AutoML Modeling Report



Olena Sergeevna Kormachova

Binary Classifier with Clean/Balanced Data

Train/Test Split I used total of 399 (199 pneumonia and 200 normal) images. How much data was used for Out of those 319 were used for training, 39 for validation and training? How much data was 40 for testing. used for testing? This table shows how often the model classified each label **Confusion Matrix** correctly (in blue), and which labels were most often confused What do each of the cells in the for that label (in gray). confusion matrix describe? Each of the cells describes (from top left to bottom right): What values did you observe True positive for "normal" class (include a screenshot)? What is False negative for "normal" class the true positive rate for the False positive for "normal" class "pneumonia" class? What is the false positive rate for the True negative for "normal" class "normal" class? The true positive rate for the "pneumonia" class is 100% and false positive rate for the "normal" class 0%



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)?

$$Precision = \frac{TP}{TP + FP}$$

A precision measures what proportion of positive identifications was actually correct.

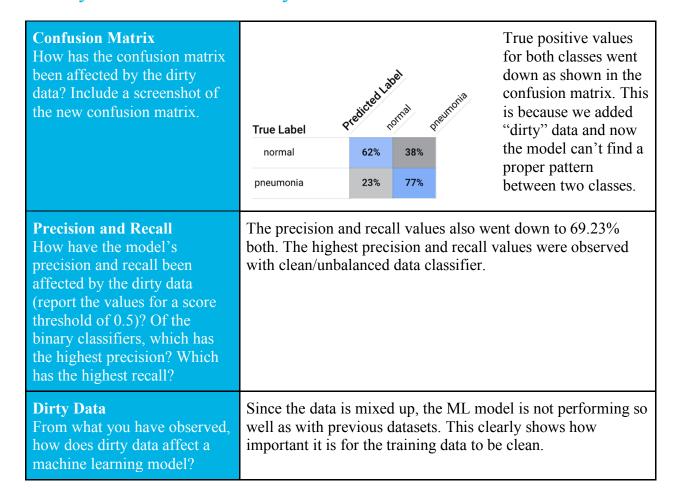
A recall measures the percentage of total relevant results correctly classified by the algorithm

	${\rm Recall} = \frac{TP}{TP + FN}$ The model achieved a precision of 92.5% and a recall of 92.5%.
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	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. After I increased the threshold the precision was 97.37% and the recall was 92.5%.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split 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.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	This time true positive for normal class went up to 100%. Pneumonia class is the same as in previous classifier. True Label pneumonia 100% - normal 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)?	Both precision and recall now report 100%
Unbalanced Classes 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



3-Class Model

Confusion Matrix Model is most likely to confuse with viral pneumonia class Summarize the 3-class and a little bit with bacterial pneumonia class. The model confusion matrix. Which predicted normal class with 100% accuracy. classes is the model most I believe likely to confuse? Which that class(es) is the model most adding likely to get right? Why might more you do to try to remedy the image True Label model's "confusion"? Include examples 100% normal a screenshot of the new to training viral pneumonia 15% 54% 31% confusion matrix. set for bacterial pneumonia 92% pneumonia classes should remedy the model's "confusion".

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)?	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.
F1 Score What is this model's F1 score?	F1 Score is the weighted average of Precision and Recall For our model it is 82.05%.