

# AutoML Modeling Report



*Olena Sergeevna Kormachova*

## Binary Classifier with Clean/Balanced Data

<b>Train/Test Split</b> How much data was used for training? How much data was used for testing?	I used total of 399 (199 pneumonia and 200 normal) images. Out of those 319 were used for training, 39 for validation and 40 for testing.											
<b>Confusion Matrix</b> 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?	<p>This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in gray).</p> <p>Each of the cells describes (from top left to bottom right):</p> <ul style="list-style-type: none"><li>- True positive for “normal” class</li><li>- False negative for “normal” class</li><li>- False positive for “normal” class</li><li>- True negative for “normal” class</li></ul> <p>The true positive rate for the “pneumonia” class is 100% and false positive rate for the “normal” class 0%</p> <div><table><tr><th rowspan="2">True Label</th><th colspan="2">Predicted Label</th></tr><tr><th>normal</th><th>pneumonia</th></tr><tr><th>normal</th><td>85%</td><td>15%</td></tr><tr><th>pneumonia</th><td>-</td><td>100%</td></tr></table></div>	True Label	Predicted Label		normal	pneumonia	normal	85%	15%	pneumonia	-	100%
True Label	Predicted Label											
	normal	pneumonia										
normal	85%	15%										
pneumonia	-	100%										
<b>Precision and Recall</b> 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)?	<div><math display="block">\text{Precision} = \frac{TP}{TP + FP}</math></div> <p>A precision measures what proportion of positive identifications was actually correct.</p> <p>A recall measures the percentage of total relevant results correctly classified by the algorithm</p>											

	$\text{Recall} = \frac{TP}{TP + FN}$ <p>The model achieved a precision of 92.5% and a recall of 92.5%.</p>
<b>Score Threshold</b> When you increase the threshold what happens to precision? What happens to recall? Why?	<p>In general, raising the classification threshold reduces false positives, thus raising precision. Raising classification threshold will cause the number of true positives to decrease or stay the same and will cause the number of false negatives to increase or stay the same. Thus, recall will either stay constant or decrease.</p> <p>After I increased the threshold the precision was 97.37% and the recall was 92.5%.</p>

## Binary Classifier with Clean/Unbalanced Data

<b>Train/Test Split</b> How much data was used for training? How much data was used for testing?	I used total of 394 (294 pneumonia and 100 normal) images. Out of those 315 were used for training, 40 for validation and 39 for testing.											
<b>Confusion Matrix</b> How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	<div>This time true positive for normal class went up to 100%. Pneumonia class is the same as in previous classifier.</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%										
<b>Precision and Recall</b> How have the model's precision and recall been affected by the unbalanced data (report the values for a score threshold of 0.5)?	Both precision and recall now report 100%											
<b>Unbalanced Classes</b> From what you have observed, how do unbalanced classed affect a machine learning model?	For unbalanced data I would expect a model to have poorer predictive performance, especially for the minority class because unbalanced data introduces bias towards the label that has more data in the training. However, from the results above, I conclude that in my case unbalanced data gave a model with better prediction performance.											

## Binary Classifier with Dirty/Balanced Data

<h3>Confusion Matrix</h3> <p>How has the confusion matrix been affected by the dirty data? Include a screenshot of the new confusion matrix.</p>	<table><tr><th>True Label</th><th colspan="2">Predicted Label</th></tr><tr><th></th><th>normal</th><th>pneumonia</th></tr><tr><th>normal</th><td>62%</td><td>38%</td></tr><tr><th>pneumonia</th><td>23%</td><td>77%</td></tr></table> <p>True positive values for both classes went down as shown in the confusion matrix. This is because we added “dirty” data and now the model can’t find a proper pattern between two classes.</p>	True Label	Predicted Label			normal	pneumonia	normal	62%	38%	pneumonia	23%	77%
True Label	Predicted Label												
	normal	pneumonia											
normal	62%	38%											
pneumonia	23%	77%											
<h3>Precision and Recall</h3> <p>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>	<p>The precision and recall values also went down to 69.23% both. The highest precision and recall values were observed with clean/unbalanced data classifier.</p>												
<h3>Dirty Data</h3> <p>From what you have observed, how does dirty data affect a machine learning model?</p>	<p>Since the data is mixed up, the ML model is not performing so well as with previous datasets. This clearly shows how important it is for the training data to be clean.</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.

Model is most likely to confuse with viral pneumonia class and a little bit with bacterial pneumonia class. The model predicted normal class with 100% accuracy. I believe that adding more image examples to training set for pneumonia classes should remedy the model’s “confusion”.

	Predicted Label		
True Label	normal	viral pneumonia	bacterial pneumonia
normal	100%	-	-
viral pneumonia	15%	54%	31%
bacterial pneumonia	-	8%	92%

<p><b>Precision and Recall</b>  What are the model's precision and recall? How are these values calculated (report the values for a score threshold of 0.5)?</p>	<p>Both precision and recall report 82.05% for this model. They are calculated for each class (individual calculation is presented in the clean/balanced section) separately and then the average of them is taken for the whole model.</p>
<p><b>F1 Score</b>  What is this model's F1 score?</p>	<p>F1 Score is the weighted average of Precision and Recall  For our model it is 82.05%.</p>