AutoML Modeling Report

Ashkan Yousefi

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	I used 271 items as normal images and 281 items as pneumonia images to be uploaded into the Google ML. I used 30% of the training set for testing and the rest for training
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?	There are two main values in the confusion matrix i.e. precision and recall. The precision describes how accurately the model can predict and the formulation is calculated as follows: TP/(TP+FP) i.e. TP (True Positive) and FP (False Positive) The recall is calculated as the correct number of identified items to the total number of items analyzed and the formulation will be as follows: TP/(TP+FN) Normal case: Precision = 100/(100+0) = 100% Recall = 100 / (100 + 4) = 96.1% The confusion matrix has four major cells. The items in

	the left side of the matrix is the actual label and the items located in the upper part of the matrix is the predicted values for the labels. For example, 100% of the normal cases are predicted correctly by the model. However, in the case of pneumonia 4% of the pneumonia cases predicted as normal. Pneumonia Precision: 96/(96+4) = 96% Recall : 96/(96+0)=100% **What is the true positive rate for the "pneumonia" class?** The true positive rate for the pneumonia case is 96% **What is the false positive rate for the "normal" class?** The false positive rate for the normal case is 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 means how accurate the model can identify the normal cases and the cases for pneumonia in the current project. In the other words, out of the determined cases for pneumonia how accurate the system can predict the pneumonia cases. Recall means how many of the actual cases with pneumonia is actually identified by the system out of all the samples. Normal case: Precision = 100/(100+0) = 100% Recall = 100 / (100 + 4) = 96.1% Pneumonia Precision: 96/(96+4) = 96% Recall : 96/(96+0)=100%
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	When we increase the threshold the precision level increases but the recall level goes down. I have compared two cases for the precision and the recall with confidence level of 0.65 and the confidence level of 0.14

Binary Classifier with Clean/Unbalanced Data

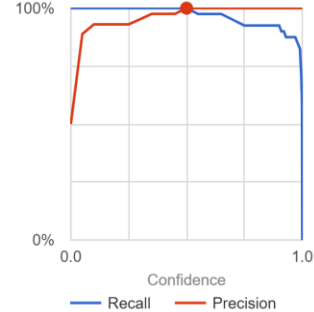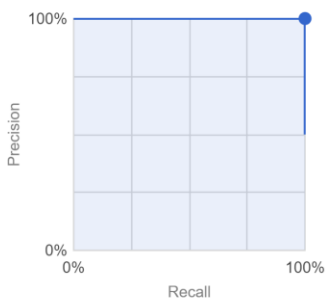|	
Train/Test Split How much data was used for training? How much data was used for testing?	The confidence level of 0.65: Total images — 496 Test items — 55 Precision ❓ — 98.18% Recall ❓ — 98.18% The confidence level of 0.14: Total images — 496 Test items — 55 Precision ❓ — 98.21% Recall ❓ — 100%
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	Total images — 359 Test items — 40 Precision ❓ — 100% Recall ❓ — 100%

	The confusion matrix improved as the data input changed to unbalance and I was not expecting the such result.
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)?	 The confidence level of the case considered to be 0.5 The confidence level of the 0.5 gives the optimum values for the Recall and the Precision
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	The unbalanced data improved the precision and the recall for the confusion matrix. However, I was expecting different results.

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 impact of the dirty data which includes the pneumonia images mixed with the normal image is presented in the following:

All labels

Total images	180
Test items	20
Precision ❓	78.95%
Recall ❓	75%

True Label	Predicted Label normal	pneumonia
normal	90%	10%
pneumonia	30%	70%

The confusion matrix with the dirty data will be different and the precision and recall values changed greatly as a result of dirty data in the data set.

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?

The other interesting point is related to the confidence level and in the confidence level of 0.5 there is balance for recall and precision. However, as we increase the confidence values the precision goes up and the recall values goes down.

Dirty Data From what you have observed, how does dirty data affect a machine learning model?	I learnt that that the dirty data could have a great impact on the model performance and I did not know that the wrong labelling of the data can reduce the performance of the model so much.

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.	

Total images	269
Test items	30
Precision ❓	76.47%
Recall ❓	86.67%

True Label	Predicted Label bacterial_pneumonia	viral_pneumonia	normal
bacterial_pneumonia	80%	10%	10%
viral_pneumonia	-	100%	-
normal	-	10%	90%

The confusion matrix related to the multi model is presented above. There are three categories to be considered which includes: bacterial and viral pneumonia as well as the normal. The model performed very well for the viral pneumonia and the normal case. However, there is a performance gain which could be achieved as a result of improvement in the bacterial pneumonia category. To improve the mentioned category, I would increase the number of cases for th bacterial pneumonia in the training set and will check the performance of the model again:

Total images	269
Test items	30
Precision ❓	76.47%
Recall ❓	86.67%

The result shows that the confidence level can still increase to achieve to the optimum point for the precision and recall values. The optimum confidence value could be 0.9

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 means how accurately the model can identify and distinguish the cases from each other and assign them to different categories. As an example, out of the cases determined as normal category how accurately the system can predict the correct normal cases. Recall means how many of the actual cases with normal category is identified by the system out of all the samples.
F1 Score What is this model's F1 score?	The In statistical analysis of binary classification, the F1 score (also F-score or F-measure) is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score: p is the number of correct positive results divided by the number of all positive results returned by the classifier, and r is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive).

The F1 score is the harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

The formula for the calculation of the F1 score is calculated as follows:

$$2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

In the current case the F1 score will be calculated as follows:

2.(0.76)(0.86) / (0.76+ 0.86) = 0.806