

AutoML Modeling Report



Alaa Alyemni

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	Total images=198 Training images=178 Testing images=20											
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?	<div>1-The true label & predicted label. 2-the labels which are normal and pneumonia. 3-the true positive rate for the “pneumonia”= 100% 4-the false positive rate for the “normal”= 0%</div> <div><table><tr><th rowspan="2">True Label</th><th colspan="2">Predicted Label</th></tr><tr><th>pneumonia</th><th>normal</th></tr><tr><th>pneumonia</th><td>100%</td><td>-</td></tr><tr><th>normal</th><td>-</td><td>100%</td></tr></table></div>	True Label	Predicted Label		pneumonia	normal	pneumonia	100%	-	normal	-	100%
True Label	Predicted Label											
	pneumonia	normal										
pneumonia	100%	-										
normal	-	100%										
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>precision is defined as the number of true positives (truly-1-1 fraudulent transaction data, in this case) over all positives, and will be the higher when the amount of false positives is .low</p> <p>recall is defined as the number of true positives over true-2 positives plus false negatives and will be higher when the .number of false negatives is low</p> <p>precision=100%-3</p> <p>recall=100%</p>											
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	<p>When the threshold =1</p> <p>The precision did not change still 100%.</p> <p>The recall decreased to 5%.</p> <p>Because threshold value is like the confidence score, when we increase it then the model will assigns a class to images only when the accuracy is100% which is very hard.</p>											

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	Total images=396 Training images=357 Testing images=39												
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	<p>I've here 10% of the normal's images predicted incorrectly.</p> <table><tr><td></td><td colspan="2">Predicted Label</td></tr><tr><td>True Label</td><td>normal</td><td>pneumonia</td></tr><tr><td>normal</td><td>90%</td><td>10%</td></tr><tr><td>pneumonia</td><td>-</td><td>100%</td></tr></table>		Predicted Label		True Label	normal	pneumonia	normal	90%	10%	pneumonia	-	100%
	Predicted Label												
True Label	normal	pneumonia											
normal	90%	10%											
pneumonia	-	100%											
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)?	Yes, if we see the balanced data we will see 100% for both precision and recall, but here we've precision=97.44% and recall=97.44%.												
Unbalanced Classes From what you have observed, how do unbalanced classes affect a machine learning model?	In previous class both data have same count of images, but here we have unbalanced data so we will got inaccurate result.												

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.	<p>The dirty data affected the accuracy of the result we got here 30% false for each.</p> <table><tr><th></th><th colspan="2">Predicted Label</th></tr><tr><th>True Label</th><th>pneumonia</th><th>normal</th></tr><tr><th>pneumonia</th><td>70%</td><td>30%</td></tr><tr><th>normal</th><td>30%</td><td>70%</td></tr></table>		Predicted Label		True Label	pneumonia	normal	pneumonia	70%	30%	normal	30%	70%
	Predicted Label												
True Label	pneumonia	normal											
pneumonia	70%	30%											
normal	30%	70%											
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?	<p>Yes, if we see the balanced data we will see 100% for both precision and recall, but here we've precision=70% and recall=70%. The accuracy had been affected by the dirty data, which will immediately affect the result of both precision and recall,</p>												
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	<p>The dirty data have negative affect on the ML. It's increase false negative & false positive as well, and the percent of the precision and recall become less.</p>												

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.

- 1- The most likely confusing class is (bacterial pneumonia)
- 2-The most likely getting right classes are viral pneumonia and normal.
- 3-Here we've notice that some of the bacterial pneumonia class has been predicted as viral pneumonia we have to do more trains for.

True Label	Predicted Label		
	viral pneumonia	normal	bacterial pneumonia
viral pneumonia	100%	-	-
normal	-	100%	-
bacterial pneumonia	50%	-	50%

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)?

The precision =83.33%The recall =83.33%.

The precision(
 $p_{Vpneumonia} = 100/150 = 0.67$
 $p_{normal} = 100/100 = 1$
 $p_{Bpneumonia} = 50/50 = 1$
 The precision = $(0.75 + 1 + 1)/3 = 0.89$
)

The recall=(
 $r_{Vpneumonia} = 100/100 = 1$
 $r_{normal} = 100/100 = 1$
 $r_{Bpneumonia} = 50/(50 + 50) = 0.5$
 The recall = $(1 + 1 + 0.5)/3 = 0.83$
)

F1 Score

What is this model's F1 score?

$F1 = 2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall})$
 $F1 = 2 * 0.89 * 0.83 / (0.89 + 0.83) =$
 $F1 = 1.78 * 1.124$
 $F1 = 2$