRAL_PNEUMONIA	40%	60%										
<p>Precision and Recall</p> <p>What are the model's precision and recall? How are these values calculated (report the values for a score threshold of 0.5)?</p>	<div><div>All labels</div><table><tr><td>Total images</td><td>269</td></tr><tr><td>Test items</td><td>30</td></tr><tr><td>Precision ?</td><td>82.14%</td></tr><tr><td>Recall ?</td><td>76.67%</td></tr></table></div> <p>Precision(P) = (P1 + P2 +P3) /3 = 0.82</p> <p>Recall (R) = (R 1 + R 2 + R 3) / 3 = 0.76</p>	Total images	269	Test items	30	Precision ?	82.14%	Recall ?	76.67%			
Total images	269											
Test items	30											
Precision ?	82.14%											
Recall ?	76.67%											
<p>F1 Score</p> <p>What is this model's F1 score?</p>	<p>F1 = 2((82.14*76.67)/(82.14+76.67)) = 79.31%</p>											