AutoML Modeling Report

<B.Akintade>

Binary Classifier with Clean/Balanced Data

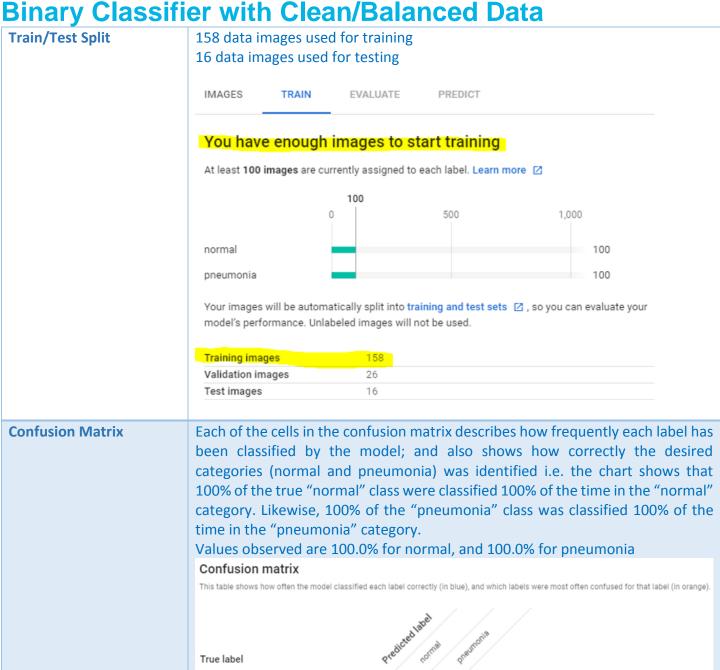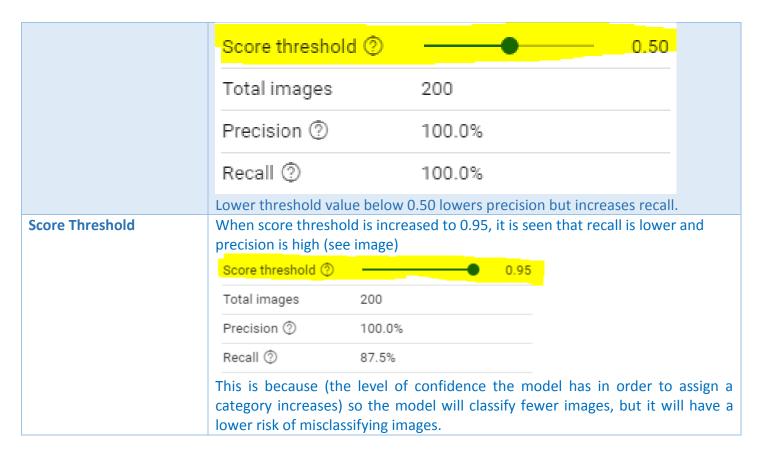|	
Train/Test Split	158 data images used for training 16 data images used for testing IMAGES **TRAIN** EVALUATE PREDICT **You have enough images to start training** At least **100 images** are currently assigned to each label. Learn more ☑ normal 100 pneumonia 100 Your images will be automatically split into training and test sets ☑ , so you can evaluate your model's performance. Unlabeled images will not be used. Training images 158 Validation images 26 Test images 16
Confusion Matrix	Each of the cells in the confusion matrix describes how frequently each label has been classified by the model; and also shows how correctly the desired categories (normal and pneumonia) was identified i.e. the chart shows that 100% of the true "normal" class were classified 100% of the time in the "normal" category. Likewise, 100% of the "pneumonia" class was classified 100% of the time in the "pneumonia" category. Values observed are 100.0% for normal, and 100.0% for pneumonia **Confusion matrix** This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in orange). Predicted label True label normal pneumonia normal 100.0% - pneumonia - 100.0% True positive rate for the "pneumonia" class: "pneumonia" was predicted positive and it's true i.e I predicted that the patient actually had "pneumonia" and it's true

	False positive rate for the "normal" class: "normal" case was positively predicted and it's false i.e I predicted that the patient is "normal" and does NOT have pneumonia, but turns out prediction was false
	Reported values for the true positive rate for the "pneumonia" class is 100% - indicating that the image is labelled as positive (for pneumonia) and the model outcome result is correct/positive i.e. pneumonia is present

Reported values for the false positive rate for the "normal" class is 100% - indicating that the image is labelled as positive (for pneumonia) but the model outcome result is negative

Both reported values of 100% for true positive rate for the "pneumonia" class and false positive rate for the "normal" class also show high precision of 100% which means the model is identifying relevant instances of pneumonia where each label is properly assigned and shows a *perfect matrix without misclassification or ambiguity.

High precision of 100% also indicate that every child identified as testing positive for pneumonia actually have pneumonia, but the model may miss some children that do have pneumonia, since it is getting false negatives. |
| **Precision & Recall** | Both precision and recall are two common measures used to evaluate a model in order to understand how the model performs for an individual class, and how it performs across classes. They also provide understanding to how accurately well a model is capturing information, and how much information it is leaving out.

Precision helps to measure exactly how many labelled images/data that are supposed to be categorized with a specific label, from all the test examples that was assigned a label ('normal' or 'pneumonia'). Modeled precision answers the question of when the model makes a prediction, how likely is that prediction to be correct?

Recall helps to measure the specific number of labels that was actually assigned from all the test examples that should have had the label assigned ('normal' or 'pneumonia'). Model recall answers the question of how good is a model at identifying actual occurrences of objects in the data. This will give us an understanding of whether or not the model can recognize these objects

For a score threshold of 0.5, the model achieved a precision of 100.0% and recall of 100.0% (see image). Since the model achieved a 'perfect' precision of 100.0% for both, this indicates that the data may not generalize well as it was too easy. |

	Score threshold ⑦ ━━━━━●━━━━━ **0.50**
	Total images 200
	Precision ⑦ 100.0%
	Recall ⑦ 100.0%
	Lower threshold value below 0.50 lowers precision but increases recall.
Score Threshold	When score threshold is increased to 0.95, it is seen that recall is lower and precision is high (see image)
	Score threshold ⑦ ━━━━━━━●━ 0.95
	Total images 200
	Precision ⑦ 100.0%
	Recall ⑦ 87.5%
	This is because (the level of confidence the model has in order to assign a category increases) so the model will classify fewer images, but it will have a lower risk of misclassifying images.

EVALUATE

IMAGES   TRAIN   **EVALUATE**   PREDICT

Model

AutoMLDataset_v20190914021900 ⌄

Created	Analyzed	Avg precision ⑦	Precision ⑦	Recall ⑦
Sep 13, 2019	200 images	1.0	100.0%	100.0%
1 compute hour	2 labels, 16 test images			

Precision and recall are based on a score threshold of 0.5

Model

AutoMLDataset_v20190914021900 ⌄

Created	Analyzed	Avg precision ⑦	Precision ⑦	Recall ⑦
Sep 13, 2019	200 images	1.0	100.0%	100.0%
1 compute hour	2 labels, 16 test images			

Precision and recall are based on a score threshold of 0.5

Score threshold ⑦   ●——— 0.50

Converting between the models output scores to a set of applicable labels requires a threshold. Lower thresholds typically increase recall but lower precision.

Use the slider to see which score threshold works best for your model on the precision-recall tradeoff curve. Learn more about these metrics and graphs ⬀

Score threshold: 0.50
Recall: 1.000
Precision: 1.000

Top-left chart

Precision@1 (y-axis): 1.00, 0.75, 0.50, 0.25, 0.00

Score threshold: 0.50
Precision: 1.000

Score threshold (x-axis): 0.0, 0.5, 1.0

Top-right chart

Precision@1 (y-axis): 1.00, 0.75, 0.50, 0.25, 0.00

The precision@1 curve shows how your model performs on the top-scored label. A higher score threshold produces fewer false positives, which increases precision.

Score threshold (x-axis): 0.0, 0.5, 1.0

(second partial axis): 1.00, 0.25, 0.00, 0.0

Score threshold

Bottom-left chart

Recall@1 (y-axis): 1.00, 0.75, 0.50, 0.25, 0.00

Score threshold: 0.50
Recall: 1.000

Score threshold (x-axis): 0.0, 0.5, 1.0

Bottom-right chart

The recall@1 curve shows how your model performs on the top-scored label. A lower score threshold produces fewer false negatives, which increases recall.

Recall@1 (y-axis): 0.75, 0.50, 0.25, 0.00

Score threshold (x-axis): 0.0, 0.5, 1.0